(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 January 2024 (18.01.2024)

(10) International Publication Number
WO 2024/013514 A2

(51) International Patent Classification:
A01K 67/027 (2024.01) *C12N 9/64* (2006.01)
C12N 9/22 (2006.01) *C12N 15/113* (2010.01)

(21) International Application Number: PCT/GB2023/051854

(22) International Filing Date: 14 July 2023 (14.07.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/368,603 15 July 2022 (15.07.2022) US

(71) Applicant: **PIG IMPROVEMENT COMPANY UK LIMITED** [GB/GB]; Matrix House, Basing View, Basingstoke Hampshire RG21 4DZ (GB).

(72) Inventors: **BURGER, Brian**; c/o Genus PLC, 1525 River Road, DeForest, Wisconsin 53532 (US). **BEATON, Benjamin**; c/o Genus PLC, 1525 River Road, DeForest, Wisconsin 53532 (US).

(74) Agent: **J A KEMP LLP**; 80 Turnmill Street, London Greater London EC1M 5QU (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2024/013514 A2

(54) **Title:** GENE EDITED LIVESTOCK ANIMALS HAVING CORONAVIRUS RESISTANCE

(57) **Abstract:** Livestock animals and progeny thereof comprising at least one edited chromosomal sequence that reduces expression or activity of an ANPEP protein are provided. Livestock animal cells that contain such edited chromosomal sequences are also provided. The livestock animals, progeny, and cells have increased resistance to coronavirus. Methods for producing coronavirus resistant livestock animals are also provided.

GENE EDITED LIVESTOCK ANIMALS HAVING CORONAVIRUS RESISTANCE

CROSS REFERENCE TO RELATED APPLICATION

5 This application claims benefit of and priority to US Provisional application 63/368,603, filed on July 15, 2022. This application is hereby incorporated by reference in its entirety.

SEQUENCE LISTING

10 The Sequence Listing, including the file named TD-14-2022-WO1.xml is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

The present teachings relate to gene edited livestock animals and the editing of 15 amino peptidase N (ANPEP) to provide coronavirus resistance.

BACKGROUND

Respiratory and enteric infections caused by coronaviruses have important impacts to both human and animal health. Infection of immunologically naïve newborn pigs with 20 transmissible gastroenteritis virus (TGEV) or porcine epidemic diarrhea virus (PEDV) can incur losses approaching 100% mortality; the result of dehydration caused by the virus-mediated destruction of enterocytes resulting in a malabsorptive diarrhea and dehydration (Madson, D. M., *et al.*, Veterinary Pathology, 2016, 53, 44-52; Saif, L. J., *et al.*, Coronaviruses, Diseases of Swine, Chapter 35, 10th Edition, 2012, Wiley-Blackwell.). 25 TGEV first appeared in the US in the 1940s (Doyle, L. P., *et al.*, Journal of the American Veterinary Medical Association, 1946, 108, 257-259). The more recent emergence of porcine epidemic diarrhea virus (PEDV) in 2013 was responsible for the death of nearly seven million pigs in the US, an estimated 10% loss in pig production (Stevenson, G. W., *et al.*, Journal of Veterinary Diagnostic Investigation, 2013, 25, 649-654). TGEV can also 30 cause 100% neonatal mortality. In older pigs, infection with TGEV or PEDV results in only mild clinical signs followed by complete recovery.

Along with the human, canine, and feline coronaviruses, PEDV and TGEV belong to the genus Alphacoronavirus in the family Coronaviridae (Lin, C.-M., *et al.*, Journal of Virology, 2015, 89, 3332-3342). Porcine respiratory coronavirus (PRCV) is also an Alphacoronavirus and is closely related to TGEV. PRCV generally causes subclinical infection or mild respiratory disease, but severe cases have been described and there is evidence that it may worsen the severity of disease when pigs are dually infected with both PRCV and another virus such as porcine respiratory and reproductive syndrome virus (PRRSV) (Killoren, K. E. *et al.*, Porcine Respiratory Coronavirus, Swine Health Information Center and Center for Food Security and Public Health, 2016, www.cfsph.iastate.edu; Van Reeth, K., *et al.*, Veterinary Microbiology, 1996, 48, 325-335). Moreover, PRCV-positive status of a herd may have economic implications, because some countries will not import animals that are PRCV-positive.

Coronaviruses are enveloped, single stranded, positive sense RNA viruses, placed in the order, Nidovirales. The characteristic hallmark of nidoviruses is the synthesis of a nested set of subgenomic mRNAs. The unique structural feature of coronaviruses is the "corona" formed by the spike proteins protruding from the surface of the virion. Even though the viral spike protein is the primary receptor protein for all coronaviruses, the corresponding cell surface receptors vary (Li, F., Journal of Virology, 2015, 89, 1954-1964). Delmas *et al.* were the first to characterize porcine aminopeptidase N (ANPEP, APN, or CD13) as a candidate receptor for TGEV (Delmas, B., et al., Nature, 1992, 357 417-420).

SUMMARY

The present teachings provide for and include livestock animals that have gene edits in the ANPEP gene confer resistance to coronaviruses.

In some embodiments, the present teachings can include a livestock animal or progeny thereof with a genome comprising an edited ANPEP gene comprising a premature stop codon. In some configurations, an ANPEP protein encoded by the edited ANPEP gene can show reduced binding to a coronavirus relative to an unedited ANPEP protein. In some configurations, the coronavirus can be transmissible gastroenteritis virus (TGEV), porcine epidemic diarrhea virus (PEDV), or Porcine respiratory coronavirus (PRCV). In some configurations, the coronavirus can be TGEV. In various configurations, the protein produced by the edited ANPEP gene shows no binding to TGEV.

In various configurations, the premature stop codon can be in exon 2, 3, 4, 16, or 18. In various configurations, the premature stop codon can be in exon 2. In various configurations, the premature stop codon can be in exon 3. In various configurations, the premature stop codon can be in exon 4. In various configurations, the premature stop codon can be in exon 16. In various configurations, the premature stop codon can be in exon 18.

In various configurations, the animal can be a pig. In some configurations, the edited ANPEP gene can comprise SEQ ID NO: 252, 253, 254, 255, 256, or 257. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 252. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 253. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 254. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 255. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 256. In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 257.

In some embodiments, a cell isolated from the animal can have an ANPEP gene as described herein. In various embodiments, the present teachings can include an isolated cell line obtained from the livestock animal as described herein.

In some embodiments, the present teachings can include an isolated fibroblast line obtained from the livestock animal as described herein.

In some embodiments, a method of producing a gene edited livestock animal comprising: introducing into an isolated cell: a Cas9 protein or a nucleic acid encoding a Cas9 protein; and a pair of gRNAs that each create a double stranded break in an ANPEP gene such that a premature stop codon is formed when the double stranded breaks are repaired. In some configurations, the method can further comprise producing an animal from the isolated cell. In various configurations, the pair of gRNAs can be SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 102 and 10, SEQ ID NOs: 58 and 52, SEQ ID NOs: 58 and 26, SEQ ID NOs: 58 and 10, SEQ ID NOs: 32 and 26, SEQ ID NOs: 32 and 10, SEQ ID NOs: 120 and 111, SEQ ID NOs: 119 and 113, SEQ ID NOs: 115 and 111, SEQ ID NOs 137 and 130, SEQ ID NOs: 133 and 130, SEQ ID NOs: 137 and 128, SEQ ID NOs: 133 and 128, SEQ ID NOs: 189 and 183, SEQ ID NOs: 231 and 208, SEQ ID NOs: 228 and 213, SEQ ID NOs: 216 and 213, or SEQ ID NOs: 216 and 208. In various configurations, the pair of gRNAs can be SEQ ID NOs: 102 and 52, SEQ ID NOs:

102 and 26, SEQ ID NOs: 58 and 52, SEQ ID Nos: 120 and 111, SEQ ID NOs: 231 and 208, or SEQ ID NOs: 228 and 213. In various configurations, the pair of gRNAs can be SEQ ID NOs: 102 and 52. In various configurations, the pair of gRNAs can be SEQ ID NOs: 102 and 26. In various configurations, the pair of gRNAs can be SEQ ID NOs: 58 and 52. In various configurations, the pair of gRNAs can be SEQ ID Nos: 120 and 111. In various configurations, the pair of gRNAs can be SEQ ID NOs: 231 and 208. In various configurations, the pair of gRNAs can be SEQ ID NOs: 228 and 213.

In various configurations, the edited ANPEP gene can comprise SEQ ID NO: 252, 253, 254, 255, 256, or 257.

In various configurations, the isolated cell can be an oocyte, a zygote, or a fibroblast cell. In various configurations, the isolated cell can be a zygote. In various configurations, the isolated cell can be a fibroblast cell. In various configurations, the isolated cell can be an oocyte. In some configurations, the producing an animal can comprise fertilizing the oocyte. In some configurations, the producing an animal can further comprise implanting the fertilized oocyte into a surrogate mother. In various configurations, the livestock animal can be resistant to coronavirus. In various configurations, the livestock animal can be a pig.

In various embodiments, the present teachings can include at least one gRNA having a sequence selected from the group consisting of SEQ ID NOs: 10, 26, 32, 52, 58, 102, 111, 113, 115, 119, 120, 128, 130, 133, 137, 183, 189, 208, 213, 216, 228, and 231. In some configurations, the at least one gRNA is a pair of gRNAs comprising, consisting of, or selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 102 and 10, SEQ ID NOs: 58 and 52, SEQ ID NOs: 58 and 26, SEQ ID NOs: 58 and 10, SEQ ID NOs: 32 and 26, SEQ ID NOs: 32 and 10, SEQ ID NOs: 120 and 111, SEQ ID NOs: 119 and 113, SEQ ID NOs: 115 and 111, SEQ ID NOs 137 and 130, SEQ ID NOs: 133 and 130, SEQ ID NOs: 137 and 128, SEQ ID NOs 133 and 128, SEQ ID NOs: 189 and 183, SEQ ID NOs: 231 and 208, SEQ ID NOs: 228 and 213, SEQ ID NOs: 216 and 213, and/or SEQ ID NOs: 216 and 208. In various configurations, the at least one gRNA is a pair of gRNAs comprising, consisting of, or selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 58 and 52, SEQ ID Nos: 120 and 111, SEQ ID NOs: 231 and 208, and/or SEQ ID NOs. 228 and 213. In various configurations the at least one gRNA is a pair of gRNAs comprising SEQ

ID NOs: 102 and 52. In various configurations, the at least one gRNA is a pair of gRNAs comprising SEQ ID NOs: 102 and 26. In various configurations, the at least one gRNA is a pair of gRNAs comprising SEQ ID NOs: 58 and 52. In various configurations, the at least one gRNA is a pair of gRNAs comprising SEQ ID Nos: 120 and 111. In various configurations, the at least one gRNA is a pair of gRNAs comprising SEQ ID NOs: 231 and 208. In various configurations, the at least one gRNA is a pair of gRNAs comprising SEQ ID NOs. 228 and 213.

DETAILED DESCRIPTION

The present disclosure relates to livestock animals and progeny thereof comprising at least one edited chromosomal sequence that reduces expression or activity of an ANPEP protein. The disclosure further relates to animal cells comprising at least one edited chromosomal sequence that reduces expression or activity of an ANPEP protein. The animals and cells have chromosomal gene edits (e.g., insertions, deletions, or substitutions) that inactivate or otherwise modulate ANPEP expression or activity. ANPEP plays important roles in coronavirus resistance. The animals and cells can be created using any number of protocols, including those that make use of gene editing.

Definitions

So that the present disclosure may be more readily understood, certain terms are first defined. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which embodiments of the present teachings pertain. Many methods and materials similar, modified, or equivalent to those described herein can be used in the practice of the embodiments of the present teachings without undue experimentation, the preferred materials and methods are described herein. In describing and claiming the embodiments of the present teachings, the following terminology will be used in accordance with the definitions set out below.

It is to be understood that all terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting in any manner or scope. For example, as used in this specification and the appended claims, the singular forms “a”, “an”, and “the” can include plural referents unless the content clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly

indicate otherwise. The word “or” means any one member of a particular list and also includes any combination of members of that list.

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid
5 sequences are written left to right in amino to carboxy orientation, respectively. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

10 Numeric ranges recited within the specification are inclusive of the numbers defining the range and include each integer within the defined range. Throughout this disclosure, various aspects of the present teachings are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the present
15 teachings. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges, fractions, and individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed sub-ranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that
20 range, for example, 1, 2, 3, 4, 5, and 6, and decimals and fractions, for example, 1.2, 3.8, 1½, and 4¾ This applies regardless of the breadth of the range.

The term “about” as used herein, refers to variation in the numerical quantity that can occur, for example, through typical measuring techniques and equipment, with respect to any quantifiable variable, including, but not limited to, mass, volume, time, and
25 temperature. Further, given solid and liquid handling procedures used in the real world, there is certain inadvertent error and variation that is likely through differences in the manufacture, source, or purity of the ingredients used to make the compositions or carry out the methods and the like. The term “about” also encompasses these variations. Whether or not modified by the term “about,” the claims include equivalents to the quantities.

30 Unless otherwise specified, the term “cell” as used herein can refer to an oocyte, a sperm cell, a zygote, an embryonic cell, a germ layer cell, or a differentiated cell. Whole genome edits are typically performed in oocytes, sperm, zygotic, or embryonic cells. For

some disease traits, tissue specific edits may be made in tissue cells or tissue progenitor cells.

The term “Cas” refers to a “CRISPR associated” protein. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cash, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof.

"Cas9" (formerly referred to as Cas5, Csn1, or Csx12) herein refers to a Cas endonuclease of a type II CRISPR system that forms a complex with a crNucleotide and a tracrNucleotide, or with a single guide polynucleotide, for specifically recognizing and cleaving all or part of a DNA target sequence.

"Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

The term "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus.

The term "conservatively edited variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively edited variants" refers to those nucleic acids which encode identical or conservatively edited variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every

position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively edited variation. Every nucleic acid sequence herein that encodes a polypeptide also, by reference 5 to the genetic code, describes every possible silent variation of the nucleic acid.

One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) can be edited to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide of the 10 present teachings is implicit in each described polypeptide sequence and is within the scope of the present disclosure.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds, or deletes a single amino acid or a small percentage of amino 15 acids in the encoded sequence is a "conservatively edited variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations can be made.

Conservatively edited variants typically provide similar biological activity as the 20 unedited polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative 25 substitutions for one another: [1] Alanine (A), Serine (S), Threonine (T); [2] Aspartic acid (D), Glutamic acid (E); [3] Asparagine (N), Glutamine (Q); [4] Arginine (R), Lysine (K); [5] Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and [6] Phenylalanine (F), Tyrosine (Y), Tryptophan (W). *See also*, Creighton (1984) Proteins W. H. Freeman and Company.

30 “Exogenous” refers to a nucleic acid sequence originating outside an organism that has been introduced into the organism. This can refer to sequences which naturally occur

in a sexual compatible species, sequences which are synthetic, or sequences from another species.

As used herein "full-length sequence" in reference to a specified polynucleotide or its encoded protein means having the entire amino acid sequence of a native (nonsynthetic), endogenous, biologically active form of the specified protein. Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots, primer extension, S1 protection, and ribonuclease protection. Comparison to known full-length homologous (orthologous and/or paralogous) sequences can also be used to identify full-length sequences of the present disclosure. Additionally, consensus sequences typically present at the 5' and 3' untranslated regions of mRNA aid in the identification of a polynucleotide as full-length. For example, the consensus sequence ANNNN***AUG***G, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

As used herein, "gene editing," "gene edited" "genetically edited" and "gene editing effectors" refer to the use of naturally occurring or artificially engineered nucleases, also referred to as "molecular scissors." The nucleases create specific double-stranded break (DSBs) at desired locations in the genome, which in some cases harnesses the cell’s endogenous mechanisms to repair the induced break by natural processes of homologous recombination (HR) and/or nonhomologous end-joining (NHEJ). Gene editing effectors include Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), the Clustered Regularly Interspaced Short Palindromic Repeats/CAS (CRISPR/Cas) system, and meganucleases re-engineered as homing endonucleases. The terms also include the use of genetic editing procedures and techniques, including, for example, where the change is relatively small and/or does not introduce DNA from a foreign species.

As used herein, the term “gene edit”, “gene edited”, or “genetically edited” refers to an organism where human intervention, such as but without limitation by using a gene editing effector, has created a genetic difference in its genome when compared to a wild type genome of the same organism. These differences can include but are not limited to nucleotide substitutions, excision of a start codon, or small deletions that do not introduce

frame shift mutations into the genome but may excise an exon or form a premature stop codon when the ends of the deletion are ligated together. A gene edit does not introduce DNA from another species into an organism.

A “gene edited animal” refers to an animal with one or more cells comprising a gene edit.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially changed from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially changed from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially changed from its original form by deliberate human intervention.

As used herein “homing DNA technology” or “homing technology” covers any mechanisms that allow a specified molecule to be targeted to a specified DNA sequence including Zinc Finger (ZF) proteins, Transcription Activator-Like Effectors (TALEs) meganucleases, and CRISPR systems (*e.g.*, CRISPR/Cas9 systems).

The terms "increased resistance" and "reduced susceptibility" herein mean, but are not limited to, a statistically significant reduction of the incidence and/or severity of clinical signs or clinical symptoms which are associated with infection by pathogen. For example, "increased resistance" or "reduced susceptibility" can refer to a statistically significant reduction of the incidence and/or severity of clinical signs or clinical symptoms which are associated with infection by a coronavirus in an animal comprising an edited chromosomal sequence as compared to a control animal having an unedited chromosomal sequence. The term "statistically significant reduction of clinical symptoms" means, but is not limited to, the frequency in the incidence of at least one clinical symptom in the edited group of subjects is at least 10%, preferably at least 20%, more preferably at least 30%, even more preferably at least 50%, and even more preferably at least 70% lower than in the non-edited control group after the challenge with the infectious agent.

The term “livestock animal” includes any animals traditionally raised in livestock farming, for example an ungulate (*e.g.*, an artiodactyl), an avian animal (*e.g.*, chickens, turkeys, ducks, geese, guinea fowl, or squabs), an equine animal (*e.g.*, horses or donkeys).

Ungulates include, but are not limited to porcine animals (*e.g.*, pigs), bovine animals (*e.g.*, beef or dairy cattle, buffalo), ovine animals, caprine animals, camels, llamas, alpacas, and deer. The term does not include rats, mice, or other rodents.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single-or double-stranded form, and unless otherwise limited, encompasses conservatively edited variants and known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (*e. g.*, peptide nucleic acids).

The term “mutation” includes alterations in the nucleotide sequence of a polynucleotide, such as for example a gene or coding DNA sequence (CDS), compared to the wild-type sequence. The term includes, without limitation, substitutions, insertions, frameshifts, deletions, inversions, translocations, duplications, splice-donor site mutations, point-mutations and the like.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also may apply to conservatively edited variants and to amino acid polymers in which one or more amino acid residues is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, the protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide" and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitization, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, the present teachings contemplate the use of both the methionine-containing and the methionine-less amino terminal variants of the protein of the present disclosure.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as testes, ovaries, or placenta. Such promoters are referred to as "tissue-preferred". Promoters which initiate transcription only in certain tissue are referred to as "tissue-specific". A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, germ cells in testes or ovaries. An "inducible" or "repressible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may affect transcription by inducible promoters include stress, and temperature. Tissue-specific, tissue-preferred, cell-type specific and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

The terms "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

"Resistance" of an animal to a disease is a characteristic of an animal, wherein the animal avoids the disease symptoms that are the outcome of animal-pathogen interactions, such as interactions between a porcine animal and coronavirus. That is, pathogens are prevented from causing animal diseases and the associated disease symptoms, or alternatively, a reduction of the incidence and/or severity of clinical signs or reduction of clinical symptoms. One of skill in the art will appreciate that the methods disclosed herein can be used with other compositions and methods available in the art for protecting animals from pathogens.

A "TALE DNA binding domain" or "TALE" is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains are involved in binding of the TALE to its cognate target DNA sequence. A single "repeat unit" (also referred to as a "repeat") is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein.

Zinc finger and TALE binding domains can be "engineered" to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of naturally occurring zinc finger or TALE proteins. Therefore, engineered DNA binding proteins (zinc fingers or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP and/or TALE designs and binding data. See, for example, U.S. Pat. Nos. 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496 and U.S. Publication No. 20110301073.

A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

A "selected" zinc finger protein or TALE is a protein not found in nature whose production results primarily from an empirical process such as phage display, interaction trap or hybrid selection. See e.g., U.S. Pat. No. 5,789,538; U.S. Pat. No. 5,925,523; U.S. Pat. No. 6,007,988; U.S. Pat. No. 6,013,453; U.S. Pat. No. 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197, WO 02/099084 and U.S. Publication No. 20110301073.

"Wild type" means those animals and blastocysts, embryos or cells derived therefrom, which have not been gene edited and are usually inbred and outbred strains developed from naturally occurring strains.

The following terms are used to describe the sequence relationships between a polynucleotide/polypeptide of the present teachings with a reference polynucleotide/polypeptide: (a)"reference sequence", (b)"comparison window", (c) "sequence identity", and (d)"percentage of sequence identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison with a polynucleotide/polypeptide of the present teachings. A

reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" includes reference to a contiguous and specified segment of a polynucleotide/polypeptide sequence, wherein the polynucleotide/polypeptide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide/polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides/amino acids residues in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide/polypeptide sequence, a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith, T.F. and Waterman, M.S., Adv. Appl. Math., 1981, 2, 482-489; by the homology alignment algorithm of Needleman, S.B. and Wunsch, C.D. (J. Mol. Biol., 1970, 48, 443-453); by the search for similarity method of Pearson, W.R. and Lipman, D.J. (Proc. Natl. Acad. Sci., 1988, 85, 2444-2448); and by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA, and TFASTA, and related programs in the GCG Wisconsin Genetics Software Package, Version 10 (available from Accelrys Inc., 9685 Scranton Road, San Diego, California, USA). The CLUSTAL program is well described in Higgins, D.G. and Sharp, P.M.,Gene, 1988, 73, 237-244; Higgins, D.G. and Sharp, P.M., CABIOS, 1989, 5, 151-153; Corpet, F., Nucleic Acids Research, 1988, 16, 10881-10890; Huang, X., *et al.*, Computer Applications in the Biosciences, 1992, 8, 155-165, and Pearson, W.R, Methods in Molecular Biology, 1994, 24, 307-331.

The BLAST family of programs that can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for

protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. *See*, Current Protocols in Molecular Biology, Chapter 19, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); Altschul, S.F., *et al.*, J. Mol. Biol., 1990, 215: 403-410; and Altschul, S.F. *et al.*, Nucleic Acids Res., 1997, 25, 3389-3402. Software for performing BLAST analyses is publicly available, for example through the National Center for Biotechnology Information (ncbi.nlm.nih.gov). This algorithm has been thoroughly described in a number of publications. *See, e.g.*, Altschul, S.F., *et al.*, Nucleic Acids Res., 1997, 25, 3389-3402; National Center for Biotechnology Information, The NCBI Handbook [Internet], Chapter 16: The BLAST Sequence Analysis Tool (McEntyre J, Ostell J, eds., 2002), available at www(dot)ncbi(dot)nlm(dot)nih(dot)gov/books/NBK21097/pdf/ch16.pdf. The BLASTP program for amino acid sequences has also been thoroughly described (see Henikoff, S., and Henikoff, J.G., Proc. Natl. Acad. Sci. USA, 1992, 89, 10915-10919).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin, S & Altschul, S.F., Proc. Nat'l. Acad. Sci. USA, 1993, 90, 5873-5877). A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten, J.C. and Federhen, S., Comput. Chem., 1993, 17, 149-163) and XNU (Claverie, J.M. and States, D.J., Comput. Chem., 1993, 17, 191-201) low-complexity filters can be employed alone or in combination.

Unless otherwise stated, nucleotide and protein identity/similarity values provided herein are calculated using GAP (GCG Version 10) under default values. GAP (Global Alignment Program) can also be used to compare a polynucleotide or polypeptide of the present teachings with a reference sequence. GAP uses the algorithm of Needleman, S.B. and Wunsch, C.D. (J. Mol. Biol., 1970, 48: 443-453), to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP represents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that

actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (*see* Henikoff, S. & Henikoff, J.G., Proc. Natl. Acad. Sci. USA, 1989, 89, 10915-10919).

Multiple alignment of the sequences can be performed using the CLUSTAL method of alignment (Higgins, D.G. and Sharp, P.M., CABIOS, 1989, 5, 151-153) with the default parameters (GAPPENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the CLUSTAL method include KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.* charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions may be calculated according to the algorithm of Meyers. E.W. and Miller, W., Computer Applic. Biol. Sci., 1988, 4: 11-17, for example as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions

or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, 5 dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The porcine amino peptidase N (ANPEP) has been reported to be involved in transmissible gastroenteritis virus (TGEV) infection. Porcine ANPEP is a 963 amino acid, 10 type II membrane metallopeptidase, that is responsible for removing N-terminal amino acids from protein substrates during digestion.

Gene description

Sus scrofa ANPEP is a membrane alanyl aminopeptidase described at ENSEMBL 15 (useast.ensembl.org/) accession number ENSSSCG00000001849, located on the reverse strand of the 7th chromosome. ENSEMBL has six transcripts. The gene was listed in ENSSSCG00000001849.5 (SEQ ID NO: 258) and the full transcript was designated as ANPEP-205, and the locations of the exons on the 7th chromosome are listed below in Table 1 (these data are from Ensembl release 109).

20

Table 1 ANPEP Gene Structure

Exon Number	LOCATION
1	55,373,881-55,373,762
2	55,366,170-55,365,354
3	55,364,844-55,364,702
4	55,364,574-55,364,435
5	55,364,011-55,363,885

6	55,363,790-55,363,636
7	55,363,406-55,363,293
8	55,363,195-55,363,052
9	55,362,874-55,362,809
10	55,362,685-55,362,620
11	55,360,729-55,360,554
12	55,360,314-55,360,554
13	55,360,072-55,359,939
14	55,359,836-55,359,781
15	55,358,851-55,358,704
16	55,354,492-55,354,401
17	55,353,921-55,353,811
18	55,353,683-55,353,516
19	55,352,790-55,352,650
20	55,351,877-55,351,796

21	55,351,673- 55,351,083

The present inventors designed edits to introduce a premature stop codon in order to make a truncated protein. gRNAs were engineered to create double stranded breaks in the gene that, when repaired via non-homologous end joining (NHEJ), create a stop codon. The present inventors developed two potential strategies: one strategy called for introduction of an early stop codon in Exon 2, 3, or 4. The gRNAs listed in SEQ ID NOs: 1-106 and 147-175 are directed to Exon 2. The gRNAs listed in SEQ ID NOs: 107-126 and 176 are directed to Exon 3. The gRNAs listed in SEQ ID NOs: 127-147 and 177-182 are directed to Exon 4. The second strategy was to introduce an exogenous stop codon late in the gene—in exon 15, 16, 17, or 18. The gRNAs listed in SEQ ID NOs: 183-190 and 240 are directed to Exon 16. The gRNAS listed in SEQ ID NOs: 191-201 and 241 are directed to Exon 17. The gRNAs listed in SEQ ID NOs: 202-239 and 242-251 are directed to Exon 18. gRNAs listed in SEQ ID NOs: 1-146 and 183-239 were designed for use with spCas9. SEQ ID NOs: 147-182 and 240-251 were designed for use with SthC3Cas9.

Gene Editing

The present disclosure provides a livestock animal or animal cell, including sperm or egg cells, with improved resistance to coronavirus, specifically transmissible gastroenteritis virus (TEGV). The livestock animals or cells comprise altered expression or activity of an ANPEP protein. The livestock animals or cells can comprise at least one edited chromosomal sequence that reduces expression or activity of the ANPEP protein. The chromosomal sequence may be (1) inactivated, (2) edited, or (3) comprise an integrated sequence. An inactivated chromosomal sequence is altered such that ANPEP protein function is impaired, reduced or eliminated. As used herein, reduced ANPEP protein function or activity (*e.g.*, coronavirus binding activity) refers to a reduction in protein function or activity relative to the function of a wild type ANPEP protein. Thus, a gene edited animal comprising an inactivated chromosomal sequence may be termed a “knock-out”. Similarly, a genetically edited animal comprising an integrated sequence may be termed a “knock in”. Furthermore, a gene edited animal comprising an edited chromosomal sequence may comprise a targeted point edit(s) or other edit such that an altered protein product is produced. Briefly, the process can comprise using a CRISPR system (*e.g.*, a CRISPR/Cas9

system) to edit the genomic sequence. To use Cas9 to edit genomic sequences, the protein can be delivered directly to a cell. Alternatively, an mRNA that encodes Cas9 can be delivered to a cell, or a gene that provides for expression of an mRNA that encodes Cas9 can be delivered to a cell. In addition, either target specific crRNA and a tracrRNA can be delivered directly to a cell or target specific sgRNA(s) can be delivered to a cell (these RNAs can alternatively be produced by a gene constructed to express these RNAs). The process of editing chromosomal sequences using a CRISPR system is rapid, precise, and highly efficient.

In some embodiments, an ANPEP locus is used as a target site for the site-specific editing. This can include insertion of an exogenous nucleic acid (e.g., a nucleic acid comprising a nucleotide sequence encoding a polypeptide of interest) or deletions of nucleic acids from the locus. In particular embodiments, insertions and/or deletions result in an edited locus. For example, integration of the premature nucleic acid and/or deletion of part of the genomic nucleic acid may edit the locus so as to produce a disrupted (*i.e.*, inactivated) ANPEP gene. The ANPEP protein encoded by the edited nucleic acid can have reduced coronavirus binding relative to an unedited ANPEP protein. In some configurations, the edited ANPEP protein may not bind to a coronavirus.

Any of the animals or cells can be an animal or cell that has been gene edited using a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system. The CRISPR/Cas system can suitably comprise any of the guide RNAs (gRNAs) described herein.

For any of the animals, progeny, or cells, the edited chromosomal sequence can be an edited chromosomal sequence that has been produced via homology directed repair (HDR). Alternatively, the edited chromosomal sequence can be an edited chromosomal sequence that has been produced via non-homologous end-joining (NHEJ).

The edited chromosomal sequence reduces the susceptibility of the animal, progeny, or cell to infection by a coronavirus, as compared to the susceptibility of a livestock animal, progeny, or cell that does not comprise the edited chromosomal sequence. In addition, in porcine animals having the edited chromosomal sequence, coronavirus nucleic acid is detected at lower titers in the nasal secretions, feces, or serum relative to a control. Alternatively or in addition, coronavirus antigen is detected at lower concentrations in the tissues of the animal (*e.g.*, in lung tissue) relative to a control. The

edited chromosomal sequence preferably substantially eliminates susceptibility of the animal, progeny, or cell to the coronavirus. The edited chromosomal sequence more preferably completely eliminates susceptibility of the animal, progeny, or cell to the coronavirus, such that animals do not show any clinical signs of disease following
5 exposure to the coronavirus. For example, porcine animals having the edited chromosomal sequence do not show any clinical signs of coronavirus (*e.g.*, fever, lethargy, anorexia, weight loss, nasal and ocular discharge, cough, sneezing, conjunctivitis, diarrhea, and/or breathing difficulties) following exposure to the coronavirus. In addition, in porcine animals having the edited chromosomal sequence, coronavirus nucleic acid cannot be
10 detected in the nasal secretions, feces, or serum, coronavirus antigen cannot be detected in the tissues of the animal (*e.g.*, in lung tissue), and serum is negative for coronavirus-specific antibody. Similarly, cells having the edited chromosomal sequence that are exposed to the pathogen do not become infected with the pathogen. Exemplary coronaviruses include TEGV, PEDV, and PRCV.
15 The edited chromosomal sequence can comprise an insertion, a deletion, a substitution, or a combination of any thereof. The insertion, the deletion, the substitution, or the combination of any thereof can result in a miscoding in the allele of the gene encoding the ANPEP protein. Where the insertion, the deletion, the substitution, or the combination of any thereof results in a miscoding in the allele of the gene encoding the
20 ANPEP protein, the miscoding can result in a premature stop codon in the allele of the gene encoding the ANPEP protein. In some embodiments, editing strategies are designed to introduce premature stop codons in conserved exonic sequences of ANPEP. In certain embodiments, premature stop codons are introduced either early in the coding sequence or in coding regions containing protease active site residues. In some configurations, these
25 editing strategies include the use of two guides to direct the creation of double stranded breaks and then, through non-homologous end joining (NHEJ) create a premature stop codon when the cut ends are ligated. The present inventors have employed this strategy to create premature stop codons early in the ANPEP gene, such as in Exons 2, 3, and 4. Pairs of guides that can be used to introduce a premature stop codon in Exon 2 include SEQ ID
30 NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 102 and 10, SEQ ID NOs: 58 and 52, SEQ ID NOs: 58 and 26, SEQ ID NOs: 58 and 10, SEQ ID NOs: 32 and 26, and SEQ ID NOs: 32 and 10. Pairs of guides that can be used to introduce a premature stop

codon in Exon 3 include SEQ ID NOs: 120 and 111, SEQ ID NOs: 119 and 113, and SEQ ID NOs: 115 and 111. Pairs of guides that can be used to introduce a premature stop codon in Exon 4 include SEQ ID NOs 137 and 130, SEQ ID NOs: 133 and 130, SEQ ID NOs: 137 and 128, and SEQ ID NOs 133 and 128. As an alternative strategy, pairs of guides were designed to introduce a premature stop codon within Exon 15 or 17. An exemplary pair of guides that can be used to introduce a premature stop codon in Exon 15 is SEQ ID NOs: 189 and 183. Pairs of guides that can be used to introduce a premature stop codon in Exon 17 include SEQ ID NOs: 231 and 208, SEQ ID NOs: 228 and 213, SEQ ID NOs: 216 and 213, and SEQ ID NOs: 216 and 208. The edited sequences produced by select guide pairs are shown in Table 2. (The sequences recite 50 base pairs on either side of the join.)

Table 2 Select Guide Pairs and Edited Sequences

Exon	Guide 1 SEQ ID NO:	Guide 2 SEQ ID NO:	Edited Genomic Sequence SEQ ID NO:
2	102	52	252
2	102	26	253
2	58	52	254
3	120	111	255
18	231	208	256
18	228	213	257

In any of the animals, progeny, or cells described herein, the edited chromosomal sequence preferably causes ANPEP protein production or activity to be reduced, as compared to ANPEP protein production or activity in an animal, progeny, or cell that lacks the edited chromosomal sequence. Specifically, the edited sequence encodes an ANPEP protein that has reduced or no binding to coronaviruses or other pathogens.

The edited chromosomal sequence can result in production of substantially no functional ANPEP protein by the animal, progeny, or cell. By "substantially no functional ANPEP protein," it is meant that the level of ANPEP protein in the animal, progeny, or cell is undetectable, or if detectable, is at least about 90% lower, at least about 95% lower, at least about 98%, lower, and even more preferably at least about 99% lower than the

level observed in an animal, progeny, or cell that does not comprise the edited chromosomal sequences. For any of the animals, progeny, or cells described herein, the animal, progeny, or cell does not produce ANPEP protein.

Alternatively, or in addition, a gene may be so edited as to delete a partial sequence in protein that responsible for an undesirable trait. For example, but without limitation, a domain of a protein may be removed that binds to a pathogen. This allows the editing of genes that are essential for host survival while still rendering the host resistant to the disease.

The ANPEP gene in the animal, progeny, or cell can comprise any combination of any of the edited chromosomal sequences described herein.

Guide RNAs

Guide RNAs (gRNAs) are provided. The gRNAs have a nucleic acid sequence that is complementary to a sequence of a gene encoding an ANPEP protein and can be used to introduce a chromosomal edit into a gene encoding an ANPEP protein.

Illustrative gRNA sequences complementary to a sequence of a gene encoding an ANPEP protein are provided in SEQ ID NOs: 1-251. Although the sequences are listed with DNA nucleotides, a person of ordinary skill would understand that the sequences are in fact RNA sequences and would be readily able to convert the DNA sequences into RNA sequences. In some embodiments, the gRNA sequences target regions early in the ANPEP coding sequence or in coding regions containing protease active site residues.

The gRNA can comprise a nucleotide sequence comprising one or more of SEQ ID NOs: 1-251. The gRNA can have a length of 100 nucleotides or fewer, 90 nucleotides or fewer, 80 nucleotides or fewer, 70 nucleotides or fewer, 60 nucleotides or fewer, 50 nucleotides or fewer, 40 nucleotides or fewer, 30 nucleotides or fewer, or 20 nucleotides or fewer. For example, the gRNA can have a length of 20 nucleotides.

In certain embodiments, guide RNAs within ANPEP are selected for their ability to generate in-frame stop codons when paired.

DNA-Binding Polypeptides

In some embodiments, site-specific integration may be accomplished by utilizing factors that are capable of recognizing and binding to particular nucleotide sequences, for example, in the genome of a host organism. For instance, many proteins comprise

polypeptide domains that are capable of recognizing and binding to DNA in a site-specific manner. A DNA sequence that is recognized by a DNA-binding polypeptide may be referred to as a "target" sequence. Polypeptide domains that are capable of recognizing and binding to DNA in a site-specific manner generally fold correctly and function independently to bind DNA in a site-specific manner, even when expressed in a polypeptide other than the protein from which the domain was originally isolated. Similarly, target sequences for recognition and binding by DNA-binding polypeptides are generally able to be recognized and bound by such polypeptides, even when present in large DNA structures (*e.g.*, a chromosome), particularly when the site where the target sequence is located is one known to be accessible to soluble cellular proteins (*e.g.*, a gene).

While DNA-binding polypeptides identified from proteins that exist in nature typically bind to a discrete nucleotide sequence or motif (e.g., a consensus recognition sequence), methods exist and are known in the art for modifying many such DNA-binding polypeptides to recognize a different nucleotide sequence or motif. DNA-binding polypeptides include, for example and without limitation: zinc finger DNA-binding domains; leucine zippers; UPA DNA-binding domains; GAL4; TAL; LexA; a Tet repressor; LacR; and a steroid hormone receptor.

In some embodiments, a DNA-binding polypeptide is a zinc finger. Individual zinc finger motifs can be designed to target and bind specifically to any of a large range of DNA sites. Canonical Cys_2His_2 (as well as non-canonical Cys_3His) zinc finger polypeptides bind DNA by inserting an α-helix into the major groove of the target DNA double helix. Recognition of DNA by a zinc finger is modular; each finger contacts primarily three consecutive base pairs in the target, and a few key residues in the polypeptide mediate recognition. By including multiple zinc finger DNA-binding domains in a targeting endonuclease, the DNA-binding specificity of the targeting endonuclease may be further increased (and hence the specificity of any gene regulatory effects conferred thereby may also be increased). See, *e.g.*, Urnov, F.D., *et al.*, Nature, 2005, 435, 646-651. Thus, one or more zinc finger DNA-binding polypeptides may be engineered and utilized such that a targeting endonuclease introduced into a host cell interacts with a DNA sequence that is unique within the genome of the host cell.

Preferably, the zinc finger protein is non-naturally occurring in that it is engineered to bind to a target site of choice. See, for example, Beerli, R.R., *et al.*, Nature Biotechnol.,

2002, 20, 135-141; Pabo, C.O., *et al.*, Ann. Rev. Biochem., 2001, 70, 313-340; Isalan, M., *et al.*, Nature Biotechnol., 2001, 19, 656-660; Segal, D.J., *et al.*, Curr. Opin. Biotechnol., 2001, 12, 632-637; Choo, Y., *et al.*, 2000, Curr. Opin. Struct. Biol., 10, 411-416; U.S. Pat. Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136;
5 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

An engineered zinc finger binding domain can have a novel binding specificity, compared to a naturally occurring zinc finger protein. Engineering methods include, but
10 are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, U.S. Pat. Nos. 6,453,242
15 and 6,534,261, incorporated by reference herein in their entireties.

Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. Pat. Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding
20 specificity for zinc finger binding domains has been described, for example, in WO 02/077227.

Selection of target sites; ZFPs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Pat. Nos. 6,140,0815; 789,538; 6,453,242; 6,534,261;
25 5,925,523; 6,007,988; 6,013,453; 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker
30 sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or

more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

In some examples, a DNA-binding polypeptide is a DNA-binding domain from GAL4. GAL4 is a modular transactivator in *Saccharomyces cerevisiae*, but it also operates as a transactivator in many other organisms. See, *e.g.*, Sadowski, I., *et al.*, Nature, 1988, 335, 563-564. In this regulatory system, the expression of genes encoding enzymes of the galactose metabolic pathway in S. cerevisiae is stringently regulated by the available carbon source. Johnston, M., Microbiol. Rev., 1987, 51, 458-476. Transcriptional control of these metabolic enzymes is mediated by the interaction between the positive regulatory protein, GAL4, and a 17 bp symmetrical DNA sequence to which GAL4 specifically binds (the UAS).

Native GAL4 consists of 881 amino acid residues, with a molecular weight of 99 kDa. GAL4 comprises functionally autonomous domains, the combined activities of which account for activity of GAL4 *in vivo* (Ma, J. and Ptashne, M., Cell, 1987, 48, 847-853; Brent, R. and Ptashne, M., Cell, 1985 43, 729-736). The N-terminal 65 amino acids of GAL4 comprise the GAL4 DNA-binding domain (Keegan, L., *et al.*, Science, 1986, 231, 699-704; Johnston, M., Nature, 1987, 328, 353-355). Sequence-specific binding requires the presence of a divalent cation coordinated by 6 Cys residues present in the DNA binding domain. The coordinated cation-containing domain interacts with and recognizes a conserved CCG triplet at each end of the 17 bp UAS via direct contacts with the major groove of the DNA helix (Marmorstein, M., *et al.*, Nature, 1992, 356, 408-414). The DNA-binding function of the protein positions C-terminal transcriptional activating domains in the vicinity of the promoter, such that the activating domains can direct transcription.

Additional DNA-binding polypeptides that may be utilized in certain embodiments include, for example and without limitation, a binding sequence from a AVRBS3-inducible gene; a consensus binding sequence from a AVRBS3-inducible gene or synthetic binding sequence engineered therefrom (*e.g.*, UPA DNA-binding domain); TAL; LexA (see, *e.g.*, Brent, R. and Ptashne, M., Cell, 1985 43, 729-736); LacR (see, *e.g.*, Labow, M.A., *et al.*, Mol. Cell. Biol., 1990, 10, 3343-3356; Baim, *S.B.*, *et al.*, Proc. Natl. Acad. Sci. USA 1991, 88, 5072-5076); a steroid hormone receptor (Elliston, J.F., *et al.*, J. Biol. Chem. 1990, 265, 11517-11521); the Tet repressor (U.S. Pat. No. 6,271,341) and a mutated Tet repressor that binds to a tet operator sequence in the presence, but not the absence, of tetracycline (Tc);

the DNA-binding domain of NF-κB; and components of the regulatory system described in Wang, Y., *et al.*, Proc. Natl. Acad. Sci. USA, 1994, 91, 8180-8184, which utilizes a fusion of GAL4, a hormone receptor, and VP16.

In certain embodiments, the DNA-binding domain of one or more of the nucleases
5 used in the methods and compositions described herein comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding domain. See, *e.g.*, U.S. Patent Publication No. 20110301073, incorporated by reference in its entirety herein.

In other embodiments, the nuclease comprises a CRISPR system (*e.g.*, CRISPR/Cas9 system). The CRISPR (clustered regularly interspaced short palindromic
10 repeats) locus, which encodes RNA components of the system, and the Cas (CRISPR-associated) locus, which encodes proteins (Jansen, R., *et al.*, Mol. Microbiol., 2002, 43, 1565-1575; Makarova, K.,S., *et al.*, Nucleic Acids Res. 2002, 30: 482-496; Makarova, K.S., *et al.*, Biol. Direct, 2006, 1, 7; Haft, D.H., *et al.*, PLoS Comput. Biol., 2005, 1, e60) make up the gene sequences of the CRISPR/Cas nuclease system. CRISPR loci in
15 microbial hosts contain a combination of Cas genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

The Type II CRISPR system is one of the most well characterized systems and carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus.
20 Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target
25 recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of exogenous DNA sequences into the CRISPR array to prevent future attacks, in a process called `adaptation`, (ii) expression of the relevant proteins, as well as expression and processing of the array, followed by (iii) RNA-mediated interference with
30 the exogenous nucleic acid. Thus, in the bacterial cell, several Cas proteins are involved with the natural function of the CRISPR/Cas system and serve roles in functions such as insertion of the foreign DNA etc.

In certain embodiments, Cas protein may be a "functional derivative" of a naturally occurring Cas protein. A "functional derivative" of a native sequence polypeptide is a compound having a qualitative biological property in common with a native sequence polypeptide. "Functional derivatives" include, but are not limited to, fragments of a native sequence and derivatives of a native sequence polypeptide and its fragments, provided that they have a biological activity in common with a corresponding native sequence polypeptide. A biological activity contemplated herein is the ability of the functional derivative to hydrolyze a DNA substrate into fragments. The term "derivative" encompasses both amino acid sequence variants of polypeptide, covalent modifications, and fusions thereof. Suitable derivatives of a Cas polypeptide or a fragment thereof include but are not limited to mutants, fusions, covalent modifications of Cas protein or a fragment thereof. Cas protein, which includes Cas protein or a fragment thereof, as well as derivatives of Cas protein or a fragment thereof, may be obtainable from a cell or synthesized chemically or by a combination of these two procedures. The cell may be a cell that naturally produces Cas protein, or a cell that naturally produces Cas protein and is genetically engineered to produce the endogenous Cas protein at a higher expression level or to produce a Cas protein from an exogenously introduced nucleic acid, which nucleic acid encodes a Cas that is same or different from the endogenous Cas. In some cases, the cell does not naturally produce Cas protein and is genetically engineered to produce a Cas protein.

Cas9 protein comprises a RuvC nuclease domain and an HNH (H-N-H) nuclease domain, each of which can cleave a single DNA strand at a target sequence (the concerted action of both domains leads to DNA double-strand cleavage, whereas activity of one domain leads to a nick). In general, the RuvC domain comprises subdomains I, II and III, where domain I is located near the N-terminus of Cas9 and subdomains II and III are located in the middle of the protein, flanking the HNH domain (Hsu, P.D., *et al.*, Cell, 2014, 157, 1262-1278). A type II CRISPR system includes a DNA cleavage system utilizing a Cas9 endonuclease in complex with at least one polynucleotide component. For example, a Cas9 can be in complex with a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). In another example, a Cas9 can be in complex with a single guide RNA that combines a crRNA and a tracrRNA into a single molecule. The amino acid sequence of a Cas9 protein described herein, as well as certain other Cas proteins herein,

may be derived from a Streptococcus (*e.g.*, *S. pyogenes*, *S. pneumoniae*, *S. thermophilus*, *S. agalactiae*, *S. parasanguinis*, *S. oralis*, *S. salivarius*, *S. macacae*, *S. dysgalactiae*, *S. anginosus*, *S. constellatus*, *S.pseudoporcinus*, *S. mutans*), Listeria (*e.g.*, *L. innocua*), Spiroplasma (*e.g.*, *S. apis*, *S. syrphidicola*), Peptostreptococcaceae, Atopobium,
5 Porphyromonas (*e.g.*, *P. catoniae*), Prevotella (*e.g.*, *P. intermedia*), Veillonella, Treponema (*e.g.*, *T socranskii*, *T denticola*), Capnocytophaga, Finegoldia (*e.g.*, *F. magna*), Coriobacteriaceae (*e.g.*, *C. bacterium*), Olsenella (*e.g.*, *O. profusa*), Haemophilus (*e.g.*, *H. sputorum, H. pittmaniae*), Pasteurella (*e.g.*, *P. bettyae*), Olivibacter (*e.g.*, *O. sitiensis*), Epilithonimonas (*e.g.*, *E. tenax*), Mesonia (*e.g.*, *M. mobilis*), Lactobacillus (*e.g.*, *L.*
10 *plantarum*), Bacillus (*e.g.*, *B. cereus*), Aquimarina (*e.g.*, *A. muelleri*), Chryseobacterium (e.g., *C. palustre*), Bacteroides (*e.g.*, *B. graminisolvens*), Neisseria (*e.g.*, *N. meningitidis*), Francisella (*e.g.*, *F. novicida*), or Flavobacterium (*e.g.* , *F. frigidarium*, *F. soli*) species, for example. As another example, a Cas9 protein can be any of the Cas9 proteins disclosed in Chylinski *et al.* (Chylinski, K., *et al.*, RNA Biology, 2013, 10, 726-737 and US patent
15 application 62/162377, filed May 15, 2015) which is incorporated herein by reference.

Accordingly, the sequence of a Cas9 protein herein can comprise, for example, any of the Cas9 amino acid sequences disclosed in GenBank Accession Nos. G3ECR1 (S. thermophilus), WP_026709422, WP_027202655, WP_027318179, WP_027347504, WP_027376815, WP_027414302, WP_027821588, WP_027886314, WP_027963583,
20 WP_028123848, WP_028298935, Q03JI6 (*S. thermophilus*), EGP66723, EGS38969, EGV05092, EHI65578 (*S. pseudoporcinus*), EIC75614 (*S. oralis*), EID22027 (*S. constellatus*), EIJ69711, EJP22331 (*S. oralis*), EJP26004 (*S. anginosus*), EJP30321, EPZ44001 (*S. pyogenes*), EPZ46028 (*S. pyogenes*), EQL78043 (*S. pyogenes*), EQL78548 (*S. pyogenes*), ERL10511, ERL12345, ERL19088 (*S. pyogenes*), ESA57807
25 (*S. pyogenes*), ESA59254 (*S. pyogenes*), ESU85303 (*S. pyogenes*), ETS96804, UC75522, EGR87316 (*S. dysgalactiae*), EGS33732, EGV01468 (*S. oralis*), EHJ52063 (*S. macacae*), EID26207 (*S. oralis*), EID33364, EIG27013 (*S. parasanguinis*), EJF37476, EJO19166 (Streptococcus sp. BS35b), EJU16049, EJU32481, YP_006298249, ERF61304, ERK04546, ETJ95568 (*S. agalactiae*), TS89875, ETS90967 (*Streptococcus* sp. SR4),
30 ETS92439, EUB27844, (*Streptococcus* sp. BS21), AFJ08616, EUC82735 (*Streptococcus* sp. CM6), EWC92088, EWC94390, EJP25691, YP_008027038, YP_008868573, AGM26527, AHK22391, AHB36273, Q927P4, G3ECR1, or Q99ZW2 (*S. pyogenes*),

which are incorporated by reference. A variant of any of these Cas9 protein sequences may be used, but should have specific binding activity, and optionally endonucleolytic activity, toward DNA when associated with an RNA component herein. Such a variant Cas9 protein should have specific binding activity, and optionally cleavage or nicking activity, toward DNA when associated with an RNA component herein. A Cas protein herein such as a Cas9 can comprise a heterologous nuclear localization sequence (NLS). A heterologous NLS amino acid sequence herein may be of sufficient strength to drive accumulation of a Cas protein in a detectable amount in the nucleus of a yeast cell herein, for example. An NLS may comprise one (monopartite) or more (*e.g.*, bipartite) short sequences (*e.g.*, 2 to 20 residues) of basic, positively charged residues *(e.g.*, lysine and/or arginine), and can be located anywhere in a Cas amino acid sequence but such that it is exposed on the protein surface. An NLS may be operably linked to the N-terminus or C-terminus of a Cas protein herein, for example. Two or more NLS sequences can be linked to a Cas protein, for example, such as on both the N- and C-termini of a Cas protein. Non-limiting examples of suitable NLS sequences herein include those disclosed in U.S. Patent No. 7309576, which is incorporated herein by reference.

The Cas endonuclease can comprise a modified form of the Cas9 polypeptide. The modified form of the Cas9 polypeptide can include an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the naturally occurring nuclease activity of the Cas9 protein. For example, in some instances, the modified form of the Cas9 protein has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9 polypeptide (US patent application US20140068797 A1, published on March 6, 2014). In some cases, the modified form of the Cas9 polypeptide has no substantial nuclease activity and is referred to as catalytically "inactivated Cas9" or "deactivated cas9 (dCas9)." Catalytically inactivated Cas9 variants include Cas9 variants that contain mutations in the HNH and RuvC nuclease domains. These catalytically inactivated Cas9 variants are capable of interacting with sgRNA and binding to the target site in vivo but cannot cleave either strand of the target DNA.

A catalytically inactive Cas9 can be fused to a heterologous sequence (US patent application US20140068797 A1, published on March 6, 2014). Suitable fusion partners include, but are not limited to, a polypeptide that provides an activity that indirectly

increases transcription by acting directly on the target DNA or on a polypeptide (e.g., a histone or other DNA-binding protein) associated with the target DNA. Additional suitable fusion partners include, but are not limited to, a polypeptide that provides for methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or demyristoylation activity. Further suitable fusion partners include, but are not limited to, a polypeptide that directly provides for increased transcription of the target nucleic acid (*e.g.*, a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, a small molecule/drug-responsive transcription regulator, etc.). A catalytically inactive Cas9 can also be fused to a Fok I nuclease to generate double-strand breaks (Guilinger, J.P., *et al.*, Nature Biotechnology, 2014, 32, 577-582).

In particular embodiments, a DNA-binding polypeptide specifically recognizes and binds to a target nucleotide sequence comprised within a genomic nucleic acid of a host organism. Any number of discrete instances of the target nucleotide sequence may be found in the host genome in some examples. The target nucleotide sequence may be rare within the genome of the organism (*e.g.*, fewer than about 10, about 9, about 8, about 7, about 6, about 5, about 4, about 3, about 2, or about 1 copy(ies) of the target sequence may exist in the genome). For example, the target nucleotide sequence may be located at a unique site within the genome of the organism. Target nucleotide sequences may be, for example and without limitation, randomly dispersed throughout the genome with respect to one another; located in different linkage groups in the genome; located in the same linkage group; located on different chromosomes; located on the same chromosome; located in the genome at sites that are expressed under similar conditions in the organism (*e.g.*, under the control of the same, or substantially functionally identical, regulatory factors); and located closely to one another in the genome (*e.g.*, target sequences may be comprised within nucleic acids integrated as concatemers at genomic loci).

Targeting Endonucleases

In particular embodiments, a DNA-binding polypeptide that specifically recognizes and binds to a target nucleotide sequence may be comprised within a chimeric polypeptide,

so as to confer specific binding to the target sequence upon the chimeric polypeptide. In examples, such a chimeric polypeptide may comprise, for example and without limitation, nuclease, recombinase, and/or ligase polypeptides, as these polypeptides are described above. Chimeric polypeptides comprising a DNA-binding polypeptide and a nuclease, recombinase, and/or ligase polypeptide may also comprise other functional polypeptide motifs and/or domains, such as for example and without limitation: a spacer sequence positioned between the functional polypeptides in the chimeric protein; a leader peptide; a peptide that targets the fusion protein to an organelle (*e.g.*, the nucleus); polypeptides that are cleaved by a cellular enzyme; peptide tags (*e.g.*, Myc, His, etc.); and other amino acid sequences that do not interfere with the function of the chimeric polypeptide.

Functional polypeptides (*e.g.*, DNA-binding polypeptides and nuclease polypeptides) in a chimeric polypeptide may be operatively linked. In some embodiments, functional polypeptides of a chimeric polypeptide may be operatively linked by their expression from a single polynucleotide encoding at least the functional polypeptides ligated to each other in-frame, so as to create a chimeric gene encoding a chimeric protein. In alternative embodiments, the functional polypeptides of a chimeric polypeptide may be operatively linked by other means, such as by cross-linkage of independently expressed polypeptides.

In some embodiments, a DNA-binding polypeptide, or guide RNA that specifically recognizes and binds to a target nucleotide sequence may be comprised within a natural isolated protein (or mutant thereof), wherein the natural isolated protein or mutant thereof also comprises a nuclease polypeptide (and may also comprise a recombinase and/or ligase polypeptide). Examples of such isolated proteins include TALENs, recombinases (*e.g.*, Cre, Hin, Tre, and FLP recombinase), CRISPR systems (*e.g.*, CRISPR/Cas9 systems), and meganucleases.

As used herein, the term "targeting endonuclease" refers to natural or engineered isolated proteins and mutants thereof that comprise a DNA-binding polypeptide or guide RNA and a nuclease polypeptide, as well as to chimeric polypeptides comprising a DNA-binding polypeptide or guide RNA and a nuclease. Any targeting endonuclease comprising a DNA-binding polypeptide or guide RNA that specifically recognizes and binds to a target nucleotide sequence comprised within an ANPEP locus (*e.g.*, either because the target sequence is comprised within the native sequence at the locus, or because the target

sequence has been introduced into the locus, for example, by recombination) may be utilized in certain embodiments.

Some examples of chimeric polypeptides that may be useful in particular embodiments of the present teachings include, without limitation, combinations of the following polypeptides: zinc finger DNA-binding polypeptides; a Fok I nuclease polypeptide; TALE domains; leucine zippers; transcription factor DNA-binding motifs; and DNA recognition and/or cleavage domains isolated from, for example and without limitation, a TALEN, a recombinase (*e.g.*, Cre, Hin, RecA, Tre, and FLP recombinases), a CRISPR system (*e.g.*, CRISPR/Cas9 system), a meganuclease; and others known to those in the art. Particular examples include a chimeric protein comprising a site-specific DNA binding polypeptide and a nuclease polypeptide. Chimeric polypeptides may be engineered by methods known to those of skill in the art to alter the recognition sequence of a DNA-binding polypeptide comprised within the chimeric polypeptide, so as to target the chimeric polypeptide to a particular nucleotide sequence of interest.

In certain embodiments, the chimeric polypeptide comprises a DNA-binding domain (*e.g.*, zinc finger, TAL-effector domain, etc.) and a nuclease (cleavage) domain. The cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a TALEN DNA-binding domain and a cleavage domain, or meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. See, for example, 2002-2003 Catalogue, New England Biolabs, Beverly, Mass.; and Belfort, M. and Roberts, J., Nucleic Acids Res., 1997, 25, 3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, 51 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; see also Linn et al. (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-

domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However, any integral number of nucleotides, or nucleotide pairs, can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding, for example, such that one or more exogenous sequences (donors/transgenes) are integrated at or near the binding (target) sites. Certain restriction enzymes (e.g., Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme Fok I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. See, for example, U.S. Pat. Nos. 5,356,802; 5,436,150 and 5,487,994; as well as Li, L., *et al.*, Proc. Natl. Acad. Sci. USA, 1992, 89, 4275-4279; Li, L., *et al.*, Proc. Natl. Acad. Sci. USA, 1993, 90, 2764-2768; Kim,Y-G., *et al.*, Proc. Natl. Acad. Sci. USA, 1994, 91, 883-887; Kim, Y-G., *et al.*, J. Biol. Chem., 1994, 269, 31978-31982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is Fok I. This particular enzyme is active as a dimer (Bitinaite, J., *et al.*, Proc. Natl. Acad. Sci. USA, 1998, 95, 10570-10575). Accordingly, for the purposes of the present disclosure, the portion of the Fok I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage

and/or targeted replacement of cellular sequences using zinc finger-Fok I fusions, two fusion proteins, each comprising a Fok I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a DNA binding domain and two Fok I cleavage half-domains can also be used.

5 A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (e.g., dimerize) to form a functional cleavage domain.

Exemplary Type IIS restriction enzymes are described in U.S. Patent Publication No. 20070134796, incorporated herein in its entirety. Additional restriction enzymes also

10 contain separable binding and cleavage domains, and these are contemplated by the present disclosure. See, for example, Roberts, R.J., *et al.*, Nucleic Acids Res., 2003, 31, 418-420.

In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Publication Nos.

15 20050064474; 20060188987 and 20080131962, the disclosures of all of which are incorporated by reference in their entireties herein.

Alternatively, nucleases may be assembled *in vivo* at the nucleic acid target site using so-called "split-enzyme" technology (see e.g. U.S. Patent Publication No. 20090068164). Components of such split enzymes may be expressed either on separate

20 expression constructs, or can be linked in one open reading frame where the individual components are separated, for example, by a self-cleaving 2A peptide or IRES sequence. Components may be individual zinc finger binding domains or domains of a meganuclease nucleic acid binding domain.

Zinc Finger Nucleases

25 In specific embodiments, a chimeric polypeptide is a custom-designed zinc finger nuclease (ZFN) that may be designed to deliver a targeted site-specific double-strand DNA break into which an exogenous nucleic acid, or donor DNA, may be integrated (See US Patent publication 20100257638, incorporated by reference herein). ZFNs are chimeric polypeptides containing a non-specific cleavage domain from a restriction endonuclease

30 (for example, Fok I) and a zinc finger DNA-binding domain polypeptide. See, *e.g.*, Huang, B., *et al.*, J. Protein Chem., 1996, 15, 481-489; Kim, J-S., *et al.* Proc. Natl. Acad. Sci. USA, 1997, 94, 3616-3620; Kim, Y-G., *et al.*, Proc. Natl. Acad. Sci. USA, 1996, 93, 1156-

1160; Kim, Y-G., *et al.*, Proc Natl. Acad. Sci. USA, 1994, 91, 883-887; Kim, Y-G., *et al.*, Proc. Natl. Acad. Sci. USA, 1997, 94, 12875-12879; Kim, Y-G., *et al.*, Gene, 1997, 203, 43-49; Kim, Y-G., *et al.*, Biol. Chem., 1998, 379, 489-495; Nahon, E. and Raveh, D., Nucleic Acids Res., 1998, 26, 1233-1239; Smith, J., *et al.*, Nucleic Acids Res., 1999, 27,
5 674-681. In some embodiments, the ZFNs comprise non-canonical zinc finger DNA binding domains (see US Patent publication 20080182332, incorporated by reference herein). The Fok I restriction endonuclease must dimerize via the nuclease domain in order to cleave DNA and introduce a double-strand break. Consequently, ZFNs containing a nuclease domain from such an endonuclease also require dimerization of the nuclease
10 domain in order to cleave target DNA (Mani, M. *et al.*, Biochem. Biophys. Res. Commun., 2005, 334, 1191-1197; Smith, J., *et al.*, Nucleic Acids Res. 2000, 28, 3361-3369). Dimerization of the ZFN can be facilitated by two adjacent, oppositely oriented DNA-binding sites (Id).

Optional Exogenous Nucleic Acids for Integration at an ANPEP Locus

15 Embodiments of the present teachings may include one or more nucleic acids selected from the group consisting of: an exogenous nucleic acid for site-specific integration in an ANPEP locus, for example and without limitation, an ORF; a nucleic acid comprising a nucleotide sequence encoding a targeting endonuclease; and a vector comprising at least one of either or both of the foregoing. Thus, particular nucleic acids for
20 use in some embodiments include nucleotide sequences encoding a polypeptide, structural nucleotide sequences, and/or DNA-binding polypeptide recognition and binding sites.

Optional Exogenous Nucleic Acid Molecules for Site-Specific Integration

As noted above, insertion of an exogenous sequence (also called a "donor sequence" or "donor") is provided, for example for expression of a polypeptide, correction
25 of a mutant gene or for increased expression of a wild-type gene. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence where it is placed. A donor sequence can contain a non-homologous sequence flanked by two regions of homology to allow for efficient HDR at the location of interest. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous
30 to the region of interest in cellular chromatin. A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. For example, for targeted insertion of sequences not normally present in a region of interest, said sequences can be

present in a donor nucleic acid molecule and flanked by regions of homology to sequence in the region of interest.

The donor polynucleotide can be DNA or RNA, single-stranded or double-stranded and can be introduced into a cell in linear or circular form. See e.g., U.S. Patent Publication
5 Nos. 20100047805, 20110281361, 20110207221 and U.S. application Ser. No. 13/889,162. If introduced in linear form, the ends of the donor sequence can be protected (e.g. from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends (see, *e.g.*
10 Chang, X-B. and Wilson, J.H., Proc. Natl. Acad. Sci. USA, 1987, 84, 4959-4963; Nehls, M., *et al.*, Science, 1996, 272, 886-889). Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues.
15 A polynucleotide can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins or promoters. Moreover, donor polynucleotides can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses; exemplary viruses include adenovirus, AAV, herpesvirus, retrovirus, lentivirus and integrase defective
20 lentivirus (IDLV).

The donor is generally integrated so that its expression is driven by the endogenous promoter at the integration site, namely the promoter that drives expression of the endogenous gene into which the donor is integrated. However, it will be apparent that the donor may comprise a promoter and/or enhancer, for example a constitutive promoter or an
25 inducible or tissue-specific promoter.

Furthermore, although not required for expression, exogenous sequences may also include transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals.

30 *Nucleic Acid Molecules Comprising a Nucleotide Sequence Encoding a Targeting Endonuclease*

In some embodiments, a nucleotide sequence encoding a targeting endonuclease may be engineered by manipulation (*e.g.*, ligation) of native nucleotide sequences encoding polypeptides comprised within the targeting endonuclease. For example, the nucleotide sequence of a gene encoding a protein comprising a DNA-binding polypeptide may be inspected to identify the nucleotide sequence of the gene that corresponds to the DNA-binding polypeptide, and that nucleotide sequence may be used as an element of a nucleotide sequence encoding a targeting endonuclease comprising the DNA-binding polypeptide. Alternatively, the amino acid sequence of a targeting endonuclease may be used to deduce a nucleotide sequence encoding the targeting endonuclease, for example, according to the degeneracy of the genetic code.

In exemplary nucleic acid molecules comprising a nucleotide sequence encoding a targeting endonuclease, the last codon of a first polynucleotide sequence encoding a nuclease polypeptide, and the first codon of a second polynucleotide sequence encoding a DNA-binding polypeptide, may be separated by any number of nucleotide triplets, e.g., without coding for an intron or a "STOP." Likewise, the last codon of a nucleotide sequence encoding a first polynucleotide sequence encoding a DNA-binding polypeptide, and the first codon of a second polynucleotide sequence encoding a nuclease polypeptide, may be separated by any number of nucleotide triplets. In these and further embodiments, the last codon of the last of a first polynucleotide sequence encoding a nuclease polypeptide, and a second polynucleotide sequence encoding a DNA-binding polypeptide, may be fused in phase-register with the first codon of a further polynucleotide coding sequence directly contiguous thereto, or separated therefrom by no more than a short peptide sequence, such as that encoded by a synthetic nucleotide linker (*e.g.*, a nucleotide linker that may have been used to achieve the fusion). Examples of such further polynucleotide sequences include, for example and without limitation, tags, targeting peptides, and enzymatic cleavage sites. Likewise, the first codon of the most 5' (in the nucleic acid sequence) of the first and second polynucleotide sequences may be fused in phase-register with the last codon of a further polynucleotide coding sequence directly contiguous thereto, or separated therefrom by no more than a short peptide sequence.

A sequence separating polynucleotide sequences encoding functional polypeptides in a targeting endonuclease (e.g., a DNA-binding polypeptide and a nuclease polypeptide) may, for example, consist of any sequence, such that the amino acid sequence encoded is

not likely to significantly alter the translation of the targeting endonuclease. Due to the autonomous nature of known nuclease polypeptides and known DNA-binding polypeptides, intervening sequences generally will not interfere with the respective functions of these structures.

5 *Other Knockout Methods*

Various other techniques known in the art can be used to inactivate genes to make knock-out animals and/or to introduce nucleic acid constructs into animals to produce founder animals, in which the knockout or nucleic acid construct is integrated into the genome. Such techniques include, without limitation, pronuclear microinjection (U.S. Pat. 10 No. 4,873,191), retrovirus mediated gene transfer into germ lines (Van der Putten H., et al., Proc. Natl. Acad. Sci. USA, 1985, 82, 6148-1652), gene targeting into embryonic stem cells (Thompson, S., et al., Cell, 1989, 56, 313-321), electroporation of embryos (Lo, W., Mol. Cell. Biol., 1983, 3, 1803-1814), sperm-mediated gene transfer (Lavitrano, M., et al., Proc. Natl. Acad. Sci. USA, 2002 99, 14230-14235; Lavitrano, M., et al., Reprod. Fert. 15 Develop., 2006, 18, 19-23), and *in vitro* transformation of somatic cells, such as cumulus or mammary cells, or adult, fetal, or embryonic stem cells, followed by nuclear transplantation (Wilmut, I, et al., Nature, 1997, 385, 810-813 and Wakayama, T., et al., Nature, 1998, 394, 369-374). Pronuclear microinjection, sperm mediated gene transfer, and somatic cell nuclear transfer are particularly useful techniques. An animal that is gene 20 edited is an animal wherein all of its cells have the genetic edit, including its germ line cells. When methods are used that produce an animal that is mosaic in its genetic edit, the animals may be bred and progeny that are gene edited may be selected. Cloning, for instance, may be used to make a mosaic animal if its cells are edited at the blastocyst state, or gene editing can take place when a single cell is edited. Animals that are edited so that 25 they exhibit reduced susceptibility to infection by a coronavirus can be homozygous or heterozygous for the edit, depending on the specific approach that is used. If a particular gene is inactivated by an edit, homozygosity would normally be required. However, the instant application provides for and includes mosaic animals and heterozygous animals that may be used to breed homozygous animals. Further, if a particular gene is inactivated by 30 an RNA interference or dominant negative strategy, then heterozygosity is often adequate.

Typically, in embryo/zygote microinjection, a nucleic acid construct or mRNA is introduced into a fertilized egg; 1 or 2 cell fertilized eggs are used as the pronuclei

containing the genetic material from the sperm head and the egg are visible within the protoplasm. Pronuclear staged fertilized eggs can be obtained *in vitro* or *in vivo* (*i.e.*, surgically recovered from the oviduct of donor animals). *In vitro* fertilized eggs can be produced as follows. For example, swine ovaries can be collected at an abattoir, and maintained at 22-28° C during transport. Ovaries can be washed and isolated for follicular aspiration, and follicles ranging from 4-8 mm can be aspirated into 50 mL conical centrifuge tubes using 18-gauge needles and under vacuum to suck the oocyte from the follicle on the ovary. Follicular fluid and aspirated oocytes can be rinsed through pre-filters with commercial TL-HEPES (Minitube, Verona, WI). Oocytes surrounded by a compact cumulus mass can be selected and placed into TCM-199 OOCYTE MATURATION MEDIUM (Minitube, Verona, WI) supplemented with 0.1 mg/mL cysteine, 10 ng/mL epidermal growth factor, 10% porcine follicular fluid, 50 µM 2-mercaptoethanol, 0.5 mg/ml cAMP, 10 IU/mL each of pregnant mare serum gonadotropin (PMSG) and human chorionic gonadotropin (hCG) for approximately 22 hours in humidified air at 38.7° C and 5% CO_2. Subsequently, the oocytes can be moved to fresh TCM-199 maturation medium, which will not contain cAMP, PMSG or hCG and then the oocytes can be incubated for an additional 22 hours. Matured oocytes can be stripped of their cumulus cells by vortexing in 0.1% hyaluronidase for 1 minute.

For swine, mature oocytes can be fertilized in 500 µl Minitube PORCPRO IVF MEDIUM SYSTEM (Minitube, Verona, Wis.) in Minitube 5-well fertilization dishes. In preparation for *in vitro* fertilization (IVF), freshly collected or frozen boar semen can be washed and resuspended in PORCPRO IVF Medium to 4×10^5 sperm. Sperm concentrations can be analyzed by computer assisted semen analysis (SPERM VISION®, Minitube, Verona, WI). Final *in vitro* insemination can be performed in a 10 µl volume at a final concentration of approximately 40 motile sperm/oocyte, depending on the boar. The oocytes can be incubated at 38.7° C in 5.0% CO_2 atmosphere for 6 hours. Six hours post-insemination, presumptive zygotes can be washed twice in NCSU-23 and moved to 0.5 mL of the same medium. This system can produce 20-30% blastocysts routinely across most boars with a 10-30% polyspermic insemination rate.

Linearized nucleic acid constructs or mRNA can be injected into one of the pronuclei or into the cytoplasm. Then the injected eggs can be transferred to a recipient female (e.g., into the oviducts of a recipient female) and allowed to develop in the recipient

female to produce the genetically edited animals. In particular, *in vitro* fertilized embryos can be centrifuged at 15,000 x g for 5 minutes to sediment lipids allowing visualization of the pronucleus. The embryos can be injected with using an EPPENDORF® FEMTOJET® injector (EPPENDORF®, Hamburg, Germany) and can be cultured until blastocyst formation. Rates of embryo cleavage and blastocyst formation and quality can be recorded.

Embryos can be surgically transferred into uteri of asynchronous recipients. Typically, 20-200 embryos can be deposited into the ampulla-isthmus junction of the oviduct using a catheter. After surgery, real-time ultrasound examination of pregnancy can be performed.

In somatic cell nuclear transfer (SCNT), a gene edited cell (*e.g.*, a gene edited pig cell) such as an embryonic blastomere, fetal fibroblast, adult ear fibroblast, or granulosa cell that includes a nucleic acid construct described above, can be introduced into an enucleated oocyte to establish a combined cell. Oocytes can be enucleated by partial zona dissection near the polar body and then pressing out cytoplasm at the dissection area. Typically, an injection pipette with a sharp beveled tip is used to inject the gene edited cell into an enucleated oocyte arrested at meiosis 2. In some conventions, oocytes arrested at meiosis-2 are termed eggs. After producing a porcine or bovine embryo (*e.g.*, by fusing and activating the oocyte), the embryo is transferred to the oviducts of a recipient female, about 20 to 24 hours after activation. See, for example, Cibelli, J.B., *et al.*, Science, 1998, 280, 1256-1258 and U.S. Pat. No. 6,548,741. For pigs, recipient females can be checked for pregnancy approximately 20-21 days after transfer of the embryos.

Standard breeding techniques can be used to create animals that are homozygous for the desired edit from the initial heterozygous or mosaic founder animals. Homozygosity may not be required, however. Gene edited animals described herein can be bred with other animals of interest. Regardless of whether or not heterozygous animals are resistant to coronavirus, these animals have utility at least in that they can be used in breeding, for example to breed homozygous animals. In order to add a gene edit to the germplasm of an elite herd, many different zygotes can be edited using methods of the present teachings. These animals can then be interbred to create a herd comprising an elite germplasm that includes the edit. This is especially important in establishing a pig herd with the edit, as it allows for the rapid integration of the edited gene into the pig herd.

In some embodiments, a nucleic acid of interest and a selectable marker can be provided on separate transposons and provided to either embryos or cells in unequal amount, where the amount of transposon containing the selectable marker far exceeds (5-10 fold excess) the transposon containing the nucleic acid of interest. Gene edited cells or animals expressing the nucleic acid of interest can be isolated based on presence and expression of the selectable marker. Because the transposons will integrate into the genome in a precise and unlinked way (independent transposition events), the nucleic acid of interest and the selectable marker are not genetically linked and can easily be separated by genetic segregation through standard breeding. Thus, gene edited animals can be produced that are not constrained to retain selectable markers in subsequent generations.

Once gene edited animals have been generated, expression of a nucleic acid can be assessed using standard techniques. Initial screening can be accomplished by ILLUMINA® sequencing (ILLUMINA®, Inc., San Diego, CA) or Southern blot analysis to determine whether or not an edit has taken place.

Expression of a nucleic acid sequence encoding a polypeptide in the tissues of gene edited animals can be assessed using techniques that include, for example, northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, western analysis, immunoassays such as enzyme-linked immunosorbent assays, and reverse-transcriptase PCR (RT-PCR).

Interfering RNAs

A variety of interfering RNA (RNAi) systems are known. Double-stranded RNA (dsRNA) induces sequence-specific degradation of homologous gene transcripts. RNA-induced silencing complex (RISC) metabolizes dsRNA to small 21-23-nucleotide small interfering RNAs (siRNAs). RISC contains a double-stranded RNAse (dsRNase, *e.g.*, Dicer) and ssRNase (*e.g.*, Argonaut 2 or Ago2). RISC utilizes antisense strand as a guide to find a cleavable target. Both siRNAs and microRNAs (miRNAs) are known. A method of inactivating a gene in a gene edited animal comprises inducing RNA interference against a target gene and/or nucleic acid such that expression of the target gene and/or nucleic acid is reduced.

For example, the exogenous nucleic acid sequence can induce RNA interference against a nucleic acid encoding a polypeptide. For example, double-stranded small interfering RNA (siRNA) or small hairpin RNA (shRNA) homologous to a target DNA

can be used to reduce expression of that DNA. Constructs for siRNA can be produced as described, for example, in Fire, A., *et al.*, Nature, 1998, 391, 806-811; Romano, N. and Masino, G., Mol. Microbiol., 1992, 6, 3343-3353; Cogoni, C., *et al.*, EMBO J., 1996, 15, 3153-3161; Cogoni, C. and Masino, G., Nature, 1999, 399, 166-169; Misquitta, L. and
5 Paterson, B.M., Proc. Natl. Acad. Sci. USA, 1999, 96, 1451-1456; and Kennerdell, K. and Carthew, C., Cell, 1998 95, 1017-1026. Constructs for shRNA can be produced as described by McIntyre, G.J. and Fanning, G.C., BMC Biotechnology, 2006, 6, 1-8. In general, shRNAs are transcribed as a single-stranded RNA molecule containing complementary regions, which can anneal and form short hairpins.

10 The probability of finding a single, individual functional siRNA or miRNA directed to a specific gene is high. The predictability of a specific sequence of siRNA, for instance, is about 50% but a number of interfering RNAs may be made with good confidence that at least one of them will be effective.

Embodiments include an *in vitro* cell, an *in vivo* cell, and a gene edited animal such
15 as a livestock animal that expresses an RNAi directed against an ANPEP gene. The RNAi may be, for instance, selected from the group consisting of siRNA, shRNA, dsRNA, and miRNA.

Vectors and Nucleic Acids

A variety of nucleic acids may be introduced into cells for knockout purposes, for
20 inactivation of a gene, to obtain expression of a gene, or for other purposes. As used herein, the term nucleic acid includes DNA, RNA, and nucleic acid analogs, and nucleic acids that are double-stranded or single-stranded *(i.e.*, a sense or an antisense single strand). Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the
25 nucleic acid. Modifications at the base moiety include deoxyuridine for deoxythymidine, and 5-methyl-2'-deoxycytidine and 5-bromo-2'-doxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a
30 six membered, morpholino ring, or peptide nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. See, Summerton, J. and Weller, D., Antisense Nucleic Acid Drug Dev., 1997, 7, 187-195; and

Hyrup, B and Nielsen, P.E., Bioorgan. Med. Chem., 1996, 4, 5-23. In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

The target nucleic acid sequence can be operably linked to a regulatory region such as a promoter. Regulatory regions can be porcine regulatory regions or can be from other species. As used herein, operably linked refers to positioning of a regulatory region relative to a nucleic acid sequence in such a way as to permit or facilitate transcription of the target nucleic acid.

Any type of promoter can be operably linked to a target nucleic acid sequence. Examples of promoters include, without limitation, tissue-specific promoters, constitutive promoters, inducible promoters, and promoters responsive or unresponsive to a particular stimulus. Suitable tissue-specific promoters can result in preferential expression of a nucleic acid transcript in beta cells and include, for example, the human insulin promoter. Other tissue-specific promoters can result in preferential expression in, for example, hepatocytes or heart tissue and can include the albumin or alpha-myosin heavy chain promoters, respectively. In other embodiments, a promoter that facilitates the expression of a nucleic acid molecule without significant tissue or temporal specificity can be used (*i.e.*, a constitutive promoter). For example, a beta-actin promoter such as the chicken beta-actin gene promoter, ubiquitin promoter, miniCAGs promoter, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter, or 3-phosphoglycerate kinase (PGK) promoter can be used, as well as viral promoters such as the herpes simplex virus thymidine kinase (HSV-TK) promoter, the SV40 promoter, or a cytomegalovirus (CMV) promoter. In some embodiments, a fusion of the chicken beta actin gene promoter and the CMV enhancer is used as a promoter. See, for example, Xu, L., *et al.*, Hum. Gene Ther., 2001, 12, 563-573; and Kiwaki, K., *et al.*, Hum. Gene Ther., 1996, 7, 821-830.

Additional regulatory regions that may be useful in nucleic acid constructs, include, but are not limited to, polyadenylation sequences, translation control sequences (*e.g.*, an internal ribosome entry segment, IRES), enhancers, inducible elements, or introns. Such regulatory regions may not be necessary, although they may increase expression by affecting transcription, stability of the mRNA, translational efficiency, or the like. Such regulatory regions can be included in a nucleic acid construct as desired to obtain optimal

expression of the nucleic acids in the cell(s). Sufficient expression, however, can sometimes be obtained without such additional elements.

Nucleic acid constructs can be introduced into embryonic, fetal, or adult animal cells of any type, including, for example, germ cells such as an oocyte or an egg, a progenitor cell, an adult or embryonic stem cell, a primordial germ cell, a kidney cell such as a PK-15 cell, an islet cell, a beta cell, a liver cell, or a fibroblast such as a dermal fibroblast, using a variety of techniques. Non-limiting examples of techniques include the use of transposon systems, recombinant viruses that can infect cells, or liposomes or other non-viral methods such as electroporation, microinjection, or calcium phosphate precipitation, that are capable of delivering nucleic acids to cells.

Nucleic acids can be incorporated into vectors. A vector is a broad term that includes any specific DNA segment that is designed to move from a carrier into a target DNA. A vector may be referred to as an expression vector, or a vector system, which is a set of components needed to bring about DNA insertion into a genome or other targeted DNA sequence such as an episome, plasmid, or even virus/phage DNA segment. Vector systems such as viral vectors (*e.g.*, retroviruses, adeno-associated virus and integrating phage viruses), and non-viral vectors (*e.g.*, transposons) used for gene delivery in animals have two basic components: 1) a vector comprised of DNA (or RNA that is reverse transcribed into a cDNA) and 2) a transposase, recombinase, or other integrase enzyme that recognizes both the vector and a DNA target sequence and inserts the vector into the target DNA sequence. Vectors most often contain one or more expression cassettes that comprise one or more expression control sequences, wherein an expression control sequence is a DNA sequence that controls and regulates the transcription and/or translation of another DNA sequence or mRNA, respectively.

Many different types of vectors are known. For example, plasmids and viral vectors, *e.g.*, retroviral vectors, are known. Mammalian expression plasmids typically have an origin of replication, a suitable promoter and optional enhancer, necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences. Examples of vectors include: plasmids (which may also be a carrier of another type of vector), adenovirus, adeno-associated virus (AAV), lentivirus (*e.g.*, modified HIV-1, SIV or FIV), retrovirus

(*e.g.*, ASV, ALV or MoMLV), and transposons (*e.g.*, Sleeping Beauty, P-elements, Tol-2, Frog Prince, piggyBac).

As used herein, the term nucleic acid refers to both RNA and DNA, including, for example, cDNA, genomic DNA, synthetic (*e.g.*, chemically synthesized) DNA, as well as
5 naturally occurring and chemically modified nucleic acids, *e.g.*, synthetic bases or alternative backbones. A nucleic acid molecule can be double-stranded or single-stranded (*i.e.*, a sense or an antisense single strand).

Founder Animals, Traits, and Reproduction

Founder animals may be produced by cloning and other methods described herein.
10 The founders can be homozygous for a gene edit, as in the case where a zygote or a primary cell undergoes a homozygous edit. Similarly, founders can also be made that are heterozygous. All of a founder's cells genome may carry the gene edit or founders may be mosaic for an edit. Mosaicism depends on the stage at which the gene editing reagents were introduced, the half life of the reagents, and at what stage editing occurred. For
15 example, viable gene editing reagents may still be in a two-cell stage embryo that was injected as a zygote. Progeny of mosaic animals may be tested to identify progeny that have the desired edit.

In livestock, many alleles are known to be linked to various traits such as production traits, type traits, workability traits, and other functional traits. Practitioners are
20 accustomed to monitoring and quantifying these traits, see *e.g.*, Visscher, P.M., *et al.*, Livestock Production Science, 1994, 40, 123-137, U.S. Pat. No. 7,709,206, US 2001/0016315, US 2011/0023140, and US 2005/0153317. An animal may include a trait chosen from a trait in the group consisting of a production trait, a type trait, a workability trait, a fertility trait, a mothering trait, and a disease resistance trait. Further traits include
25 expression of a recombinant gene product. Animals with a desired trait or traits may be edited to increase their resistance to infection with a coronavirus.

In addition to monitoring traits, artisans can look at the genetic background of the animal as a whole. Gene edits such as those of the present teachings may be made in elite genetic backgrounds. Elite PIC™ (Pig Improvement Company, Limited, Basingstoke,
30 UK) lines 2, 3, 15, 19, 27, 62 and 65 are lines selected for superior commercial phenotypes. Fibroblast cell lines can be grown from collagenase treated ear notch samples extracted from the animals from these cell lines using methods known in the art.

Other potential porcine lines include lines can be PIC™ Line 15, PIC™ Line 17, PIC™ Line 27, PIC™ Line 65, PIC™ Line 14, PIC™ Line 62, PIC337, PIC800, PIC280, PIC327, PIC408, PIC™ 399, PIC410, PIC415, PIC359, PIC380, PIC837, PIC260, PIC265, PIC210, PIC™ Line 2, PIC™ Line 3, PIC™ Line 4, PIC™ Line 5, PIC™ Line 18, PIC™
5 Line 19, PIC™ Line 92, PIC95, PIC™ CAMBOROUGH® (Pig Improvement Company, Limited, Basingstoke, UK), PIC1070, PIC™ CAMBOROUGH® 40, PIC™ CAMBOROUGH® 22, PIC1050, PIC™ CAMBOROUGH® 29, PIC™ CAMBOROUGH® 48, or PIC™ CAMBOROUGH® x54.

In various aspects, PIC™ Line 65 is sold under the trade name PIC337. In various
10 aspects, PIC™ line 62 is sold under the tradename PIC408. In various aspects, hybrid pigs made by crossing PIC™ lines 15 and 17 are sold under the tradenames PIC800 or PIC280. In various aspects, PIC™ Line 27 is sold under the tradename PIC327. In various aspects, hybrids created from crossing PIC™ Line 65 and PIC™ Line 62 are sold under the tradenames PIC399, PIC410, or PIC415. In various aspects, hybrids created from crossing
15 PIC™ Line 65 and Pic Line 27 are sold under the tradename PIC359. In various aspects, hybrids prepared from crossing PIC™ Line 800 pigs (which is a hybrid of PIC™ Line 15 and PIC™ Line 17) to PIC™ Line 65 pigs are sold under the tradenames PIC380 or PIC837. In various aspects, PIC™ Line 14 is sold under the trade name PIC260. In various aspects, hybrids created from crossing PIC™ Line 14 and PIC™ Line 65 are sold under
20 the tradename PIC265. In various aspects, hybrids created by crossing PIC™ Line 2 and PIC™ Line 3 are sold under the tradenames PIC210, PIC™ CAMBOROUGH®, and PIC1050. In various aspects, hybrids of PIC™ Line 3 and PIC™ Line 92 are sold under the tradename PIC95. In various configurations, hybrids made from crossing PIC™ Line 19 and PIC™ Line 3 are sold under the tradename PIC1070. In various aspects, hybrids
25 created by crossing PIC™ Line 18 and PIC™ Line 3 are sold under the tradename PIC™ CAMBOROUGH® 40. In various aspects, hybrids created from crossing PIC™ Line 19 and PIC1050 (which is itself a hybrid of PIC™ lines 2 and 3) are sold under the tradename PIC™ CAMBOROUGH® 22. In various aspects, hybrids created from crossing PIC™ Line 2 and PIC1070 (which is itself a hybrid of PIC™ lines 19 and 3) are sold under the
30 tradename PIC™ CAMBOROUGH® 29. In various aspects, hybrids created from crossing PIC™ Line 18 and PIC1050 (which is itself a hybrid of PIC™ lines 2 and 3) are sold under the tradename PIC™ CAMBOROUGH® 48. In various aspects, hybrids created from

crossing PIC™ Line 4 and PIC™ Line 5 are sold under the tradename PIC™ CAMBOROUGH® x54.

Confirmation of Edits

One useful method of detecting the desired edit is to use real-time PCR. PCR primers flanking the region of interest and a probe that specifically anneals to the region of interest can be designed. The probe can be labelled with both a fluorophore and a quencher. In the PCR reaction, the primers and probe hybridize in a sequence-dependent manner to the complementary DNA strand of the region of interest. Because the probe is intact, the fluorophore and quencher are in close proximity and the quencher absorbs fluorescence emitted by the fluorophore. The polymerase extends from the primers and begins DNA synthesis. When the polymerase reaches the probe, the exonuclease activity of the polymerase cleaves the hybridized probe. As a result of cleavage, the fluorophore is separated from the quencher and fluoresces. This fluorescence is detected by the real time instrument. These steps are repeated for each PCR cycle and allow detection of specific products. A commercial real-time PCR kit can be used to probe various animals for the desired edit. A variety of commercial real-time PCR kits exist including, such as, but without limitation, PRIMETIME® from IDT, TAQMAN® (Roche Molecular Systems, Inc, Pleasonton, CA) from Applied Biosystems, and various kits from Qiagen and Bio-Rad. Skilled persons will recognize that any such kit can be used with the primers and methods of the present teachings to achieve like results.

EXAMPLES

The present teachings including descriptions provided in the Examples that are not intended to limit the scope of any claim or embodiment. The following non-limiting examples are provided to further illustrate the present teachings. Those of skill in the art, in light of the present disclosure, will appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the present teachings.

Example 1

This example demonstrates the use of gRNAs, in combination with Cas protein activity, to generate double-stranded breaks in DNA, which upon host-cell mediated non-homologous end-joining, results in the introduction of a premature stop codons in coding

regions of ANPEP. Upon protein translation of the resultant mRNA encoded by the new ANPEP gene, premature stop codons in ANPEP mRNA would result in premature termination of protein translation and ultimately truncated and non-functional proteins, and/or initiate nonsense-mediated mRNA decay resulting in elimination of ANPEP mRNA
5 transcripts. Premature stop codons can be introduced via the homology directed repair (HDR) pathway by inclusion of a single- or double-stranded DNA template in editing experiments. However, exogenously added DNA can integrate randomly in the genome. Therefore, it would be advantageous to identify single gRNAs or gRNA pairs that promote the formation of in-frame stop codons, without the introduction of non-wild type amino
10 acids. To accomplish this, one or two gRNAs can be used to direct nuclease cut sites, which are then repaired by NHEJ to affect the desired edit.

Guides within ANPEP were tested computationally for their ability to generate in-frame stop codons when paired. Computational predictions were subsequently tested in porcine fetal fibroblasts (PFFs), as described below. The gRNAs listed in Table 3 below
15 were generated by *in vitro* transcription and complexed with SpyCas9 in water, using 3.2 µg of Cas9 protein and 2.2 µg of gRNA in a total volume of 2.23 µl. Resulting ribonucleoprotein (RNP) complexes were then combined 1:1 in a total volume of 2.23 µl to generate gRNA pairs, as indicated in Table 3, and nucleofected into PFFs using a Lonza electroporator. In preparation for nucleofection, PFF cells were harvested using TRYPLE
20 EXPRESS™ (ThermoFisher, recombinant Trypsin): the culture medium was removed from cells, cells were washed once with Hank's Balanced Salt Solution (HBSS) or Dulbecco's Phosphate-Buffered Saline (DPBS), and incubated for 3–5 minutes at 38.5° C in the presence of TrypLE. Cells were then harvested with complete medium. Cells were pelleted via centrifugation (300 g x 5 minutes at room temperature), supernatant was
25 discarded, and then the cells were resuspended in 10 mL PBS to obtain single cell suspensions to allow cell counting using trypan blue staining. After counting, cells were pelleted via centrifugation, the supernatant was discarded, and the cells were resuspended in nucleofection buffer P3 at a final concentration of 7.5×10^6 cells/ml. 20 µl of the cell suspension was added to each well of a nucleofection cuvette containing the RNP mixture
30 and then mixed gently to resuspend the cells. The RNP/cell mixture was then nucleofected with program CM138 (provided by the manufacturer). Following nucleofection, 80 µl of warm Embryonic Fibroblast Medium (EFM) (Dulbecco's Modified Eagle's Medium

(DMEM) containing 2.77 mM glucose, 1.99 mM L-glutamine, and 0.5 mM sodium pyruvate, supplemented with 100 µM 2-Mercaptoethanol, 1X Eagle's minimum essential medium non-essential amino acids (MEM NEAA), 100 µg/mL Penicillin-Streptomycin, and 12% Fetal Bovine Serum) was added to each well. The suspensions were mixed gently
5 by pipetting, and then 100 µl was transferred to a 12-well plate containing 900 µl of EFM pre-incubated at 38.5°C. The plate was then incubated at 38.5°C, 5% CO_2, 5% O_2, 90% N_2 for 48 hours. Forty-eight hours after nucleofection, genomic DNA was prepared from transfected and control PFF cells: 15 µl of QUICKEXTRACT™ DNA Extraction Solution (Lucigen) was added to pelleted cells, and the cells were then lysed by incubating for 10
10 minutes at 37° C, for 8 minutes at 65° C, and for 5 minutes at 95° C. Lysate was held at 4° C until used for DNA sequencing.

To evaluate NHEJ repair outcomes at ANPEP target sites mediated by the guide RNA/Cas endonuclease system, a region of approximately 250 bp of genomic DNA surrounding the target site was amplified by PCR and then the PCR product was examined
15 by amplicon deep sequencing for the presence and nature of repairs. After transfection in triplicate, PFF genomic DNA was extracted and the region surrounding the intended target site was PCR amplified with Q5 Polymerase (NEB) adding sequences necessary for amplicon-specific barcodes and ILLUMINA® sequencing using tailed primers through two rounds of PCR, respectively. The resulting PCR amplification products were deep
20 sequenced on an ILLUMINA® MISEQ® Personal Sequencer (ILLUMINA®, Inc., San Diego, CA). The resulting reads were examined for the presence and nature of repairs at the expected sites of cleavage by comparison to control experiments where the Cas9 protein and guide RNA were omitted from the transfection or by comparison to the reference genome. To calculate the frequency of NHEJ mutations for a target site/Cas9
25 protein/guide RNA combination, the total number of mutant reads (amplicon sequences containing insertions or deletions when compared to the DNA sequences from control treatments or reference genome) was divided by total read number (wild-type plus mutant reads) of an appropriate length containing a perfect match to the barcode and forward primer. NHEJ activity is expressed as the average (n=#) mutant fraction in Table 3
30 A subset of gRNA pairs screened in porcine fetal fibroblasts were additionally tested for their ability to introduce a premature stop codon in the ANPEP coding regions in porcine embryos. The subset of guides to be tested in porcine embryos was chosen based

on their efficacy in generating premature stop codons in porcine fibroblasts. Edited porcine embryos were generated as described below. Briefly, oocytes recovered from slaughterhouse ovaries were *in vitro* fertilized. The sgRNP solution was injected into the cytoplasm of presumptive zygotes at 16–17 hours post-fertilization by using a single pulse from a FEMTOJET® 4i microinjector (EPPENDORF®; Hamburg, Germany). Microinjection was performed in TL-Hepes (ABT360, LLC) supplemented with 3 mg/ml BSA (Proliant) on the heated stage of an inverted microscope equipped with Narishige micromanipulators (Narishige International USA, Amityville, NY). Following injections, presumptive zygotes were cultured for 7 days in PZM5 (Cosmo Bio, Co LTD, Tokyo, Japan) in an incubator environment of 5% CO_2, 5% O_2, 90% N_2. Mutation frequency of blastocysts was determined by ILLUMINA® sequencing as described for fetal fibroblasts above. The frequencies of the desired end-to-end NHEJ repairs resulting in premature stop codons in ANPEP are shown in Table 3.

Table 3 ANPEP Guides used to generate in-frame stop codons in porcine fibroblasts and blastocysts

Exon	Guide 1 SEQ ID No.	Guide 2 SEQ ID No.	Deletion Size	% desired editing fibroblasts	% desired editing blastocysts	Desired Edit SEQ ID NO.
2	102	52	317	66.7	9.5	252
2	102	26	423	68.7	10.8	253
2	102	10	510	17	ND	
2	58	52	44	40.3	50	254
2	58	26	150	21.3	ND	
2	58	10	237	3.3	ND	
2	32	26	27	<2	ND	
2	32	10	114	2.7	ND	
3	120	111	55	77.7	38.5	255
3	119	113	50	19.7	ND	
3	115	111	26	<2	ND	

4	137	130	84	13.3	ND	
4	133	130	42	<2	ND	
4	137	128	94	<2	ND	
4	133	128	52	5	ND	
16	189	183	52	<2	ND	
18	231	208	100	20.7	25	256
18	228	213	59	36.7	11.9	257

The guides encoded by Guide 1 and Guide 2 each create a double strand break in conjunction with Cas9, and then the blunt ends are joined through NHEJ to create a premature stop codon at the ligated breakpoint. (ND = not done)

5 This example demonstrates that the porcine ANPEP gene nucleotide sequences can be edited through the stimulation of double-strand breaks mediated by transfecting Cas9 protein with single or paired guide RNAs to effect a *de novo* in-frame stop codon.

Example 2

This example illustrates generation of pigs having a premature stop codon in

10 ANPEP.

Porcine oocytes will be isolated, fertilized, and then the resulting zygotes will be edited as described in Example 1 to generate gene edited pigs.

ANPEP guide RNA pairs selected from Table 3 will be used to form RNP complexes as described in The present teachings including descriptions provided in the

15 Examples that are not intended to limit the scope of any claim or embodiment. The following non-limiting examples are provided to further illustrate the present teachings. Those of skill in the art, in light of the present disclosure, will appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the present teachings.

20 Example *1* and will be microinjected into the cytoplasm of *in vivo* or *in vitro* fertilized PIC™ porcine one-cell zygotes or MII oocytes. These zygotes will then be incubated to generate edited multicellular embryos and transferred to surrogate gilts via standard methods to birth gene edited pigs. To prepare embryo donors and surrogates, PIC™ pubertal gilts will be subjected to estrus synchronization by treatment with 0.22%

altrenogest solution (20–36 mg/animal) for 14 days. Follicular growth will be induced by the administration of PMSG 36 hours following the last dose of Matrix, and ovulation will be induced by the administration of hCG 82 hours after PMSG administration. To generate *in vivo* fertilized zygotes, females in standing heat will be artificially inseminated (AI) with
5 PIC™ boar semen. *In vivo* derived zygotes will be recovered surgically 12-24 hours after AI by retrograde flushing the oviduct with sterile TL-HEPES medium supplemented with 0.3% BSA (w/v). Fertilized zygotes will be subjected to a single 2 - 50 picoliter (pl) cytoplasmic injection of Cas9 protein and guide RNA complex (25 - 50 ng/µl and 12.5 - 35 ng/µl) targeting ANPEP and cultured in PZM5 medium (Yoshioka, K., *et al.*, Biol.
10 Reprod., 2002, 60: 112-119; Suzuki, C., et al., Reprod. Fertil. Dev., 2006 18, 789-795; Yoshioka, K., J. Reprod. Dev. 2008, 54, 208-213). Injected zygotes will be surgically implanted into the oviducts of estrus synchronized, un-mated surrogate females by a mid-line laparotomy under general anesthesia (each surrogate will receive 20-60 injected embryos).

15 *In vitro* fertilized embryos for gene editing will be derived from non-fertilized PIC™ oocytes. Immature oocytes from estrus synchronized PIC™ gilts will be collected from medium size (3–6 mm) follicles. Oocytes with evenly dark cytoplasm and intact surrounding cumulus cells will be selected for maturation. Cumulus oocyte complexes will be placed in a well containing 500 µl of maturation medium, TCM-199 (Invitrogen) with
20 3.05 mM glucose, 0.91 mM sodium pyruvate, 0.57 mM cysteine, 10 ng/ml EGF, 0.5 µg/ml luteinizing hormone (LH), 0.5 µg/ml FSH, 10 ng/ml gentamicin (Sigma), and 10% follicular fluid for 42–44 h at 38.5° C and 5% CO_2, in humidified air. At the end of the maturation, the surrounding cumulus cells will be removed from the oocytes by vortexing for 3 min in the presence of 0.1% hyaluronidase. Then, *in vitro* matured oocytes will be
25 placed in 100 µl droplets of IVF medium (modified Tris-buffered medium containing 113.1 mM NaCl, 3 mM KCl, 7.5 mM $CaCl_2$, 11 mM glucose, 20 mM Tris, 2 mM caffeine, 5 mM sodium pyruvate, and 2 mg/ml bovine serum albumin (BSA)) in groups of 25–30 oocytes and will be fertilized according to established protocol (Abeydeera, L.R. and Day, B.N., *Biol. Reprod.*, 1997, 57:729–734) using fresh extended boar semen. One ml of
30 extended semen will be mixed with Dulbecco's Phosphate Buffered Saline (DPBS) containing 1 mg/ml BSA to a final volume of 10 ml and centrifuged at 1000 × g, 25° C for 4 minutes, and spermatozoa will be washed in DPBS three times. After the final wash,

spermatozoa will be re-suspended in mTBM medium and added to oocytes at a final concentration of 1×10^5 spermatozoa/ml, and co-incubated for 4-5 h at 38.5° C and 5% CO_2. Presumptive zygotes will be microinjected 5 hours post IVF and transferred to a surrogate female after 18-42 hours (1-4 cell stage). Each surrogate receives 20-60 injected embryos. Pregnancies will be confirmed by lack of return to estrus (21 days) and ultrasound at 28 days post embryo transfer.

Example 3

This example illustrates molecular characterization of pigs by sequencing.

A tissue sample will be taken from an animal with a genome that was edited according to the examples herein. Tail, ear notch, or blood samples are suitable tissue types. The tissue sample will be frozen at -20°C within 1 hour of sampling to preserve integrity of the DNA in the tissue sample.

DNA will be extracted from tissue samples after proteinase K digestion in lysis buffer. Characterization will be performed on two different sequence platforms, short sequence reads using the ILLUMINA® platform (ILLUMINA®, Inc., San Diego, CA) and long sequence reads on an Oxford NANOPORE™ platform (Oxford NANOPORE™ Technologies, Oxford, UK).

For short sequence reads, two-step PCR will be used to amplify and sequence the region of interest. The first step is a locus-specific PCR which amplifies the locus of interest from the DNA sample using a combined locus-specific primer with a vendor-specific primer. The second step attaches the sequencing index and adaptor sequences to the amplicon from the first step so that sequencing can occur.

The locus-specific primers for the first step PCR are chosen so that they amplify a region <300bp such that ILLUMINA® paired-end sequencing reads could span the amplified fragment. Multiple amplicons are preferred to provide redundancy should deletions or naturally occurring point mutations prevent primers from correctly binding. Sequence data for the amplicon will be generated using an ILLUMINA® sequencing platform (MISEQ®, ILLUMINA®, Inc., San Diego, CA). Sequence reads are analyzed to characterize the outcome of the editing process.

For long sequence reads, two-step PCR will be used to amplify and sequence the region of interest. The first step is a locus-specific PCR which amplifies the locus of interest from the DNA sample using a combined locus-specific primer with a vendor-

specific adapter. The second step PCR attaches the sequencing index to the amplicon from the first-step PCR so that the DNA is ready for preparing a sequencing library. The step 2 PCR products undergo a set of chemical reactions from a vendor kit to polish the ends of the DNA and ligate on the adapter containing the motor protein to allow access to the pores 5 for DNA strand-based sequencing.

The locus specific primers for the first step PCR range will be designed to amplify different regions of the ANPEP gene and will amplify regions different in length. Normalized DNA is then mixed with vendor supplied loading buffer and is loaded onto the NANOPORE™ flowcell.

10 Long sequence reads, while having lower per base accuracy than short reads, are very useful for observing the long-range context of the sequence around the target site.

Example 4

This example illustrates viral challenge of ANPEP edited pigs.

PIC™ pigs will be edited with guides as described in Example 2. Edits will be 15 confirmed as described in Example 3. Edited pigs will then be crossbred to create pigs that are homozygous for the edit. PEDV viral challenge of homozygous edited pigs will be performed as described in Whitworth, K.M., et al., Transgenic Research, 28, 21-32. Briefly, homozygous edited pigs will be inoculated with PEDV. Fecal swabs will be collected at intervals beginning prior to inoculation. Realtime PCR to determine the 20 presence of virus in the feces will be performed using standard methods (see, *e.g.* Whitworth, K.M., et al., Transgenic research, 28, 21-32). It is expected that the viral amounts for edited pigs will be significantly lower than control pigs without the edit.

Example 5

This example illustrates an assay for real time PCR verification of desired edits.

25 There will be two assays created for detection of ANPEP edits using real time PCR (rtPCR). For the first assay, two sets of primers and two probes will be designed. One set of primers will flank the spacer sequence of the first guide RNA. A probe, labeled with a fluorescent moiety, will be designed to anneal to the unedited version of the spacer sequence. The other set of primers will be designed to flank the desired edit sequence. A 30 probe, labeled with a different fluorescent moiety, will be designed to anneal to nucleotides spanning the joining region of the edit. For validation, quantitative real time PCR will be performed using a commercial kit using DNA isolated from pigs of confirmed status from

sequencing, and fluorescence will be charted. Validation will occur if, as expected, the homozygotes are close to the y axis (representing the fluorescent moiety for the probe annealing to the desired edit), the heterozygotes group near the center of the chart, and the wild type pigs group close to the X axis (representing the fluorescent moiety for the probe annealing to the spacer sequence).

For the second assay, two sets of primers and two probes will also be designed. One set of primers will flank the second spacer sequence. A probe, labeled with a fluorescent moiety, will be designed to anneal to the unedited version of the spacer sequence. The other set of primers will be designed to flank the desired edit sequence—these may be the same primers and probe used for the first assay. A probe, labeled with a different fluorescent moiety, will be designed to anneal to nucleotides spanning the joining region of the edit. For validation, real time PCR will be performed using a commercial kit using DNA isolated from pigs of confirmed status from sequencing, and fluorescence will be charted. Validation will occur if, as expected, the homozygotes are close to the y axis (representing the fluorescent moiety for the probe annealing to the desired edit), the heterozygotes group near the center of the chart, and the wild type pigs group close to the X axis (representing the fluorescent moiety for the probe annealing to the spacer sequence).

All references, including publications, patents, and patent applications, cited herein are hereby incorporated by reference to the extent they are not inconsistent with the explicit details of this disclosure, and are so incorporated to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein. Examples disclosed herein are provided by way of exemplification and are not intended to limit the scope of the present teachings.

What is claimed is:

1. A livestock animal or progeny thereof with a genome comprising an edited ANPEP gene comprising a premature stop codon.

5 2. The livestock animal according to claim 1, wherein an ANPEP protein encoded by the edited ANPEP gene shows reduced binding to a coronavirus relative to an ANPEP protein encoded by an unedited ANPEP gene.

3. The livestock animal according to claim 1, wherein the premature stop codon is in exon 10 2, 3, 4, 16, or 18.

4. The livestock animal according to claim 1, wherein the livestock animal is resistant to a coronavirus.

15 5. The livestock animal according to claim 1, wherein the livestock animal is a pig.

6. The pig according to claim 5, wherein the pig is resistant to transmissible gastroenteritis virus (TGEV), porcine epidemic diarrhea virus (PEDV), or Porcine respiratory coronavirus (PRCV).

20

7. The pig according to claim 5, wherein the edited ANPEP gene comprises SEQ ID NO: 252, 253, 254, 255, 256, or 257.

8. A cell isolated from the livestock animal according to claim 1.

25

9. An isolated cell line obtained from the livestock animal of claim 1.

10. An isolated fibroblast line obtained from the livestock animal of claim 1.

30 11. A method of producing a gene edited livestock animal comprising:

introducing into an isolated cell:

a Cas9 protein or a nucleic acid encoding a Cas9 protein; and

a pair of gRNAs that each create a double stranded break in an ANPEP gene such that a premature stop codon is formed when the double stranded breaks are repaired.

5 12. The method of claim 11, further comprising producing an animal from the isolated cell.

13. The method according to claim 11, wherein the pair of gRNAs is selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 102 and 10, SEQ ID NOs: 58 and 52, SEQ ID NOs: 58 and 26, SEQ ID NOs: 58 and 10,
10 SEQ ID NOs: 32 and 26, SEQ ID NOs: 32 and 10, SEQ ID NOs: 120 and 111, SEQ ID NOs: 119 and 113, SEQ ID NOs: 115 and 111, SEQ ID NOs 137 and 130, SEQ ID NOs: 133 and 130, SEQ ID NOs: 137 and 128, SEQ ID NOs 133 and 128, SEQ ID NOs: 189 and 183, SEQ ID NOs: 231 and 208, SEQ ID NOs: 228 and 213, SEQ ID NOs: 216 and 213, and SEQ ID NOs: 216 and 208.
15

14. The method according to claim 11, wherein pair of gRNAs is selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 58 and 52, SEQ ID Nos: 120 and 111, SEQ ID NOs: 231 and 208, and SEQ ID NOs: 228 and 213.

20 15. The method according to claim 11, wherein the edited ANPEP gene comprises SEQ ID NO: 252, 253, 254, 255, 256, or 257.

16. The method according to claim 11, wherein the isolated cell is an oocyte, a zygote, or a fibroblast cell.
25

17. The method according to claim 12, wherein the isolated cell is an oocyte.

18. The method according to claim 12, wherein the producing an animal comprises fertilizing the oocyte.
30

19. The method according to claim 18, wherein the producing an animal further comprises implanting the fertilized oocyte into a surrogate mother.

20. The method according to claim 11, wherein the livestock animal is resistant to a coronavirus.

5 21. The method according to claim 11, wherein the livestock animal is a pig.

22. At least one gRNA having a sequence selected from the group consisting of SEQ ID NOs: 10, 26, 32, 52, 58, 102, 111, 113, 115, 119, 120, 128, 130, 133, 137, 183, 189, 208, 213, 216, 228, and 231.

10

23. The at least one gRNA according to claim 22, wherein the at least one gRNA is a pair of gRNAs having sequences selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 102 and 10, SEQ ID NOs: 58 and 52, SEQ ID NOs: 58 and 26, SEQ ID NOs: 58 and 10, SEQ ID NOs: 32 and 26, SEQ ID NOs: 32 and
15 10, SEQ ID NOs: 120 and 111, SEQ ID NOs: 119 and 113, SEQ ID NOs: 115 and 111, SEQ ID NOs 137 and 130, SEQ ID NOs: 133 and 130, SEQ ID NOs: 137 and 128, SEQ ID NOs: 133 and 128, SEQ ID NOs: 189 and 183, SEQ ID NOs: 231 and 208, SEQ ID NOs: 228 and 213, SEQ ID NOs: 216 and 213, and SEQ ID NOs: 216 and 208.

20 24. The at least one gRNA according to claim 22, wherein the at least one gRNA is a pair of gRNAs selected from the group consisting of SEQ ID NOs: 102 and 52, SEQ ID NOs: 102 and 26, SEQ ID NOs: 58 and 52, SEQ ID Nos: 120 and 111, SEQ ID NOs: 231 and 208, and SEQ ID NOs: 228 and 213.