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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 June 2017 (08.06.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/094015 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/85* (2006.01)

(21) International Application Number:
PCT/IL2016/051291

(22) International Filing Date:
1 December 2016 (01.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/262,409 3 December 2015 (03.12.2015) US

(71) Applicant: **EGGXYT LTD** [IL/IL]; Keren HaYessod 36, 9100760 Jerusalem (IL).

(72) Inventor: **OFFEN, Daniel**; 2 Hayeshiva St., 3895500 Kfar HaRoe (IL).

(74) Agent: **ZRIHAN-LICHT, Sheila**; Reinhold Cohn and Partners, P.O.B. 13239, 61131 Tel-Aviv (IL).

(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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2017/094015 A1

(54) Title: METHODS FOR GENDER DETERMINATION OF AVIAN EMBRYOS IN UNHATCHED EGGS AND MEANS THEREOF

(57) Abstract: The present invention relates to methods of gender determination and identification in avian subjects. More specifically, the invention provides non-invasive methods using transgenic avian animals that comprise at least one reporter gene integrated into at least one gender chromosome Z or W. The transgenic avian animals of the invention are used for gender determination and selection of embryos in unhatched avian eggs.

METHODS FOR GENDER DETERMINATION OF AVIAN EMBRYOS IN UNHATCHED EGGS AND MEANS THEREOF

FIELD OF THE INVENTION

The present invention relates to methods of gender determination and identification in avian subjects. More specifically, the invention provides non-invasive methods and transgenic avian animals for gender determination and selection of embryos in unhatched avian eggs.

BACKGROUND ART

References considered to be relevant as background to the presently disclosed subject matter are listed below:

- WO 2010/103111
- WO 2014/0296707
- US 6244214
- 06124456A2
- US2014069336A
- WO16005539
- WO 96/39505
- WO 97/49806
- Quansah, E., Long, J.A., Donovan, D.M., Becker, S.C., Telugu, B., Foster Frey, J.A., Urwin, N. 2014. Sperm-mediated transgenesis in chicken using a PiggyBac transposon system. Poultry Science Association Meeting Abstract. BARC Poster Day.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096), 816-821.
- Cong, L., & Zhang, F. (2015). Genome engineering using CRISPR-Cas9 system. *Chromosomal Mutagenesis,* 197-217.

Acknowledgement of the above references herein is not to be inferred as meaning that these are in any way relevant to the patentability of the presently disclosed subject matter.

BACKGROUND OF THE INVENTION

In the food industry, chicks are culled by billions on a daily basis via suffocation or grinding. The males are terminated since they are not useful for laying eggs or to be bread for meat and the weak or unhealthy females are being terminated as well. A method for in-ovo, or embryo sex-determination prior to hatching is thus highly desired due to both ethical and economic considerations.

Specifically, visually identifying poultry fertile eggs is important to allow removal of unfertile eggs to save hatching costs (by prevention of hatching an unfertile egg), and to lower the bio-security risks involved in the continuation of the incubation of these contamination-prone unfertile eggs alongside the fertile eggs.

Visually identifying egg fertility at an early stage of the embryo, while inside the unhatched egg, involves outer light source candling and can be difficult, virtually impossible in early embryonic stages. An even greater challenge is to identify the sex of embryos, and currently there is no available method for the discrimination between males and females in unhatched eggs that are found fertile. However, identification of fertility at an early embryonic stage and the sex determination of poultry are vital for aviculture, scientific research, and conservation. The determination of sex in young birds by morphological features is extremely challenging for most species. The gender may be determined by individual vent sexing which involves manually squeezing the feces out of the chick, which opens up the chicks' anal vent slightly, allowing to see if the chick has a small "bump", which would indicate that the chick is a male. However, this method represents high risk of bird injury and mistakes in sex determination, together with cumbersome work conducted manually by trained personal.

Vent sexing or chick sexing is the method of distinguishing the sex of chicken and other hatchlings, usually by a trained person called a chick sexer or chicken sexer. Chicken sexing is practiced mostly by large commercial hatcheries, who have to know the difference between the sexes in order to separate them into sex groups, and in order to take them into different programs, which can include the growing of one group and culling of the other group because due to being a sex which does not meet the

commercial needs. (In example, a male hatched from an egg that comes from an egg layer commercial line of breed. That male will not have a good meat yield and will not lay eggs; therefore it will be culled after sexing. After the sexing, the relevant sex will continue its course to serve his or her purpose while the other sex or most of it will be culled within days of hatching being irrelevant to egg production.

In farms that produce eggs, males are unwanted, and chicks of an unwanted sex are killed almost immediately to reduce costs to the breeder. Chicks are moved down a conveyer belt, where chick sexers separate out the males and toss them into a chute where they are usually ground up alive in a meat grinder.

Identification and determination of the fertility of an egg and the sex of the embryos in eggs prior to their hatching, will enable the elimination of unfertile eggs, and the unwanted type of embryos while in their eggs, and thus will immensely reduce incubation costs (which includes the energy and efficiency costs alongside with air pollution and energy consumption). In addition, chicks' suffering will cease and pollution from culling will be prevented. An automated sexing device will additionally result in reduced eggs production costs by eliminating the need for chick sexers, as well as reduce the size of the hatcheries needed since at early stage 50% of the eggs will be reduced deducted from the process, thus reducing the costs of hatching these eggs, and later on the need for any elaborate killing procedures.

In all commercial types of birds intended for breeding, laying, or meat production, there is a need to determine fertility and the sex of the embryo. There are great economic returns; in energy saving, biosecurity risk reduction, garbage disposal, sexing labor costs and sexing errors, culling costs and disposal, and animal welfare.

WO 2010/103111 describes an invasive method comprising a series of steps, among them introducing into the egg a labeled antibody, specifically designed to match a sex-specific antigen on the embryo.

WO 2014/0296707 describes luminance composition designed to serve as a biomarker for quantifying or evaluating efficiency of vaccination being injected into the bird's egg.

No sex determination is described or even hinted in this disclosure. In-ovo injection apparatus and detection methods was disclosed by US 6244214.

WO 06124456A2 discloses invasive methods of in-ovo sex determining of an avian embryo by determining the presence of an estrogenic steroid compound in a sample of embryonic fluid (e.g., allantoic fluid or blood) from the avian egg. Determining the presence of the compound is done by measuring analytes in samples obtained from said avian egg by competitive immunoassay utilizing fluorescence microscopy.

Spectroscopic approaches were also described, among them US2014069336A which is based on screening the avian embryo feather color (pre-hatching) and determining the sex of the avian embryo, based on the feather color or WO16005539 which disclose a device obtaining a shell-specific spectral response to an incident light signal

Further genetic approaches for this problem include DNA sequencing of DNA samples obtained from fertilized eggs for detecting two specific genes located on the Z and W chromosomes of birds (WO 96/39505), or the use of oligonucleotide probes which hybridize to specific sequences of the female W chromosome (WO 97/49806). These methods are invasive and therefore do not provide a safe strategy.

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system is the state of the art gene editing system, allowing a simple construct design with high success rate (J. Doudna, 2012).

Niu et al. (2014) injected guide RNA (gRNA) and Cas9 RNA into monkey oocytes to modify three target genes, and Hwang et al. (2013) modified the drd3 and gsk3b genes in zebrafish embryos to obtain a two-locus mutant. Cong and Zhang (2015) have modified the CRISPR system to edit any gene in living cells.

Veron and coworkers (2015), demonstrated that expression levels of somatic cells in chicken embryos were modified by electroporation of CRISPR gRNA plasmids directed against the PAX7 transcription factor (Nadège et al. 2015), Bai and coworkers edited the PPAR-g, ATP synthase epsilon subunit (ATP5E).

Quansah, E. et. al., disclosed sperm mediated transgenesis in chicken using a PiggyBac transposon system. In particular, they disclose that aGFP plasmid and Lipofectamine LTXTM 9LPX) combination had no effect on viability, mobility or fertility of chicken sperm.

Thus, effective and non-invasive methods for sex identification during the egg stage, prior to the hatching of the chick are currently not available. There is therefore a long-felt need for a method enabling accurate and safe sex identification of the embryos in unhatched eggs.

SUMMARY OF THE INVENTION

A first aspect of the invention relates to a method of gender determination of avian, or avian embryo in an unhatched egg, specifically, a fertilized unhatched egg. In some specific embodiments, the method may comprise the step of:

First, in step (a), providing or obtaining at least one transgenic avian subject or animal comprising at least one exogenous reporter gene integrated into at least one position or location (also referred to herein as locus) in at least one of gender chromosome Z and W. In a second step (b) obtaining at least one fertilized egg from the transgenic avian subject, specifically animal or of any cells thereof.

The next step (c) involves determining in the egg if at least one detectable signal is detected. In more specific embodiments, detection of at least one detectable signal indicates the expression of said at least one reporter gene, thereby the presence of the W chromosome or Z chromosome in the avian embryo.

In a second aspect, the invention relates to an avian transgenic animal comprising, in at least one cell thereof, at least one exogenous reporter gene integrated into at least one position or location (also referred to herein as locus) in at least one of gender chromosome Z and W.

In yet another aspect, the invention relates to a cell comprising at least one exogenous reporter gene integrated into at least one position or locus in at least one of gender chromosome Z and W.

In yet a further aspect, the invention provides a kit comprising:
(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and (b) at least one second nucleic acid sequence comprising at least one said reporter gene.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to better understand the subject matter that is disclosed herein and to exemplify how it may be carried out in practice, embodiments will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Figure 1A-1B. ***Luciferase reporter gene signal penetrates the egg shell***
Luciferase expressing transgenic mice were injected subcutaneously with luciferin. Ear (**Fig. 1A**) and tail (**Fig. 1B**) are excised 10 min thereafter and incorporated into unfertilized eggs. Eggs were imaged using the bio-space photon Imager (Bio space lab, USA).

Figure 2A-2B. ***Luciferase reporter gene signal is formed in a fertilized egg and penetrates the egg shell***
Ear (**Fig. 2A**) and tail (**Fig. 2B**), excised from luciferase expressing transgenic mice, were incorporated into a fertilized carrying a 10 days old chicken embryo. Luciferin is subsequently injected to induce bioluminescence. Images were taken 10 minutes thereafter using the bio-space photon Imager (Bio space lab, USA).

Figure 3A-3B. ***GFP reporter gene signal is not detectable through the egg shell***
Tail from GFP-expressing transgenic mice were incorporated into Chicken embryo (10 days) or placed outside of the shell. Only tail placed outside of the egg shell (Fig. 3A) can be observed with GFP fluorescence, whereas no signal is detected when placed

inside the egg (Fig. 3B) Images were taken after 5 minutes thereafter using the Maestro 2.2 Imager (Cambridge Research & Instrumentation, Inc. USA).

Figure 4. ***Detection of female avian embryo***

Luciferase reporter gene (star) is incorporated into the W chromosome of a female transgenic chicken (hen). Only female egg that carry the W and Z chromosomes, provide the reporter gene, specifically, luciferase signal.

Figure 5. ***Detection of male avian embryo***

Luciferase reporter gene (star) is incorporated into the Z chromosomes of female transgenic chicken (hen). Male embryos are detected via the luciferase signal and females are free of foreign DNA.

DETAILED DESCRIPTION OF THE INVENTION

Each day, billions of male chicks are being terminated via suffocation or grinding since they are not useful for laying eggs or to be bread for meat. The ability to determine the sex of the embryo before hatching is of high importance both ethically and financially.
In the chicken- the genetic make-up of the sex chromosomes is ZZ for males and ZW for females. Meaning the W chromosome determines the gender of the female. This is unlike humans, in which it is the Y from the father that determines the male gender.

The invention provides a non-invasive efficient method for gender determination, using a reporter gene integrated in a gender specific chromosomes of transgenic avian subjects. Expression of this reporter gene in an embryo of an unhatched egg clearly and accurately identify the gender of said embryo.

Thus, a first aspect of the invention relates to a method of gender determination and optionally of selection of avian, or avian embryo in an unhatched egg, specifically, a fertilized unhatched egg. In some specific embodiments, the method may comprise the step of:
First, in step (a), providing or obtaining at least one transgenic avian subject or animal comprising at least one exogenous reporter gene integrated into at least one position or

location in at least one of gender chromosome Z and W. In a second step (b) obtaining at least one fertilized egg from the transgenic avian subject, specifically animal or of any cells thereof.

The next step (c) involves determining in the egg if at least one detectable signal is detected. In more specific embodiments, detection of at least one detectable signal indicates the expression of the at least one reporter gene, thereby the presence of the W chromosome or Z chromosome in the avian embryo. Thus, in case the reporter gene has been integrated into the Z chromosome of a female transgenic avian, identification of a detectable signal in the examined egg indicate that the embryo has a maternal Z chromosome having a reporter gene integrated therein, and the embryo is thereby identified as male. Alternatively, in case the reporter gene has been integrated into the W chromosome of a female transgenic avian, identification of a detectable signal in the examined egg indicate that the embryo carries a maternal W chromosome and is therefore determined as female, thereby providing gender determination thereof.

It should be appreciated that the transgenic avian provided by the invention may be either a female or a male, as described in more detail herein after. In more specific embodiments, where the transgenic avian subject is a female, the egg identified by the method of the invention is laid by the transgenic female avian provided by the invention. In more specific embodiment, the transgenic female may be fertilized either by a transgenic male or by a wild type avian male. Still further, fertilization may occur either by mating or by insemination of the transgenic avian female with sperms obtained from a transgenic or wild type avian male. In yet other embodiments, where the transgenic avian is a male, egg identified by the method of the invention may be laid by either a wild type or transgenic female mated with the transgenic male provided by the invention, or inseminated by any cells thereof, specifically sperm cells that comprise the exogenous reporter gene of the invention integrated into the gender chromosomes thereof.

The invention thus provides a method for detecting a gender of an avian embryo within an unhatched fertilized egg. It should be appreciated that the method of the invention may be applicable for unhatched eggs of any embryonic stage of an avian embryo.

It should be noted that "**Embryonic development stage or step of avian embryo**" , as used herein refers to the stage of **day 1** wherein the germinal disc is at the blastodermal stage and the segmentation cavity takes on the shape of a dark ring; the stage of **day 2** wherein the first groove appears at the center of the blastoderm and the vitelline membrane appears; the stage of **day 3** wherein blood circulation starts, the head and trunk can be discerned, as well as the brain and the cardiac structures which begins to beat; the stage of **day 4** wherein the amniotic cavity is developing to surround the embryo and the allantoic vesicle appears; the stage of **day 5** wherein the embryo takes a C shape and limbs are extending; the stage of **day 6** wherein fingers of the upper and lower limbs becomes distinct; the stage of **day 7** wherein the neck clearly separates the head from the body, the beak is formed and the brain progressively enters the cephalic region; the stage of **day 8** wherein eye pigmentation is readily visible, the wings and legs are differentiated and the external auditory canal is opening; the stage of **day 9** wherein claws appears and the first feather follicles are budding; the stage of **day 10** wherein the nostrils are present, eyelids grow and the egg-tooth appears; the stage of **day 11** wherein the palpebral aperture has an elliptic shape and the embryo has the aspect of a chick; the stage of **day 12** wherein feather follicles surround the external auditory meatus and cover the upper eyelid whereas the lower eyelid covers major part of the cornea; the stage of **day 13** wherein the allantois becomes the chorioallantoic membrane while claws and leg scales becomes apparent; the stage of **days 14 to 16** wherein the whole body grows rapidly, vitellus shrinking accelerates and the egg white progressively disappears; the stage of **day 17** wherein the renal system produces urates, the beak points to the air cell and the egg white is fully resorbed; the stage of **day 18** wherein the vitellus internalized and the amount of amniotic fluid is reduced; the stage of **day 19** wherein vitellus resorption accelerates and the beak is ready to pierce the inner shell membrane; the stage of **day 20** wherein the vitellus is fully resorbed, the umbilicus is closed, the chick pierces the inner shell membrane, breathes in the air cell and is ready to hatch; the stage of **day 21** wherein the chick pierces the shell in a circular way by means of its egg-tooth, extricates itself from the shell in 12 to 18 hours and lets its down dry off.

More specifically, the method of the invention may be applicable in determining the gender of an avian embryo in-ovo, inside the egg, at every stage of the embryonic

developmental process. More specifically, from day 1, from day 2, from day 3, from day 4, from day 5, from day 6, from day 7, from day 8, from day 9, from day 10, from day 11, from day 12, from day 13, from day 14, from day 15, from day 16, from day 17, from day 18, from day 19, from day 20 and from day 21. More specifically, the method of the invention may be applicable for early detection of the embryo's gender, specifically, from day 1 to day 10, more specifically, between days 1 to 5..

As noted above, the method of the invention may be applicable for fertilized unhatched eggs. The term "**fertilized egg**" refers hereinafter to an egg laid by a hen wherein the hen has been mated by a rooster within two weeks, allowing deposit of male sperm into the female infundibulum and fertilization event to occur upon release of the ovum from the ovary. "**Unhatched egg**" as used herein, relates to an egg containing and embryo (also referred to herein as a fertile egg) within a structurally integral (not broken) shell.

The method of the invention is based on determination of a detectable signal formed by a reporter gene integrated into specific loci of the transgenic avian female or male laying the examined egg.

The "**Integration of foreign or exogenous DNA/gene into chromosome**" as used herein, refers hereinafter to a permanent modification of the nucleotide sequence of an organism chromosome. This modification is further transferred during cell division and if occurring in germinal cell lines, it will be transmitted also to offspring. In this case, the integrated reporter gene may be transferred to the embryo within the unhatched egg. The term "**exogenous**" as used herein, refers to originating from outside an organism that has been introduced into an organism for example by transformation or transfection with specifically manipulated vectors, viruses or any other vehicle. The integrated exogenous gene according to certain embodiments, may be a reporter gene. The term "**reporter gene**" relates to gene which encodes a polypeptide, whose expression can be detected in a variety of known assays and wherein the level of the detected signal indicates the presence of said reported.

As noted above, the exogenous reporter gene may be integrated into the avian gender chromosomes Z or W. The avian **"gender chromosome Z or W**" as used herein refers

to the chromosomal system that determines the sex of offspring in chicken wherein males are the homogametic sex (ZZ), while females are the heterogametic sex (ZW). The presence of the W chromosome in the ovum determines the sex of the offspring while the Z chromosome is known to be larger and to possess more genes.

The method of the invention is based on the detection of a detectable signal that indicates and reflects the presence of the reporter gene and thereby the presence of a specific gender chromosome. "**Detectable signal**" refers hereinafter to a change in that is perceptible either by observation or instrumentally. Without limitations, the signal can be detected directly or only in the presence of a reagent. In some embodiments, detectable response is an optical signal including, but are not limited to chemiluminescent groups.

It should be appreciated that in some specific embodiments, at least one transgenic avian subject provided by the method of the invention, may comprise at least two different reporter genes, each reporter gene may be integrated into at least one position or location in one of gender chromosome Z or W. In case of at least two different reporter genes, each of the gender chromosomes may be labeled differently. The evaluation of the detectable signal formed, may indicate the gender of the examined embryo.

In yet some specific embodiments, the reporter gene comprised within the transgenic avian of the invention may be at least one bioluminescence reporter gene. Thus, in some embodiments the expressed polypeptide is a bioluminescence protein and accordingly the assay measures the levels of light emitted from bioluminescent reaction.
The term "**bioluminescence**" refers to the emission of light by biological molecules, such as proteins. Bioluminescence involves a molecular oxygen, an oxygenase, and a luciferase, which acts on a substrate, a luciferin, as will be described in more detail herein after.

In more specific embodiments, the reporter gene may be luciferase. The term "**Luciferase**" refers hereinafter to a class of oxidative enzymes that produce bioluminescence (photon emission). The emitted photon can be detected by light

sensitive apparatus such as a luminometer or modified optical microscopes. Luciferase can be produced through genetic engineering in a variety of organisms mostly for use as a reporter gene. Luciferases occur naturally in bacteria, algae, fungi, jellyfish, insects, shrimp, and squid. In bacteria, the genes responsible for the light-emitting reaction (the ***lux*** genes encoded into the lux operon) have been isolated and used extensively in the construction of bio reporters that emit a blue-green light with a maximum intensity at 490 nm. Three variants of lux are available, one that functions at < 30°C, another at < 37°C, and a third at < 45°C. The *lux* genetic system consists of five genes, *luxA*, *luxB*, *luxC*, *luxD*, and *luxE*. Depending on the combination of these genes used, several different types of bioluminescent bioreporters can be constructed. The luciferase protein is a heterodimer formed by the *luxA* and *luxB* gene products. The *luxC*, *luxD*, and *luxE* gene products encode for a reductase, transferase, and synthase respectively, that work together in a single complex to generate an aldehyde substrate for the bioluminescent reaction. *luxAB* bioreporters contain only the *luxA* and *luxB* genes, which are able to generate the light signal. However, to fully complete the light-emitting reaction, the substrate (long chain aldehyde) must be supplied to the cell.

On the other hand, l*uxCDABE* bioreporters contain all five genes of the lux cassette, thereby allowing for a completely independent light generating system that requires no extraneous additions of substrate nor any excitation by an external light source. Due to their rapidity and ease of use, along with the ability to perform the bioassay repetitively in real time and on-line, makes *luxCDABE* bioreporters extremely attractive. Thus, in certain embodiments, the method of the invention may use as the reporter gene, the *luxCDABE* bioreporters.

In yet some further embodiments, the method of the invention may use as a reporter gene, the *luc* gene. Firefly luciferase (*luc* gene) catalyzes a reaction that produces visible light in the 550-575 nm range. A click-beetle luciferase is also available that produces light at a peak closer to 595 nm. Both luciferases require the addition of an exogenous substrate (luciferin) for the light reaction to occur.

It should be appreciated that any of the luciferases described herein, of any source known in the art, may be applicable for the methods and kits of the invention.

In yet some specific embodiments, the luciferase that may be used by the methods of the invention may be *Gaussia princeps* luciferase. In yet more specific embodiments, the luciferase used by the invention may be the luciferase encoded by the nucleic acid sequence as disclosed by GenBank: AY015993.1, having the amino acid sequence as disclosed by GenBank: AAG54095.1. In yet some further specific embodiments, the luciferase used by the methods and kits of the invention may be encoded by a nucleic acid sequence comprising the sequence as denoted by SEQ ID NO. 22. In yet some further embodiments, such luciferase may comprise the amino acid sequence as denoted by SEQ ID NO. 23, or any homologs, mutants or derivatives thereof.

In yet some further embodiments, luciferase used by the invention may be ***P. pyralis* (firefly) luciferase**. In some specific embodiments such luciferase may be the luciferase encoded by the nucleic acid sequence as disclosed by GenBank: M15077.1, having the amino acid sequence as disclosed by GenBank: AAA29795.1. In yet some further specific embodiments, the luciferase used by the methods and kits of the invention may be encoded by a nucleic acid sequence comprising the sequence as denoted by SEQ ID NO. 20. In yet some further embodiments, such luciferase may comprise the amino acid sequence as denoted by SEQ ID NO. 21, or any homologs, mutants or derivatives thereof.

As noted above, the luciferase used by the method of the invention may require supplementing additional reagents, specifically, a substrate.

Thus, in yet some further embodiments, the method may further comprise the step of providing to said egg of step (b), at least one of substrate and enzyme compatible to the bioluminescence reporter gene. It should be noted that such substrate or enzyme may be required for the formation of the detectable signal detected at step (c). More specifically, the method of the invention may comprise the step of providing to the egg of step (b), for example by injection, a substrate for luciferase. In some specific embodiments, such substrate may be luciferin. Luciferin, as used herein is a generic term for the light-emitting compound found in organisms that generate bioluminescence. Luciferins typically undergo an enzyme-catalyzed oxidation and the resulting excited state intermediate emits light upon decaying to its basal state. In yet

some further embodiments, the substrate luciferin, that is an essential element in formation of said detectable signal, is injected to said egg, specifically, prior to measurement and determination of said signal, as performed in step (c). In some embodiments, the substrate may be injected at day 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or 21 of embryonal development of said avian subject, specifically animal. In yet some further embodiments, the substrate and/or further enzyme required for formation of the detectable signal may be provided to the fertilized egg as nucleic acid sequence encoding said substrate and/or enzyme, operably liked to said reporter gene. Such specific embodiments may refer for example to the use of the *luxCDABE* bioreporters as described above. LuxCDABE system contain five genes of the lux cassette, thereby allowing for a completely independent light generating system that requires no extraneous additions of substrate nor any excitation by an external light source.

In some embodiments, it should be noted that the detectable signal, specifically, the bioluminescent signal may be detected using suitable bioluminescent means. In some embodiments, the detectable signal formed by the luciferase reporter gene may be detected by light sensitive apparatus such as a luminometer or modified optical microscopes or Charge Coupled Device (CCD), a highly sensitive photon detector.

In still further embodiments, the at least one transgenic avian subject or animal provided by the method of the invention may be a female avian subject or animal. In more specific embodiments, where the at least one reporter gene is integrated into at least one position of female chromosome Z, detection of a detectable signal indicates that the embryo in the unhatched egg is male.

In yet some further embodiments, at least one transgenic avian subject or animal provided by the method of the invention may be a female avian subject or animal. In some specific embodiments, where the at least one reporter gene is integrated into at least one position of female chromosome W, detection of a detectable signal, indicates that the embryo in the unhatched egg is female.

In yet some further embodiments, the transgenic animal provided by the method of the invention may be a male subject having the reporter gene integrated into the Z chromosomes thereof. In such case, a detectable signal determined in an egg fertilized by such transgenic male or any sperms thereof, indicates that the embryo carries a paternal Z chromosome comprising the transgenic reporter gene, and is therefore male. In still further embodiments, detection of a detectable signal in an egg laid by a transgenic female avian fertilized by a transgenic male avian, both carrying the reporter gene of the invention integrated into the Z chromosomes thereof, may indicate in case of an intense signal that the embryo carries two copies of a reporter gene integrated into the female and male Z chromosomes thereof. In case of a less intense signal, the egg may be determined as a female.

As indicated herein before, the method of the invention involves the provision of transgenic avian animals. The preparation of transgenic avian animals, requires the use of genetic engineering approach that may use specific nucleases.

Thus, in yet more specific embodiments, the at least one reporter gene may be integrated into the gender chromosome of the transgenic avian subject or animal provided by the method of the invention using at least one programmable engineered nuclease (PEN). The term "**programmable engineered nucleases (PEN)**" as used herein, refers to synthetic enzymes that cut specific DNA sequences, derived from natural occurring nucleases involved in DNA repair of double strand DNA lesions and enabling direct genome editing.

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Type II system is a bacterial immune system that has been modified for genome engineering.
It should be appreciated however that other genome engineering approaches, like zinc finger nucleases (ZFNs) or transcription-activator-like effector nucleases (TALENs) that relay upon the use of customizable DNA-binding protein nucleases that require design and generation of specific nuclease-pair for every genomic target may be also applicable herein.

As used herein, CRISPR arrays also known as SPIDRs (Spacer Interspersed Direct Repeats) constitute a family of recently described DNA loci that are usually specific to a particular bacterial species. The CRISPR array is a distinct class of interspersed short sequence repeats (SSRs) that were first recognized in *E. coli*. In subsequent years, similar CRISPR arrays were found in *Mycobacterium tuberculosis*, *Haloferax mediterranei*, *Methanocaldococcus jannaschii*, *Thermotoga maritima* and other bacteria and archaea. It should be understood that the invention contemplates the use of any of the known CRISPR systems, particularly and of the CRISPR systems disclosed herein. The CRISPR-Cas system has evolved in prokaryotes to protect against phage attack and undesired plasmid replication by targeting foreign DNA or RNA. The CRISPR-Cas system, targets DNA molecules based on short homologous DNA sequences, called spacers that exist between repeats. These spacers guide CRISPR-associated (Cas) proteins to matching (and/or complementary) sequences within the foreign DNA, called proto-spacers, which are subsequently cleaved. The spacers can be rationally designed to target any DNA sequence. Moreover, this recognition element may be designed separately to recognize and target any desired target. With respect to CRISPR systems, as will be recognized by those skilled in the art, the structure of a naturally occurring CRISPR locus includes a number of short repeating sequences generally referred to as "repeats". The repeats occur in clusters and are usually regularly spaced by unique intervening sequences referred to as "spacers." Typically, CRISPR repeats vary from about 24 to 47 base pair (bp) in length and are partially palindromic. The spacers are located between two repeats and typically each spacer has unique sequences that are from about 20 or less to 72 or more bp in length. Thus, in certain embodiments the CRISPR spacers used in the sequence encoding at least one gRNA of the methods and kits of the invention may comprise between 10 to 75 nucleotides (nt) each. More specifically, about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75 or more. In some specific embodiments the spacers comprise about 20 to 25 nucleotides, more specifically, about 20 nucleobases.

In addition to at least one repeat and at least one spacer, a CRISPR locus also includes a leader sequence and optionally, a sequence encoding at least one tracrRNA. The leader

sequence typically is an AT-rich sequence of up to 550 bp directly adjoining the 5' end of the first repeat.

In more specific embodiments, PEN may be a clustered regularly interspaced short palindromic repeat (CRISPR) type II system.

More specifically, three major types of CRISPR–Cas system are delineated: Type I, Type II and Type III.

The type II CRISPR–Cas systems include the `HNH'-type system (Streptococcus-like; also known as the Nmeni subtype, for *Neisseria meningitidis serogroup A str.* Z2491, or CASS4), in which Cas9, a single, very large protein, seems to be sufficient for generating crRNA and cleaving the target DNA, in addition to the ubiquitous Cas1 and Cas2. Cas9 contains at least two nuclease domains, a RuvC-like nuclease domain near the amino terminus and the HNH (or McrA-like) nuclease domain in the middle of the protein, but the function of these domains remains to be elucidated. However, as the HNH nuclease domain is abundant in restriction enzymes and possesses endonuclease activity responsible for target cleavage.

Type II systems cleave the pre-crRNA through an unusual mechanism that involves duplex formation between a tracrRNA and part of the repeat in the pre-crRNA; the first cleavage in the pre-crRNA processing pathway subsequently occurs in this repeat region. Still further, it should be noted that type II system comprise at least one of *cas9, cas1, cas2 csn2, and cas4* genes. It should be appreciated that any type II CRISPR-Cas systems may be applicable in the present invention, specifically, any one of type II-A or B.

Thus, in yet some further and alternative embodiments, the at least one *cas* gene used in the methods and kits of the invention may be at least one *cas* gene of type II CRISPR system (either typeII-A or typeII-B). In more particular embodiments, at least one *cas* gene of type II CRISPR system used by the methods and kits of the invention may be the *cas9* gene. It should be appreciated that such system may further comprise at least one of *cas1, cas2, csn2* and *cas4* genes.

Double-stranded DNA (dsDNA) cleavage by Cas9 is a hallmark of ***"type II CRISPR-Cas"*** immune systems. The CRISPR-associated protein Cas9 is an RNA-guided DNA endonuclease that uses RNA:DNA complementarity to identify target sites for sequence-specific double stranded DNA (dsDNA) cleavage, creating the double strand brakes (DSBs) required for the HDR that results in the integration of the reporter gene into the specific target sequence, for example, a specific target within the avian gender chromosomes W and Z. The targeted DNA sequences are specified by the CRISPR array, which is a series of about 30 to 40 bp spacers separated by short palindromic repeats. The array is transcribed as a pre-crRNA and is processed into shorter crRNAs that associate with the Cas protein complex to target complementary DNA sequences known as proto-spacers. These proto-spacer targets must also have an additional neighboring sequence known as a proto-spacer adjacent motif (PAM) that is required for target recognition. After binding, a Cas protein complex serves as a DNA endonuclease to cut both strands at the target and subsequent DNA degradation occurs via exonuclease activity.

CRISPR type II system as used herein requires the inclusion of two essential components: a "guide" RNA (gRNA) and a non-specific CRISPR-associated endonuclease (Cas9). The gRNA is a short synthetic RNA composed of a "scaffold" sequence necessary for Cas9-binding and about 20 nucleotide long "spacer" or "targeting" sequence which defines the genomic target to be modified. Thus, one can change the genomic target of Cas9 by simply changing the targeting sequence present in the gRNA. Guide RNA (gRNA), as used herein refers to a synthetic fusion of the endogenous bacterial crRNA and tracrRNA, providing both targeting specificity and scaffolding/binding ability for Cas9 nuclease. Also referred to as "single guide RNA" or "sgRNA". CRISPR was originally employed to "knock-out" target genes in various cell types and organisms, but modifications to the Cas9 enzyme have extended the application of CRISPR to "knock-in" target genes, selectively activate or repress target genes, purify specific regions of DNA, and even image DNA in live cells using fluorescence microscopy. Furthermore, the ease of generating gRNAs makes CRISPR one of the most scalable genome editing technologies and has been recently utilized for genome-wide screens.

The target within the genome to be edited, specifically, the specific target loci within the gender chromosomes Z or W, where the reporter gene of the invention is to be integrated, should be present immediately upstream of a Protospacer Adjacent Motif (PAM).

The PAM sequence is absolutely necessary for target binding and the exact sequence is dependent upon the species of Cas9 (5′ NGG 3′ for *Streptococcus pyogenes* Cas9). In certain embodiments, Cas9 from *S. pyogenes* is used in the methods and kits of the invention. Nevertheless, it should be appreciated that any known Cas9 may be applicable. Non-limiting examples for Cas9 useful in the present disclosure include but are not limited to *Streptococcus pyogenes* (SP), also indicated herein as SpCas9, *Staphylococcus aureus* (SA), also indicated herein as SaCas9, *Neisseria meningitidis* (NM), also indicated herein as NmCas9, *Streptococcus thermophilus* (ST), also indicated herein as StCas9 and *Treponema denticola* (TD), also indicated herein as TdCas9. In some specific embodiments, the Cas9 of Streptococcus pyogenes M1 GAS, specifically, the Cas9 of protein id: AAK33936.1, may be applicable in the methods and kits of the invention. In some embodiments, the Cas9 protein may be encoded by the nucleic acid sequence as denoted by SEQ ID NO. 24. In further specific embodiments, the Cas9 protein may comprise the amino acid sequence as denoted by SEQ ID NO. 25, or any derivatives, mutants or variants thereof. Once expressed, the Cas9 protein and the gRNA, form a riboprotein complex through interactions between the gRNA "scaffold" domain and surface-exposed positively-charged grooves on Cas9. Cas9 undergoes a conformational change upon gRNA binding that shifts the molecule from an inactive, non-DNA binding conformation, into an active DNA-binding conformation. Importantly, the "spacer" sequence of the gRNA remains free to interact with target DNA. The Cas9-gRNA complex binds any genomic sequence with a PAM, but the extent to which the gRNA spacer matches the target DNA determines whether Cas9 will cut. Once the Cas9-gRNA complex binds a putative DNA target, a "seed" sequence at the 3′ end of the gRNA targeting sequence begins to anneal to the target DNA. If the seed and target DNA sequences match, the gRNA continues to anneal to the target DNA in a 3′ to 5′ direction.

Cas9 will only cleave the target if sufficient homology exists between the gRNA spacer and target sequences. Still further, the Cas9 nuclease has two functional endonuclease domains: RuvC and HNH. Cas9 undergoes a second conformational change upon target binding that positions the nuclease domains to cleave opposite strands of the target DNA. The end result of Cas9-mediated DNA cleavage is a double strand break (DSB) within the target DNA that occurs about 3 to 4 nucleotides upstream of the PAM sequence.

The resulting DSB may be then repaired by one of two general repair pathways, the efficient but error-prone Non-Homologous End Joining (NHEJ) pathway and the less efficient but high-fidelity Homology Directed Repair (HDR) pathway. In some embodiments, the insertion that results in the specific integration of the reporter gene of the invention to the specific target loci within the gender chromosomes W or Z, is a result of repair of DSBs caused by Cas9. In some specific embodiments, the reporter gene of the invention is integrated, or knocked-in the target loci by HDR.

The term "**Homology directed repair (HDR)**", as used herein refers to a mechanism in cells to repair double strand DNA lesions. The most common form of HDR is homologous recombination. The HDR repair mechanism can only be used by the cell when there is a homologue piece of DNA present in the nucleus, mostly in G2 and S phase of the cell cycle. When the homologue DNA piece is absent, another process called non-homologous end joining (NHEJ) can take place instead. Programmable engineered nucleases (PEN) strategies for genome editing, are based on cell activation of the HDR mechanism following specific double stranded DNA cleavage.

As discussed previously, Cas9 generates double strand breaks (DSBs) through the combined activity of two nuclease domains, RuvC and HNH. The exact amino acid residues within each nuclease domain that are critical for endonuclease activity are known (D10A for HNH and H840A for RuvC in *S. pyogenes* Cas9) and modified versions of the Cas9 enzyme containing only one active catalytic domain (called "Cas9 nickase") have been generated. Cas9 nickases still bind DNA based on gRNA specificity, but nickases are only capable of cutting one of the DNA strands, resulting in a "nick", or single strand break, instead of a DSB. DNA nicks are rapidly repaired by HDR (homology directed repair) using the intact complementary DNA strand as the

template. Thus, two nickases targeting opposite strands are required to generate a DSB within the target DNA (often referred to as a "double nick" or "dual nickase" CRISPR system). This requirement dramatically increases target specificity, since it is unlikely that two off-target nicks will be generated within close enough proximity to cause a DSB. It should be therefore understood, that the invention further encompasses the use of the dual nickase approach to create a double nick-induced DSB for increasing specificity and reducing off-target effects.

Thus, in certain embodiments, the at least one reporter gene may be integrated into the gender chromosome of the transgenic avian subject, specifically animal by homology directed repair (HDR) mediated by at least one CRISPR/CRISPR-associated endonuclease 9 (Cas9) system.

In some further embodiments, the gRNA of the kit of the invention may comprise at least one CRISPR RNA (crRNA) and at least one trans-activating crRNA (tracrRNA).

In some alternative embodiments the kit of the invention may comprise nucleic acid sequence encoding the at least one gRNA. Such nucleic acid sequence may comprise a CRISPR array comprising at least one spacer sequence that targets and is therefore identical to at least one protospacer in a target genomic DNA sequence. It should be note that the nucleic acid sequence further comprises a sequence encoding at least one tracrRNA.

In some embodiments the CRISPR array according to the present disclosure comprises at least one spacer and at least one repeat. In yet another embodiment, the invention further encompasses the option of providing a pre-crRNA that can be processed to several final gRNA products that may target identical or different targets.

In yet some more specific embodiments, the crRNA comprised within the gRNA of the invention may be a single-stranded ribonucleic acid (ssRNA) sequence complementary to a target genomic DNA sequence. In some specific embodiments, the target genomic DNA sequence may be located immediately upstream of a protospacer adjacent motif (PAM) sequence and further.

As indicated herein, the gRNA of the kit of the invention may be complementary, at least in part, to the target genomic DNA. In certain embodiments, "**Complementarity**" refers to a relationship between two structures each following the lock-and-key principle. In nature complementarity is the base principle of DNA replication and transcription as it is a property shared between two DNA or RNA sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position in the sequences will be complementary (e.g., A and T or U, C and G).
As indicated above, the genomic DNA sequence targeted by the gRNA of the kit of the invention is located immediately upstream to a PAM sequence. In some embodiments, such PAM sequence may be of the nucleic acid sequence NGG.

In certain embodiments, the PAM sequence referred to by the invention may comprise N, that is any nucleotide, specifically, any one of Adenine (A), Guanine (G), Cytosine (C) or Thymine (T). In yet some further embodiments the PAM sequence according to the invention is composed of A, G, C, or T and two Guanines.

According to one embodiment, the polynucleotide encoding the gRNA of the invention may comprise at least one spacer and optionally, at least one repeat. In yet some further embodiments, the DNA encoding the gRNA of the invention may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more, specifically, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200 or more spacers. In some embodiments, each spacer is located between two repeats. It should be further understood that the spacers of the nucleic acid sequence encoding the gRNA of the invention may be either identical or different spacers. In more embodiments, these spacers may target either an identical or different target genomic DNA. In yet some other embodiments, such spacer may target at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77,

78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more target genomic DNA sequence. These target sequences may be derived from a single locus or alternatively, from several target loci.

As used herein, the term "**spacer**" refers to a non-repetitive spacer sequence that is designed to target a specific sequence and is located between multiple short direct repeats (i.e., CRISPR repeats) of CRISPR arrays. In some specific embodiments, spacers may comprise between about 15 to about 30 nucleotides, specifically, about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides. More specifically, about 20-25 nucleotides.

The guide or targeting RNA encoded by the CRISPR system of the invention may comprise a CRISPR RNA (crRNA) and a trans activating RNA (tracrRNA). The sequence of the targeting RNA encoded by the CRISPR spacers is not particularly limited, other than by the requirement for it to be directed to (i.e., having a segment that is the same as or complementarity to) a target sequence in avian genomic DNA that is also referred to herein as a "**proto-spacer**". Such proto-spacers comprise nucleic acid sequence having sufficient complementarity to a targeting RNA encoded by the CRISPR spacers comprised within the nucleic acid sequence encoding the gRNA of the methods and kits of the invention.

In some embodiments, a crRNA comprises or consists of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nt of the spacer (targeting) sequence followed by 19-36 nt of repeat sequence. In specific and non-limiting embodiments, the targeting spacer may comprise or consist of a segment that targets any one of the genomic DNA sequence for which representative spacer sequences are indicated herein.

It should be noted that in some specific embodiments, the spacers of the CRISPR system of the invention may encode a targeting guide RNA (gRNA). A "**gRNA**" or "**targeting RNA**" is an RNA that, when transcribed from the portion of the CRISPR system encoding it, comprises at least one segment of RNA sequence that is identical to (with the exception of replacing T for U in the case of RNA) or complementary to (and thus "targets") a DNA sequence in the target genomic DNA. The CRISPR systems of

the present disclosure may optionally encode more than one targeting RNA, and the targeting RNAs be directed to one or more target sequences in the genomic DNA.

Still further, in some embodiments, the at least one reporter gene may be integrated into a gender chromosome of the transgenic avian subject, specifically animal by co-transfecting at least one cell of the avian subject, specifically animal or at least one cell introduced into the avian subject, specifically animal, with: (a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and (b) at least one second nucleic acid sequence comprising at least one reporter gene.

Thus, for the preparation of a transgenic avian animal used by the methods of the invention, at least two nucleic acid molecules should be provided.
As used herein, "**nucleic acids or nucleic acid molecules**" is interchangeable with the term "**polynucleotide(s)**" and it generally refers to any polyribonucleotide or poly-deoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA or any combination thereof. "Nucleic acids" include, without limitation, single- and double-stranded nucleic acids. As used herein, the term "nucleic acid(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. As used herein, the term "**oligonucleotide**" is defined as a molecule comprised of two or more deoxyribonucleotides and/or ribonucleotides, and preferably more than three. Its exact size will depend upon many factors which in turn, depend upon the ultimate function and use of the oligonucleotide. The oligonucleotides may be from about 8 to about 1,000 nucleotides long. More specifically, the oligonucleotide molecule/s used by the kit of the invention may comprise any one of 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more bases in length.

Nucleic acid molecules can be composed of monomers that are naturally-occurring nucleotides (such as DNA and RNA), or analogs of naturally-occurring nucleotides (e.g., alpha-enantiomeric forms of naturally-occurring nucleotides), or modified

nucleotides or any combination thereof. Herein this term also encompasses a cDNA, i.e. complementary or copy DNA produced from an RNA template by the action of reverse transcriptase (RNA-dependent DNA polymerase).

In this connection an "**isolated polynucleotide**" is a nucleic acid molecule that is separated from the genome of an organism. For example, a DNA molecule that encodes the reporter gene used by the methods and kits of the invention or any derivatives or homologs thereof, as well as the sequences encoding the CRISPR/Cas9 and gRNAs of the methods and kits of the invention, that has been separated from the genomic DNA of a cell is an isolated DNA molecule. Another example of an isolated nucleic acid molecule is a chemically-synthesized nucleic acid molecule that is not integrated in the genome of an organism. A nucleic acid molecule that has been isolated from a particular species is smaller than the complete DNA molecule of a chromosome from that species. In some embodiments, the nucleic acid sequences used by the methods and kits of the invention, specifically, nucleic acid sequences comprising sequences encoding the Cas9 and gRNA, or alternatively the reporter gene of the invention, may be provided constructed within a vector. The invention thus further relates to recombinant DNA constructs comprising the polynucleotides of the invention, and optionally, further additional elements such as promoters, regulatory and control elements, translation, expression and other signals, operably linked to the nucleic acid sequence of the invention.

As used herein, the terms "**recombinant DNA**", "**recombinant nucleic acid sequence**" or "**recombinant gene**" refer to a nucleic acid comprising an open reading frame encoding one of the CRISPR system of the invention, specifically, the CRISPR/Cas9 type II, along with the gRNA of the invention that target the Cas9 to specific locus within avian chromosomes Z and/or W. In yet another embodiments, recombinant DNA as used herein further refers to a nucleic acid comprising an open reading frame encoding the reporter gene of the invention, specifically, transgene.

As referred to herein, by the term "**gene**" or "**transgene**" is meant a nucleic acid, either naturally occurring or synthetic, which encodes a protein product. The term "nucleic acid" is intended to mean natural and/or synthetic linear, circular and sequential arrays

of nucleotides and nucleosides, e.g., cDNA, genomic DNA (gDNA), mRNA, and RNA, oligonucleotides, oligonucleosides, and derivatives thereof.

The phrase "**operatively-linked**" is intended to mean attached in a manner which allows for transgene transcription. The term "**encoding**" is intended to mean that the subject nucleic acid may be transcribed and translated into either the desired polypeptide or the subject protein in an appropriate expression system, e.g., when the subject nucleic acid is linked to appropriate control sequences such as promoter and enhancer elements in a suitable vector (e.g., an expression vector) and when the vector is introduced into an appropriate system or cell.

It should be appreciated that in some embodiments, at least one of the first and the second nucleic acid sequences provided and used by the methods and kits of the invention may be constructed and comprised within a vector. "**Vectors**" or "**Vehicles**", as used herein, encompass vectors such as plasmids, phagemides, viruses, integratable DNA fragments, and other vehicles, which enable the integration of DNA fragments into the genome of the host, or alternatively, enable expression of genetic elements that are not integrated. Vectors are typically self-replicating DNA or RNA constructs containing the desired nucleic acid sequences, and operably linked genetic control elements that are recognized in a suitable host cell and effect the translation of the desired spacers. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system. Such system typically includes a transcriptional promoter, transcription enhancers to elevate the level of RNA expression. Vectors usually contain an origin of replication that allows the vector to replicate independently of the host cell. In yet some alternative embodiments, the expression vectors used by the invention may comprise elements necessary for integration of the desired reporter gene of the invention into the avian gender specific chromosomes W and/or Z.

Accordingly, the term "**control and regulatory elements**" includes promoters, terminators and other expression control elements. Such regulatory elements are described in Goeddel; [Goeddel., et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990)]. For instance, any of a wide variety of expression control sequences that control the expression of a DNA

sequence when operatively linked to it may be used in these vectors to express DNA sequences encoding any desired protein using the method of this invention.
A vector may additionally include appropriate restriction sites, antibiotic resistance or other markers for selection of vector-containing cells. Plasmids are the most commonly used form of vector but other forms of vectors which serve an equivalent function and which are, or become, known in the art are suitable for use herein. See, e.g., Pouwels et al., Cloning Vectors: a Laboratory Manual (1985 and supplements), Elsevier, N.Y.; and Rodriquez, et al. (eds.) Vectors: a Survey of Molecular Cloning Vectors and their Uses, Buttersworth, Boston, Mass (1988), which are incorporated herein by reference.

To create the transgenic avian animal used by the methods of the invention, an avian cell comprising the reporter gene integrated into specific loci within the gender chromosomes Z or W thereof must be prepared. Such cell may be prepared by co-transfecting the cell with the first and second nucleic acid sequences provided by the methods and kits of the invention or with any construct comprising the same. "**Transfection**" as used herein is meant the process of inserting genetic material, such as DNA and double stranded RNA, into mammalian cells. The insertion of DNA into a cell enables the expression, or production, of proteins using the cells own machinery. Thus, co-transfection as used herein refers to simultaneous transfection of at least two different nucleic acid molecules or any vector comprising the same to each single cell. Still further, the nucleic acid sequences to be transfected can be transiently expressed for a short period of time, or become incorporated into the genomic DNA, where the change is passed on from cell to cell as it divides.

The invention therefore provides methods for an in-ovo gender determination of an avian embryo in-ovo based on expression of a reporter gene, specifically, luciferase. "**Expression**" generally refers to the process by which gene-encoded information is converted into the structures present and operating in the cell. Therefore, according to the invention "expression" of a reporter gene, specifically, may refer to transcription into a polynucleotide, translation into a protein, or even posttranslational modification of the protein.

In yet some further specific embodiments, the at least one reporter gene in the second nucleic acid sequence may be flanked at 5' and 3' thereof by homologous arms. It should be appreciated that in some embodiments, these arms are required and therefore facilitate HDR of the reporter gene at the integration site.

In more specific embodiments, the reporter gene in the second nucleic acid sequence used by the method of the invention, may be flanked with two arms that are homologous or show homology or identity of about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% to at least one nucleic acid sequence comprised within the target loci within the gender chromosomes Z or W, that serves as the integration site to facilitate specific integration via HDR. In certain embodiments, the target sequence is also referred to herein as at least one "proto-spacer" that is recognized by the "spacer" sequences that are part of the gRNA used by the invention, and provided by the first nucleic acid sequence.

The term "**Homologous arms**", as used herein refers to HDR templates introduced into specific vectors or viruses, used to create specific mutations or insertion of new elements into a gene, that possess a certain amount of homology surrounding the target sequence to be modified (depending on which PEN is used). In yet some further specific embodiments, where CRISPR is used as a PEN, the arms sequences (left, upstream and right, downstream) may comprise between about 10 to 5000 bp, specifically, between about 50 to 1000 bp, between 100 to 500, specifically, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000bp.

In yet some further embodiments, the targeting sequence within the gRNA encoded by the first nucleic acid sequence provided by the methods and kits of the invention, also referred to herein as the "spacer" sequence, exhibits homology or identity of about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% to at least one nucleic acid sequence comprised within the target loci within the gender chromosomes Z or W, referred to herein as the "proto-spacer".

In some embodiments, the at least one reporter gene in the second nucleic acid sequence may be operably linked to any one of a gender specific promoter, an embryonal specific promoter (for example α-Globin Promoter as referred in Mason et al. 1996) and an inducible promoter (for example light-inducible promoters derived from the soybean SSU gene claimed into US 5750385, or derived from parsley chalcone synthase CHS promoter as referred in Weisshaar et al. 1991, or an engineered version of EL222, a bacterial Light-Oxygen-Voltage protein that activates expression when illuminated with blue light cited from Metta-Mena et al. 2014). In yet more specific embodiments, the reporter gene is under the control of an embryonic promoter, thereby limiting the expression of the transgenic reporter gene to the embryonal stage, with no expression in the adult chick. In such embodiment, the reporter transgene is used and expressed only at the embryonal stage, for diagnostic purposes.

More specifically, "**Promoter**" as used herein, refers to a particular region of the DNA that has the ability to control the expression of the gene which is placed downstream. Thus, "**Promoter specific for gender in chicks**" refers hereinafter to a promoter that will activate the expression of a gene, only in a specific chick gender (i.e. male or female). Still further, "**Promoter specific for development in chicks**" refers to a promoter that will activate the expression of a gene, only at specific stages of the chick development.

In some specific embodiments, the at least one reporter gene may be inserted and thereby integrated into at least one non-coding region of the target gender chromosome. Such approach avoids the disruption of genes that may be required for development and maturation of the unhatched embryo.

"**Non-coding region**" as used herein, refers to components of an organism's DNA that do not encode protein sequences. Some noncoding DNA region is transcribed into functional non-coding RNA molecules, other functions of noncoding DNA regions include the transcriptional and translational regulation of protein-coding sequences, scaffold attachment regions, origins of DNA replication, centromeres and telomeres. The hypothesized non-functional portion (or DNA of unknown function) has often been referred to as "junk DNA".

In some specific embodiments, the at least one reporter gene may be integrated into at least one site at gender W chromosome. In more specific embodiments, the specific locus in the W chromosome may be location 1022859-1024215. In some specific embodiments, the target locus may comprise the nucleic acid sequence as denoted by SEQ ID NO. 3.

In more specific embodiments, the at least one gRNA required to target the reporter gene to such specific location within the W chromosome may comprises the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2, these gRNAs are designated herein as gRNA1 and gRNA2, respectively.

In yet some more specific embodiments, the gRNA used by the method of the invention to prepare the transgenic avian female may comprise the nucleic acid sequence as denoted by SEQ ID NO. 1 (gRNA1). In such case, the at least one reporter gene comprised within said second nucleic acid sequence may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, that facilitate the integration thereof to said specific loci in W chromosome, respectively. It should be appreciated that these arms are also referred to herein as left and right arms, respectively.

In yet some alternative embodiments, the gRNA used for preparing the transgenic avian female of the invention may comprise the nucleic acid sequence as denoted by SEQ ID NO. 2 (gRNA2). In such case the at least one reporter gene comprised within the second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively. It should be appreciated that these arms are also referred to herein as left and right arms, respectively.

In yet some further alternative embodiments, the at least one reporter gene used by the method of the invention for preparing the transgenic avian animal, may be integrated into at least one site at gender Z chromosome. In more specific embodiments, the specific loci in the Z chromosome may be any one of regions 9156874-9161874, as

denoted by SEQ ID NO:15, 27764943-27769943, as denoted by SEQ ID NO:16, 42172748-42177748, as denoted by SEQ ID NO:17, 63363656-63368656, as denoted by SEQ ID NO:18 and 78777477-78782477, as denoted by SEQ ID NO:19 of Chromosome Z of female chicken.

In more specific embodiments, the at least one gRNA required to target the reporter gene to such specific location within the Z chromosome may comprises the nucleic acid sequence as denoted by any one of gRNA3: ACAGACCTATGATATGT, as denoted by SEQ ID NO. 11; gRNA4: CGATTATCACTCACAAG, as denoted by SEQ ID NO. 12; gRNA5: CTGGTTAGCATGGGGAC, as denoted by SEQ ID NO. 13; gRNA6: GTAAAGAGTCAGATACA, as denoted by SEQ ID NO. 14.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 41 and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 42 may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA3 of SEQ ID NO: 11. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 43, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 44, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA4 of SEQ ID NO:12. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 45, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 46, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA5 of SEQ ID NO:13. In some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 47, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 48, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA6 of SEQ ID NO:14.

Further non-limiting examples for gRNA sequences suitable for integration into specific loci within the Z chromosome, may include but are not limited to gRNA7 of Z chromosome locus chrZ_42174515_-1, comprising the nucleic acid sequence

GTAATACAGAGCTAAACCAG, as also denoted by SEQ ID NO:26, gRNA8 of Z chromosome locus chrZ_9157091_1, comprising the nucleic acid sequence ACAGACCTATGATATGTGAG, as also denoted by SEQ ID NO:27, gRNA9 of Z chromosome locus chrZ_27767602_-1, comprising the nucleic acid sequence GAGCTTGTGAGTGATAATCG, as also denoted by SEQ ID NO:28, gRNA10 of Z chromosome locus chrZ_78779927_1, comprising the nucleic acid sequence GTAAAGAGTCAGATACACAG, as also denoted by SEQ ID NO: 29, and gRNA11 of Z chromosome locus chrZ_63364946_-1, comprising the nucleic acid sequence CAGTGGGTACTGAAGCTGTG as also denoted by SEQ ID NO: 30.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 31, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 32, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA7 of SEQ ID NO:26. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 33, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 34, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA8 of SEQ ID NO:27. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 35, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 36, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA9 of SEQ ID NO:28. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 37, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 38, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA10 of SEQ ID NO:29. In yet a further embodiment, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 39, and right arm comprising the nucleic acid

sequence as denoted by SEQ ID NO. 40, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA11 of SEQ ID NO:30.

When genetic loci of zygote cells of an avian host, have been targeted and/or transfected with exogenous sequences, specifically, the reporter gene used by the invention, it may be desirable to use such cells to generate transgenic animals. For such a procedure, following the introduction of the targeting construct into the embryonic stem (ES) cells, the cells may be plated onto a feeder layer in an appropriate medium, for example, DMEM supplemented with growth factors and cytokines, fetal bovine serum and antibiotics. The embryonic stem cells may have a single targeted locus (heterozygotic) or both loci targeted (homozygotic). Cells containing the construct may be detected by employing a selective medium and after sufficient time for colonies to grow, colonies may be picked and analyzed for the occurrence of gene targeting. In some specific embodiments, PCR may be applied to verify the integration of the desired exogenous sequences into the target loci, using primers within and outside the construct sequence. Colonies which show gene targeting may then be used for injection into avian embryos. The ES cells can then be trypsinized and the modified cells can be injected through an opening made in the side of the egg. After sealing the eggs, the eggs can be incubated under appropriate conditions until hatching. Newly hatched avian can be tested for the presence of the target construct sequences, for example by examining a biological sample thereof, e.g., a blood sample. After the avian have reached maturity, they are bred and their progeny may be examined to determine whether the exogenous integrated sequences are transmitted through the germ line.

Chimeric avian are generated which are derived in part from the modified embryonic stem cells or zygote cells, capable of transmitting the genetic modifications through the germ line. Mating avian strains containing exogenous sequences, specifically, the reporter gene used by the invention, or portions thereof, with strains in which the avian wild type loci, or portions thereof, is restored, should result in progenies displaying an in-ovo detectable gender.

Still further, transgenic avian can also be produced by other methods, some of which are discussed below. Among the avian cells suitable for transformation for generating

transgenic animals are primordial germ cells (PGC), sperm cells and zygote cells (including embryonic stem cells). Sperm cells can be transformed with DNA constructs by any suitable method, including electroporation, microparticle bombardment, lipofection and the like. The sperm can be used for artificial insemination of avian. Progeny of the inseminated avian can be examined for the exogenous sequence as described above.

Alternatively, primordial germ cells may be isolated from avian eggs, transfected with the exogenous reporter gene of the invention by any appropriate method, and transferred or inserted into new embryos, where they can become incorporated into the developing gonads. Hatched avian and their progeny can be examined for the exogenous reporter gene sequence as described by the invention.

In yet another approach, dispersed blastodermal cells isolated from eggs can be transfected by any appropriate means with the exogenous reporter gene sequence, or portions thereof, integrated to the gender specific chromosomes Z or W, followed by injection into the subgerminal cavity of intact eggs. Hatched avian subjects and their progeny may be examined for the exogenous reporter gene as described above.

Chicken primordial germ cells (PGCs) are the precursors for ova and spermatozoa. Thus, in some aspects thereof, the invention provides the production of transgenic chickens via a germline transmission system using PGCs co-transfected with the reporter gene construct and with the CRISPR/Cas9 gRNA construct that directs the integration of the reporter gene into the gender specific chromosomes W and Z. PGCs are sorted and transferred into the bloodstream of 2.5-day recipient embryos for germline transmission.

Thus, in some specific embodiments, the "**Preparation of transgenic avian animal**" refers to a multi-step method involving genetic engineering techniques for production of chicken with genomic modifications wherein a) Primordial Germ Cells (PGCs) are isolated from the blood of two days-old chick embryos; b) a transgene construct is incorporated into cultured PGCs by using lentiviral system, Piggybac transposon vectors, TALENS or CRISPR/Cas9 techniques; (c) transgenic PGCs are identified and

injected into the circulatory system of embryos and migrate to the developing gonads; d) recipient embryos are incubated at 37°C until hatching (d) hatched males are reared to sexual maturity and crossed with wild-type hens (e) offspring are screened to identify those derived from the transgenic PGCs.

Thus, in a second aspect, the invention relates to an avian transgenic animal comprising, in at least one cell thereof, at least one exogenous reporter gene integrated into at least one position or location (also referred to herein as locus) in at least one of gender chromosome Z and W.

The term "**avian**" relates to any species derived from birds characterized by feathers, toothless beaked jaws, the laying of hard-shelled eggs, a high metabolic rate, a four-chambered heart, and a lightweight but strong skeleton. Avian species includes, without limitation, chicken, quail, turkey, duck, Gallinacea sp, goose, pheasant and other fowl. The term "hen" includes all females of the avian species. A "**transgenic avian**" generally refers to an avian that has had a heterologous DNA sequence, or one or more additional DNA sequences normally endogenous to the avian (collectively referred to herein as "transgenes") chromosomally integrated into the germ cells of the avian. As a result of such transfer and integration, the transferred sequence may be transmitted through germ cells to the offspring of a transgenic avian. The transgenic avian (including its progeny) also have the transgene integrated into the gender chromosomes of somatic cells.

In some specific embodiments, the at least one transgenic animal of the invention may comprise at least two different reporter genes. In such case, each reporter gene may be integrated into at least one position or location in one of gender chromosome Z or W.

In yet some further embodiments, the reporter gene comprised within the transgenic animal of the invention, may be at least one bioluminescence reporter gene.

In more specific embodiments, such bioluminescence reporter gene may comprise or may be luciferase.

In certain embodiments, the at least one transgenic avian animal provided by the invention, may be female. In more specific embodiments, the at least one reporter gene in such transgenic avian female may be integrated into at least one position of the female chromosome Z.

In yet some alternative embodiments, the at least one transgenic avian animal may be female, having at least one reporter gene integrated into at least one position of the female chromosome W.

In some specific embodiments, the at least one reporter gene may be integrated into the gender chromosome of the transgenic animal of the invention using at least one PEN.

More specifically, such PEN may be in certain embodiments, a CRISPR type II system.

In yet more specific embodiments, the at least one reporter gene may be integrated into the gender chromosome of the transgenic avian animal of the invention by HDR mediated by at least one CRISPR/Cas9 system.

In more specific embodiments, the at least one reporter gene may be integrated into a gender chromosome of the transgenic avian animal of the invention by co-transfecting at least one cell of this avian animal, or at least one cell that is to be introduced into said avian animal with at least two nucleic acid sequences. More specifically, such cell may be co-transfected with (a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one gRNA, thereby providing a CRISPR mediated integration; and (b) at least one second nucleic acid sequence comprising at least one reporter gene.

In more specific embodiments, the at least one reporter gene in the second nucleic acid sequence may be flanked at 5' and 3' thereof by homologous arms. These arms exhibit homology to the integration target site within the target gender chromosome, thereby facilitating HDR at the integration site.

In yet more specific embodiments, the at least one reporter gene in the second nucleic acid sequence may be operably linked to any one of a gender specific promoter, an embryonal specific promoter and an inducible promoter. Such promoter should limit the expression of the reporter gene of the invention to the specific desired gender (in case of gender specific promoter), the specific embryonic stage (embryonic specific promoter) or specific conditions (inducible conditions).

In yet some further specific embodiments, the at least one reporter gene comprised within the transgenic avian animal of the invention may be integrated into at least one non-coding region of one of its gender chromosomes.

In certain embodiments, the at least one reporter gene may be integrated into at least one site at gender W chromosome. In some particular embodiments, the integration site may be located at locus 1022859-1024215 at the W chromosome, specifically, galGal5_dna range of chromosome W:1022859-1024215. In yet some further specific embodiments, such loci comprises the nucleic acid sequence as denoted by SEQ ID NO. 3.

For specific integration of the reporter gene of the invention at any position within the loci described above, specific gRNAs may be required. Therefore, in some particular and non-limiting embodiments, appropriate gRNAs used for the preparation of the transgenic avian animal of the invention may comprise the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2. In some specific embodiments, these gRNAs are referred to herein as gRNA1 and gRNA2, respectively.

In some particular embodiments, the transgenic avian animal provided by the invention has been prepared using a gRNA1 that comprises the nucleic acid sequence as denoted by SEQ ID NO. 1. To enable integration of the reporter gene of the invention in such specific location, the reporter gene that should be integrated, must carry in certain embodiments, particular arms facilitating incorporation thereof in the target integration site directed by the gRNA used. Thus, in some specific embodiments, the at least one reporter gene may be comprised within the second nucleic acid sequence, where this

reporter gene is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

In yet some alternative embodiments, the transgenic avian animal provided by the invention may be prepared using a gRNA2 that comprises the nucleic acid sequence as denoted by SEQ ID NO. 2. In such case, to enable integration of the reporter gene of the invention at the specific site recognized by said gRNA2, the at least one reporter gene comprised within the second nucleic acid sequence may be according to specific embodiments, flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

In yet some further alternative embodiments, the transgenic avian animal of the invention may comprise at least one reporter gene integrated into at least one site at gender Z chromosome. In some particular and non-limiting embodiments, such avian transgenic animal may be female that carry the transgenic reporter gene integrated into the Z chromosome. In more specific embodiments, the specific loci in the Z chromosome may be any one of regions 9156874-9161874, as denoted by SEQ ID NO:15, 27764943-27769943, as denoted by SEQ ID NO:16, 42172748-42177748, as denoted by SEQ ID NO:17, 63363656-63368656, as denoted by SEQ ID NO:18 and 78777477-78782477, as denoted by SEQ ID NO:19 of Chromosome Z of female chicken.

In more specific embodiments, the at least one gRNA required to target the reporter gene to such specific location within the Z chromosome may comprises the nucleic acid sequence as denoted by any one of gRNA3: ACAGACCTATGATATGT, as denoted by SEQ ID NO. 11; gRNA4: CGATTATCACTCACAAG, as denoted by SEQ ID NO. 12; gRNA5: CTGGTTAGCATGGGGAC, as denoted by SEQ ID NO. 13 ; gRNA6: GTAAAGAGTCAGATACA, as denoted by SEQ ID NO. 14.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 41, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 42, may be used to integrate the reporter gene of

the invention to the specific loci directed by gRNA3 of SEQ ID NO:11. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 43, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 44, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA4 of SEQ ID NO:12. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 45, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 46, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA5 of SEQ ID NO:13. In some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 47, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 48, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA6 of SEQ ID NO:14.

Further non-limiting examples for gRNA sequences suitable for integration into specific loci within the Z chromosome, may include but are not limited to gRNA7 of Z chromosome locus chrZ_42174515_-1, comprising the nucleic acid sequence GTAATACAGAGCTAAACCAG, as also denoted by SEQ ID NO:26, gRNA8 of Z chromosome locus chrZ_9157091_1, comprising the nucleic acid sequence ACAGACCTATGATATGTGAG, as also denoted by SEQ ID NO:27, gRNA9 of Z chromosome locus chrZ_27767602_-1, comprising the nucleic acid sequence GAGCTTGTGAGTGATAATCG, as also denoted by SEQ ID NO:28, gRNA10 of Z chromosome locus chrZ_78779927_1, comprising the nucleic acid sequence GTAAAGAGTCAGATACACAG, as also denoted by SEQ ID NO: 29, and gRNA11 of Z chromosome locus chrZ_63364946_-1, comprising the nucleic acid sequence CAGTGGGTACTGAAGCTGTG as also denoted by SEQ ID NO: 30.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 31, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 32, may be used to integrate the reporter gene of

the invention to the specific loci directed by gRNA7 of SEQ ID NO:26. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 33, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 34, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA8 of SEQ ID NO:27. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 35, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 36, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA9 of SEQ ID NO:28. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 37, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 38, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA10 of SEQ ID NO:29. In yet a further embodiment, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 39, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 40, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA11 of SEQ ID NO:30.

In yet another aspect, the invention relates to a cell comprising at least one exogenous reporter gene integrated into at least one position or location in at least one of gender chromosome Z and W.

In some specific embodiments, the cell provided by the invention may be an avian cell.

In some particular embodiments, the avian cell provided by the invention may be a primordial germ cell (PGC).

The term "**germ cells**" refers to an embryonic cell that upon uniting with another germ cells develops into a gamete. "**Primordial germ cells (PGCs)**", as used herein relates to germline stem cells that serve as progenitors of the gametes and give rise to

pluripotent embryonic stem cells. The cells in the gastrulating embryo that are signaled to become PGCs during embryogenesis, migrate into the genital ridges which becomes the gonads, and differentiate into mature gametes.

In yet some further embodiments, the cell provided by the invention may comprise at least one reporter gene integrated into a gender chromosome of the cell. In more specific embodiments, such specific integration of the reporter gene may be enabled by co-transfecting the cell with: (a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and (b) at least one second nucleic acid sequence comprising at least one said reporter gene.

In certain embodiments, the at least one reporter gene in the second nucleic acid sequence co-transfected to the cell of the invention, may be flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

In some particular embodiments, to target the integration of the reporter gene to chromosome W in the cell provided by the invention, specific gRNAs should be used. In further particular embodiments, the gRNA may comprise the nucleic acid sequence as denoted by SEQ ID NO. 1 referred to herein as gRNA1. In such case, the at least one reporter gene comprised within the second nucleic acid sequence, may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

In yet some further alternative embodiments, the cell provided by the invention may be prepared by using gRNA referred to herein as gRNA2. In certain embodiments, gRNA2 may comprise the nucleic acid sequence as denoted by SEQ ID NO. 2. In such specific embodiments, the at least one reporter gene comprised within the second nucleic acid sequence may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

In yet some further alternative embodiments, the cell provided by the invention may be prepared by integrating the at least one reporter gene of the invention into the Z

chromosome of the cell. In certain embodiments, for preparing the cell of the invention, the at least one reporter gene may be integrated into at least one site at gender Z chromosome. In more specific embodiments, the specific loci in the Z chromosome may be any one of regions 9156874-9161874, as denoted by SEQ ID NO:15, 27764943-27769943, as denoted by SEQ ID NO:16, 42172748-42177748, as denoted by SEQ ID NO:17, 63363656-63368656, as denoted by SEQ ID NO:18 and 78777477-78782477, as denoted by SEQ ID NO:19 of Chromosome Z of female chicken.

In more specific embodiments, the at least one gRNA required to target the reporter gene to such specific location within the Z chromosome of the cell of the invention may comprises the nucleic acid sequence as denoted by any one of gRNA3: ACAGACCTATGATATGT, as denoted by SEQ ID NO. 11; gRNA4: CGATTATCACTCACAAG, as denoted by SEQ ID NO. 12; gRNA5: CTGGTTAGCATGGGGAC, as denoted by SEQ ID NO. 13; gRNA6: GTAAAGAGTCAGATACA, as denoted by SEQ ID NO. 14.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 41, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 42, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA3 of SEQ ID NO:11. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 43, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 44, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA4 of SEQ ID NO:12. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 45, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 46, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA5 of SEQ ID NO:13. In some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 47, and right arm comprising the

nucleic acid sequence as denoted by SEQ ID NO. 48, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA6 of SEQ ID NO:14. Further non-limiting examples for gRNA sequences suitable for integration into specific loci within the Z chromosome, may include but are not limited to gRNA7 of Z chromosome locus chrZ_42174515_-1, comprising the nucleic acid sequence GTAATACAGAGCTAAACCAG, as also denoted by SEQ ID NO:26, gRNA8 of Z chromosome locus chrZ_9157091_1, comprising the nucleic acid sequence ACAGACCTATGATATGTGAG, as also denoted by SEQ ID NO:27, gRNA9 of Z chromosome locus chrZ_27767602_-1, comprising the nucleic acid sequence GAGCTTGTGAGTGATAATCG, as also denoted by SEQ ID NO:28, gRNA10 of Z chromosome locus chrZ_78779927_1, comprising the nucleic acid sequence GTAAAGAGTCAGATACACAG, as also denoted by SEQ ID NO: 29, and gRNA11 of Z chromosome locus chrZ_63364946_-1, comprising the nucleic acid sequence CAGTGGGTACTGAAGCTGTG as also denoted by SEQ ID NO: 30.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, of the cell of the invention, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 31, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 32, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA7 of SEQ ID NO:26. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 33, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 34, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA8 of SEQ ID NO:27. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 35, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 36, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA9 of SEQ ID NO:28. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 37, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 38,

may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA10 of SEQ ID NO:29. In yet a further embodiment, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 39, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 40, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA11 of SEQ ID NO:30.

In yet a further aspect, the invention provides a kit comprising:
(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and (b) at least one second nucleic acid sequence comprising at least one said reporter gene.

In some embodiments, the at least one reporter gene in the second nucleic acid sequence comprised within the kit of the invention, may be flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

In yet some further specific embodiments, the at least one reporter gene in the second nucleic acid sequence of the kit of the invention may be operably linked to any one of a gender specific promoter, an embryonic specific promoter and an inducible promoter.

In certain embodiments, the at least one reporter gene may be integrated into at least one non-coding region of the gender chromosome, specifically, to chromosome W. In such case, the first nucleic acid sequence of the kit of the invention may encode at least one gRNA comprising the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2, also referred to herein as gRNA1 and gRNA2, respectively.

In some specific embodiments, the first nucleic acid sequence of the kit of the invention may comprise a gRNA, being gRNA1. In some embodiments, such gRNA1 may comprise the nucleic acid sequence as denoted by SEQ ID NO. 1. In such case, the reporter gene comprised within said second nucleic acid sequence of the kit of the

invention, may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

In yet some further alternative embodiments, the kit of the invention may comprise in the first nucleic acid sequence thereof, a sequence encoding gRNA2. In some specific embodiments, such sequence encodes the nucleic acid sequence as denoted by SEQ ID NO. 2. In yet some further embodiments, the least one reporter gene comprised within the second nucleic acid sequence of the kit of the invention , may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

In yet some further alternative embodiments, the at least one reporter gene may be integrated into at least one site at gender Z chromosome. In more specific embodiments, the specific loci in the Z chromosome may be any one of regions 9156874-9161874, as denoted by SEQ ID NO:15, 27764943-27769943, as denoted by SEQ ID NO:16, 42172748-42177748, as denoted by SEQ ID NO:17, 63363656-63368656, as denoted by SEQ ID NO:18 and 78777477-78782477, as denoted by SEQ ID NO:19 of Chromosome Z of female chicken.

Thus, in more specific embodiments, the first nucleic acid sequence of the kit of the invention may comprise a gRNA, being the at least one of gRNA3: ACAGACCTATGATATGT, as denoted by SEQ ID NO. 11; gRNA4: CGATTATCACTCACAAG, as denoted by SEQ ID NO. 12; gRNA5: CTGGTTAGCATGGGGAC, as denoted by SEQ ID NO. 13 ; gRNA6: GTAAAGAGTCAGATACA, as denoted by SEQ ID NO. 14.

In further embodiments, the at least one reporter gene comprised within the second nucleic acid sequence of the kit of the invention, may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by any one of SEQ ID NO. 41-48. More specifically, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 41, and right arm comprising the nucleic acid

sequence as denoted by SEQ ID NO. 42, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA3 of SEQ ID NO:11. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 43, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 44, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA4 of SEQ ID NO:12. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 45, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 46, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA5 of SEQ ID NO:13. In some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 47, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 48, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA6 of SEQ ID NO:14.

Further non-limiting examples of the first nucleic acid sequence of the kit of the invention may comprise a gRNA, being the at least one of gRNA7 of Z chromosome locus chrZ_42174515_-1, comprising the nucleic acid sequence GTAATACAGAGCTAAACCAG, as also denoted by SEQ ID NO:26, gRNA8 of Z chromosome locus chrZ_9157091_1, comprising the nucleic acid sequence ACAGACCTATGATATGTGAG, as also denoted by SEQ ID NO:27, gRNA9 of Z chromosome locus chrZ_27767602_-1, comprising the nucleic acid sequence GAGCTTGTGAGTGATAATCG, as also denoted by SEQ ID NO:28, gRNA10 of Z chromosome locus chrZ_78779927_1, comprising the nucleic acid sequence GTAAAGAGTCAGATACACAG, as also denoted by SEQ ID NO: 29, and gRNA11 of Z chromosome locus chrZ_63364946_-1, comprising the nucleic acid sequence CAGTGGGTACTGAAGCTGTG as also denoted by SEQ ID NO: 30.

In further embodiments, the at least one reporter gene comprised within the second nucleic acid sequence of the kit of the invention, may be flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by any one of SEQ

ID NO. 31 to 40. More specifically, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 31, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 32, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA7 of SEQ ID NO:26. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 33, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 34, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA8 of SEQ ID NO:27. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 35, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 36, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA9 of SEQ ID NO:28. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 37, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 38, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA10 of SEQ ID NO:29. In yet a further embodiment, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 39, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 40, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA11 of SEQ ID NO:30.

In some embodiments, the reporter gene comprised within the second nucleic acid sequence of the kit of the invention may be at least one bioluminescence reporter gene.

In yet some further embodiments, the kit of the invention may be suitable for use in the preparation of a transgenic avian animal comprising at least one exogenous reporter gene integrated into at least one position or location in at least one of gender chromosome Z and W.

In some embodiments, the method of the invention may use any of the kits of the invention as described herein.

Still further, it must be appreciated that the kits of the invention may further comprise any reagent, buffer, media or material required for the preparation of the transgenic avian animals of the invention. The kit of the invention may further comprise instructions as well as containers for the different components thereof.

It should be appreciated that in certain embodiments, the oligonucleotide/s or polynucleotide/s used by the kit/s and method/s of the invention are isolated and/or purified molecules. As used herein, "**isolated**" or "**purified**" when used in reference to a nucleic acid means that a naturally occurring sequence has been removed from its normal cellular (e.g., chromosomal) environment or is synthesized in a non-natural environment (e.g., artificially synthesized). Thus, an "isolated" or "purified" sequence may be in a cell-free solution or placed in a different cellular environment. The term "purified" does not imply that the sequence is the only nucleotide present, but that it is essentially free (about 90-95% pure) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

All scientific and technical terms used herein have meanings commonly used in the art unless otherwise specified. The definitions provided herein are to facilitate understanding of certain terms used frequently herein and are not meant to limit the scope of the present disclosure.

Before specific aspects and embodiments of the invention are described in detail, it is to be understood that this invention is not limited to particular methods, and experimental conditions described, as such methods and conditions may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus for

example, references to "a method" includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

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 this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word “comprise”, and variations such as “comprises” and “comprising”, will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps. More specifically, the terms "comprises", "comprising", "includes", "including", "having" and their conjugates mean "including but not limited to". This term encompasses the terms "consisting of" and "consisting essentially of". The phrase "consisting essentially of" means that the composition or method may include additional ingredients and/or steps, but only if the additional ingredients and/or steps do not materially alter the basic and novel characteristics of the claimed composition or method.

The term "about" as used herein indicates values that may deviate up to 1%, more specifically 5%, more specifically 10%, more specifically 15%, and in some cases up to 20% higher or lower than the value referred to, the deviation range including integer values, and, if applicable, non-integer values as well, constituting a continuous range. As used herein the term "about" refers to ± 10 %.

It should be noted that various embodiments of this invention may be 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 invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub ranges as well as 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 range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range. Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases "ranging/ranges between" a first indicate number and a second indicate number and "ranging/ranges from" a first indicate number "to" a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals there between.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the methods and compositions of the invention, and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

The examples are representative of techniques employed by the inventors in carrying out aspects of the present invention. It should be appreciated that while these techniques are exemplary of preferred embodiments for the practice of the invention, those of skill in the art, in light of the present disclosure, will recognize that numerous modifications can be made without departing from the spirit and intended scope of the invention.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub combination or as suitable in any other described embodiment of the invention. Certain features described in the context of various embodiments are not to be considered essential features of those embodiments, unless the embodiment is inoperative without those elements.

Various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below find experimental support in the following examples.

Disclosed and described, it is to be understood that this invention is not limited to the particular examples, methods steps, and compositions disclosed herein as such methods steps and compositions may vary somewhat. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only and not intended to be limiting since the scope of the present invention will be limited only by the appended claims and equivalents thereof.
It must be noted that, as used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise.

EXAMPLES

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the claimed invention in any way.

Standard molecular biology protocols known in the art not specifically described herein are generally followed essentially as in Sambrook et al., Molecular cloning: A laboratory manual, Cold Springs Harbor Laboratory, New-York (1989,1992), and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, Maryland (1988).

Reagents

Animals:

Commercial White Leghorn chickens are obtained from Hendrix ISA, and Minnesota. Marker Line chickens are from the Pacific Agri-Food Research Centre, Agassiz, British Columbia, Canada.

Transgenic mice CAG-luc-eGFP are from The Jackson Laboratory (catalogue number L2G85).

Transgenic mice C57BL/6-Tg(CAG-EGFP)1Osb/J are from The Jackson Laboratory (catalogue number 00329).

Animal experiments were done in strict accordance to IACUC approved protocols and under supervision of the Crystal Bioscience IACUC committee ensuring that no animal suffers from illness nor dies during the course of the experiments.

Vectors:

Cas9 SmartNuclease™ All-in-one tagged vectors is ordered from System Bioscience Inc., catalogue number CAS8/9xx series.

pCMV-Gluc 2 vector is ordered from New England Biolabs Inc., catalogue number N8081S.

Cell lines:

Female cells are of *Gallus gallus*, chicken T lymphocyte cells.

Male cells are of *Gallus gallus, chicken Liver*, ordered from ATCC® Number: CRL-2118™.

Primordial Germ Cell (PGC) line.

Experimental procedures

Reporter gene bioluminescence detection through the egg shell

D-Luciferin (Sigma-Aldrich Co. LLC, Israel, catalogue number 2591-17-5) is dissolved at room temperature in DPBS to a final concentration of 15 mg/mL.

An amount of 0.1 ml of luciferin or saline solution (negative control) is injected subcutaneously in the loose skin around the neck and shoulder area of transgenic luciferase-expressing mice. Ear and tail are excised after 10 min and introduced into Chicken embryo (10 days).

Alternatively, excised ear and tail from transgenic luciferase-expressing transgenic mice is incorporated into the Chicken embryo prior to direct injection to the egg of 0.1 ml of luciferin or saline solution.

Bioluminescence is observed by using Bio-space photon Imager (Bio space lab, USA).

Restriction-free (RF) cloning

The insertion of gRNAs into Cas9-SmartNuclease™ vector is performed by applying the Restriction Free method (Peleg Y et al., 2010). Primers are ordered from Sigma-Genosys (Rehovot, Israel) and subsequent RF reactions were carried out using Phusion polymerase (Thermo Scientific, Hudson, NH, USA). Plasmid purification is carried out using the MEGAspin kit and DNA-spin plasmid DNA purification kit, respectively (Intron Biotechnology Biotechnology, Daejoen, South Korea).

Cell culture

PGCs are grown in KO-DMEM (Life Technologies), of which 40% is preconditioned on buffalo rat liver cells (BRL, ATCC), and supplemented with 7.5% fetal bovine serum (Hyclone), 2.5% irradiated chicken serum, 1X non-essential amino acids, 2mM glutamine, 1mM sodium pyruvate, 0.1mMβ-mercaptoethanol (all from Life Technologies), 4ng/ml recombinant human fibroblast growth factor, 6ng/ml recombinant mouse stem cell factor (both from R&D Systems) and grow on an irradiated feeder layer of BRL cells. The cells are passaged 3 times per week onto fresh feeder layers.

Transfection

For stable transfectant targeting of the above-mentioned loci of chromosomes W or Z, 15 µg of vector-containing gRNA and 15 µg of circular luciferase-containing vector were added to $5x10^6$ cells and brought to volume of 100µl with V-buffer (lonza, Walkersville). The cell suspension is transferred to a 2 mm cuvette and subjected to 8 square wave pulses of 350 volts/100µsec (BTX 830 electroporator). Cells are then plated with Neomycin-resistant irradiated BRLs and seeded in a 48-well plate at a density of 10^5 cells per well. After 3 days, 40µg/ml Neomycin is added to select for cells with a stable integration of luciferase reporter gene.

Preparation of transgenic chickens

Concentrated vehicle (that may be either lentivirus at a titer of about 10^7 MOI) or plasmid DNA) is injected to 25 embryos in new laid eggs. Injections are carried out weekly three injections. The injected embryos hatch 3 weeks after injection. These are G0 birds. Immediately after hatch, the DNA is extracted from CAM samples of the hatched chicks and detection of the presence/absence of vector DNA is carried out by semi-quantitative PCR. Blood sample G0 chicks at 2-3 weeks of age and repeat PCR screen. G0 birds are raised to sexual maturity, 16-20 weeks for males, 20-24 weeks for females. Cockerels are tested for semen production from approximately 16 weeks.

Hens are inseminated, fertile eggs collected daily. The G1 chicks are hatch 3 weeks later and each individual chick wing banded and a chick chorioallantoic membrane (CAM) sample taken from the shell. Extract DNA from CAM samples and carry out PCR screen for presence of transgene, predicted to be single copy level. Repeat screen to confirm and sex chicks on DNA from blood sample 2-3 weeks later.

At a few weeks of age a blood sample is taken from G1 birds to prepare genomic DNA for PCR analysis. G1 birds are used for breeding G2.

Example 1

Selection of reporter gene for visual gender identification in poultry

In order to demonstrate the feasibility of visually identify gender of in-ovo poultry, the use of bioluminescent as compared to fluorescent reporter genes was evaluated. Therefore, transgenic mice expressing reporter genes such as firefly luciferase (having a nucleic acid sequence as denoted by SEQ ID NO:20; encoding the amino acid sequence as denoted by SEQ ID NO:21)and green fluorescent protein (eGFP), were first employed.

For observation of luciferase activity, in **Figure 1** luciferin was injected subcutaneously to luciferase-expressing transgenic mice, tails and ears were then excised and introduced through a 5mm hole in the egg shell of an unfertilized egg. As shown in the figure, the luciferase detectable signal is clearly observed in tail and ear samples (Fig. 1A, 1B) through the egg shell. The inventors therefore next examined the feasibility of inducing luciferase reaction in-ovo. Therefore, ears and tails of luciferase-expressing

transgenic mice were excised, introduced through a hole into a fertilized egg that carry a 10-days old chicken embryo and luciferin was subsequently injected. As clearly shown in Figures 2A and 2B, an in-ovo luciferase reaction successfully resulted in a detectable signal that was able to penetrate the egg shell.

On the other hand, similar experiments performed using GFP as the reporter gene, clearly indicated that GFP signal is not detectable following incorporation of tails and ears of GFP-expressing transgenic mice into Chicken embryo as seen in **Figure 3**.

Luciferase reporter gene, specifically, firefly luciferase (comprising the amino acid sequence as denoted by SEQ ID NO. 21, encoded by the nucleic acid sequence as denoted by SEQ ID NO:20) was thus further selected for incorporation into sex chromosomes W and Z.

Figure 4 represents a schematic illustration of the method of the invention for identification of embryo's gender in-ovo. More specifically, a transgenic avian female hen containing a gender specific chromosome (W) with the luciferase reporter gene integrated therein is provided. In eggs laid by said hen, expression of the reporter gene observed by a detectable signal indicates that the embryo carry the W gender chromosome and is therefore female. This enables the selection for continued incubation of male while females that carry the reporter gene are discarded. This selection is probably more relevant for Poultry.

Figure 5 schematically presents yet a further alternative that facilitates determination of male embryo, in-ovo. More specifically, the provision of transgenic female chickens carrying the gender specific Z chromosome with a reporter gene integrated therein, results in female embryos (that received the maternal wild type W chromosome) without reporter gene or male embryos (that received the maternal labeled Z chromosome) expressing the transgenic luciferase gene.

Example 2

Design of guide RNAs vector

In order to incorporate the luciferase reporter gene into the gender chromosomes W or Z, the CRISPR/Cas9 mediated HDR method is selected. Relevant gRNA sites are then sought from both gender chromosomes.

The region 1022859-1024215 of Chromosome W of female chicken, comprising the nucleic acid sequence as denoted by SEQ ID NO. 3, is analyzed for guide RNA design. Two guide RNAs are selected, synthesized and cloned separately into the Cas9 SmartNuclease vector containing the wild type Cas9 nuclease (Horizon) by Restriction free cloning protocol: gRNA1: GCACTAGGAACCAGCAGCAG, as denoted by SEQ ID NO. 1 and gRNA2: GTAGCCCCAAGAGGGCTAGG, as denoted by SEQ ID NO. 2.

The predicted parameters of these two gRNAs are presented in Table 1:

Table 1

gRNA parameters	gRNA1	gRNA2
sgRNA designer	0.506	0.63
sscore	0.8677	0.5323
sgRNA scorer	94.8	99.9

The regions 9156874-9161874, as denoted by SEQ ID NO:15, 27764943-27769943, , as denoted by SEQ ID NO:16, 42172748-42177748, as denoted by SEQ ID NO:17, 63363656-63368656, as denoted by SEQ ID NO:18 and 78777477-78782477, as denoted by SEQ ID NO:19 of Chromosome Z of female chicken are analyzed for guide RNA design. Four guide RNAs are selected, synthesized and cloned separately into the Cas9 SmartNuclease vector containing the wild type Cas9 nuclease (Horizon) by Restriction free cloning protocol: gRNA3: ACAGACCTATGATATGT, as denoted by

SEQ ID NO. 11; gRNA4: CGATTATCACTCACAAG, as denoted by SEQ ID NO. 12; gRNA5: CTGGTTAGCATGGGGAC, as denoted by SEQ ID NO. 13 ; gRNA6: GTAAAGAGTCAGATACA, as denoted by SEQ ID NO. 14.

Further non-limiting examples for gRNA sequences suitable for integration into specific loci within the Z chromosome, may include but are not limited to gRNA7 of Z chromosome locus chrZ_42174515_-1, comprising the nucleic acid sequence GTAATACAGAGCTAAACCAG, as also denoted by SEQ ID NO:26, gRNA8 of Z chromosome locus chrZ_9157091_1, comprising the nucleic acid sequence ACAGACCTATGATATGTGAG, as also denoted by SEQ ID NO:27, gRNA9 of Z chromosome locus chrZ_27767602_-1, comprising the nucleic acid sequence GAGCTTGTGAGTGATAATCG, as also denoted by SEQ ID NO:28, gRNA10 of Z chromosome locus chrZ_78779927_1, comprising the nucleic acid sequence GTAAAGAGTCAGATACACAG, as also denoted by SEQ ID NO: 29, and gRNA11 of Z chromosome locus chrZ_63364946_-1, comprising the nucleic acid sequence CAGTGGGTACTGAAGCTGTG as also denoted by SEQ ID NO: 30.

These gRNAs have few predicted off-target sites, none of which were in known coding sequences.

Example 3

Design of Luciferase targeting vector

Flanking sequences homological of the appropriate flanking sequences indicated above of female W chromosome or of the female Z chromosome loci, are introduced into the luciferase-expressing vector upstream to the CMV-promoter and downstream the Neomycin-resistance or alternatively downstream the polyA site (ordered synthetic DNA, Integrated DNA Technologies, Inc., USA).

For the female W chromosome, the reporter gene, specifically Luciferase may be cloned for using either the Guide 1 (gRNA1), as denoted by SEQ ID NO. 1 or Guide 2 (gRNA2): as denoted by SEQ ID NO. 2. For cloning using the gRNA1, "Left arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 4, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 5 are provided. For cloning using the gRNA2, "Left arm" comprising the nucleic acid sequence as

denoted by SEQ ID NO. 6, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 7 are provided.

Still further, a "left arm" for the region upstream to the CMV-promoter comprises the nucleic acid sequence as denoted by SEQ ID NO. 8, and a "right arm" for the region downstream the Neomycin-resistance, may comprise the nucleic acid sequence as denoted by SEQ ID NO. 9, or SEQ ID NO.10 for the region downstream the polyA site.

For the female Z chromosome, the Luciferase reporter gene may be cloned for using either the gRNA3, as denoted by SEQ ID NO. 11, gRNA4 : as denoted by SEQ ID NO. 12, gRNA5, as denoted by SEQ ID NO. 13, gRNA6, as denoted by SEQ ID NO. 14.

For cloning using the gRNA3, "Left arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 41, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 42 are provided. For cloning using the gRNA4, "Left arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 43, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 44 are provided. For cloning using the gRNA5, "Left arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 45, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 46 are provided. For cloning using the gRNA6, "Left arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 47, and the "Right arm" comprising the nucleic acid sequence as denoted by SEQ ID NO. 48 are provided.

In yet some further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 31, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 32, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA7 of SEQ ID NO:26. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 33, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 34, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA8 of SEQ ID NO:27. In still further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 35, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 36, may be used to

integrate the reporter gene of the invention to the specific loci directed by gRNA9 of SEQ ID NO:28. In further embodiments, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 37, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 38, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA10 of SEQ ID NO:29. In yet a further embodiment, for integrating the reporter gene of the invention into the specific locus within the Z chromosome, left arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 39, and right arm comprising the nucleic acid sequence as denoted by SEQ ID NO. 40, may be used to integrate the reporter gene of the invention to the specific loci directed by gRNA11 of SEQ ID NO:30.

Example 4

Germline transmission of CRISPR-treated cells

The two above described vectors, specifically, the gRNA/Cas9 and the reporter-gene vectors are co-transfected to PGCs as detailed in experimental procedures. After stable clones are identified, the cells are expanded and confirmed for the luciferase integration by PCR. Confirmed clones are injected into recipient chicken embryos at Stage 14–16 (H&H). The injected embryos are transferred to surrogate shells and incubated until hatch at 37°C. The sex of the chicks is determined after hatch by PCR for the W-chromosome.

Female and Male chimeras are grown to sexual maturity and bred to wild type male and female chickens. Hatched chicks are evaluated for the expression of luciferase, and the germline progeny are confirmed by PCR to carry targeted luciferase.

CLAIMS:

1. A method of gender determination of avian fertilized unhatched egg, the method comprising the step of:

(a) providing/obtaining at least one transgenic avian animal comprising at least one exogenous reporter gene integrated into at least one position or location in at least one of gender chromosome Z and W;

(b) obtaining at least one fertilized egg from said transgenic avian subject, specifically animal or of any cells thereof;

(c) determining in said egg if at least one detectable signal is detected, wherein detection of said at least one detectable signal indicates the expression of said at least one reporter gene, thereby the presence of said W chromosome or Z chromosome in said avian embryo.

2. The method according to claim 1, wherein said at least one transgenic avian subject comprises at least two different reporter genes, each reporter gene is integrated into at least one locus in one of gender chromosome Z or W.

3. The method according to any one of claims 1 and claim 2, wherein said reporter gene is at least one bioluminescence reporter gene.

4. The method according to claim 3, wherein said reporter gene is luciferase.

5. The method according to claims 1, 3 and 4, wherein said method further comprises the step of providing to said egg of step (b), at least one of substrate and enzyme compatible to said bioluminescence reporter gene, for formation of said detectable signal detected at step (c).

6. The method according to claim 1, wherein said at least one transgenic avian subject, specifically animal is a female avian animal, and wherein said at least one reporter gene is integrated into at least one position of female chromosome Z, thereby detection of a detectable signal indicates that said embryo in said unhatched egg is male.

7. The method according to claim 1, wherein said at least one transgenic avian subject, specifically animal is a female avian animal, and wherein said at least one reporter gene is integrated into at least one position of female chromosome W, thereby detection of a detectable signal, indicates that said embryo in said unhatched egg is female.

8. The method according to claim 1, wherein said at least one reporter gene is integrated into said gender chromosome of said transgenic avian subject, specifically animal using at least one programmable engineered nuclease (PEN).

9. The method according to claim 8, wherein said PEN is a clustered regularly interspaced short palindromic repeat (CRISPR) type II system.

10. The method according to claim 9, wherein said at least one reporter gene is integrated into said gender chromosome of said transgenic avian animal by homology directed repair (HDR) mediated by at least one CRISPR/CRISPR-associated endonuclease 9 (Cas9) system.

11. The method according to claim 10, wherein said at least one reporter gene is integrated into a gender chromosome of said transgenic avian animal by co-transfecting at least one cell of said avian animal or at least one cell introduced into said avian animal, with:

(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and

(b) at least one second nucleic acid sequence comprising at least one said reporter gene.

12. The method according to claim 11, wherein said at least one reporter gene in said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

13. The method according to any one of claims 11 and 12, wherein said at least one reporter gene in said second nucleic acid sequence is operably linked to any one of a gender specific promoter, an embryonal specific promoter and an inducible promoter.

14. The method according to any one of claims 11 to 13, wherein said at least one reporter gene is integrated into at least one non-coding region of said gender chromosome.

15. The method according to claim 14, wherein said at least one reporter gene is integrated into at least one site at gender W chromosome locus 1022859-1024215.

16. The method according to claims 11 and 15, wherein said at least one gRNA comprises the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2.

17. The method according to claim 16, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 1, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

18. The method according to claim 16, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 2, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

19. An avian transgenic animal comprising at least one exogenous reporter gene integrated into at least one locus in at least one of gender chromosome Z and W.

20. The transgenic animal according to claim 19, wherein said at least one transgenic animal comprises at least two different reporter genes, each reporter gene is integrated into at least one locus in one of gender chromosome Z or W.

21. The transgenic animal according to any one of claims 19 and claim 20, wherein said reporter gene is at least one bioluminescence reporter gene.

22. The transgenic animal according to claim 21, wherein said reporter gene is luciferase.

23. The transgenic animal according to claim 19, wherein said at least one transgenic avian animal is female, and wherein said at least one reporter gene is integrated into at least one position of female chromosome Z.

24. The transgenic animal according to claim 19, wherein said at least one transgenic avian animal is female, and wherein said at least one reporter gene is integrated into at least one position of female chromosome W.

25. The transgenic animal according to claim 19, wherein said at least one reporter gene is integrated into said gender chromosome of said transgenic animal using at least one PEN.

26. The transgenic animal according to claim 25, wherein said PEN is a CRISPR type II system.

27. The transgenic animal according to claim 26, wherein said at least one reporter gene is integrated into said gender chromosome of said transgenic avian animal by HDR mediated by at least one CRISPR/Cas9 system.

28. The transgenic animal according to claim 27, wherein said at least one reporter gene is integrated into a gender chromosome of said transgenic avian subject, specifically animal by co-transfecting at least one cell of said avian animal or at least one cell introduced into said avian animal:

(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one gRNA; and

(b) at least one second nucleic acid sequence comprising at least one said reporter gene.

29. The transgenic animal according to claim 28, wherein at least one reporter gene in said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

30. The transgenic animal according to any one of claims 28 and 29, wherein at least one reporter gene in said second nucleic acid sequence is operably linked to any one of a gender specific promoter, an embryonal specific promoter and an inducible promoter.

31. The transgenic animal according to any one of claims 28 to 30, wherein said at least one reporter gene is integrated into at least one non-coding region of said gender chromosome.

32. The transgenic animal according to claim 31, said at least one reporter gene is integrated into at least one site at gender W chromosome locus 1022859-1024215.

33. The transgenic animal according to claims 28 and 32, wherein said at least one gRNA comprises the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2.

34. The transgenic animal according to claim 33, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 1, and wherein at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

35. The transgenic animal according to claim 33, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 2, and wherein at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

36. A cell comprising at least one exogenous reporter gene integrated into at least one locus in at least one of gender chromosome Z and W.

37. The cell according to claim 36, wherein said cell is an avian cell.

38. The cell according to claim 37, wherein said avian cell is a primordial germ cell (PGC).

39. The cell according to any one of claims 36 to 38, wherein said at least one reporter gene is integrated into a gender chromosome of said cell, by co-transfecting said cell with:

(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and

(b) at least one second nucleic acid sequence comprising at least one said reporter gene

40. The cell according to claim 39, wherein said at least one reporter gene in said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

41. The cell according to any one of claims 39 and 40, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 1, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

42. The cell according to any one of claims 39 and 40, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 2, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

43. A kit comprising:

(a) at least one first nucleic acid sequence comprising at least one nucleic acid sequence encoding at least one Cas9 protein and at least one nucleic acid sequence encoding at least one guide RNA (gRNA); and

(b) at least one second nucleic acid sequence comprising at least one said reporter gene.

44. The kit according to claim 43, wherein said at least one reporter gene in said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms for HDR at the integration site.

45. The kit according to any one of claims 43 and 44, wherein said at least one reporter gene in said second nucleic acid sequence is operably linked to any one of a gender specific promoter, an embryonic specific promoter and an inducible promoter.

46. The kit according to any one of claims 43 to 45, wherein said at least one reporter gene is integrated into at least one non-coding region of said gender chromosome, and wherein said at least one gRNA comprises the nucleic acid sequence as denoted by any one of SEQ ID NO. 1 and 2.

47. The kit according to claim 46, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 1, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 4 and 5, respectively.

48. The kit according to claim 46, wherein said gRNA comprises the nucleic acid sequence as denoted by SEQ ID NO. 2, and wherein said at least one reporter gene comprised within said second nucleic acid sequence is flanked at 5' and 3' thereof by homologous arms comprising the amino acid sequence as denoted by SEQ ID NO. 6 and 7, respectively.

49. The kit according to any one of claims 43 to 48, wherein said reporter gene is at least one bioluminescence reporter gene.

50. The kit according to any one of claims 43 to 49, for use in the preparation of a transgenic one transgenic avian animal comprising at least one exogenous reporter gene integrated into at least one locus in at least one of gender chromosome Z and W.

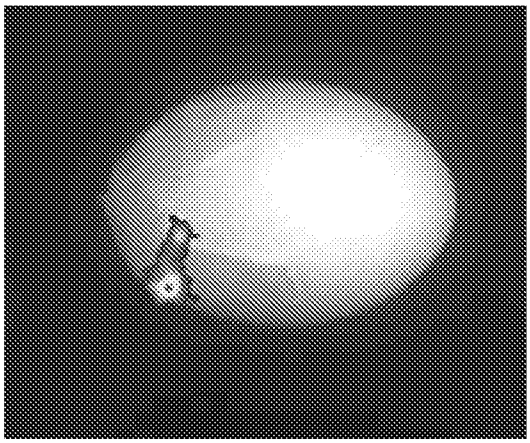


Fig. 1A

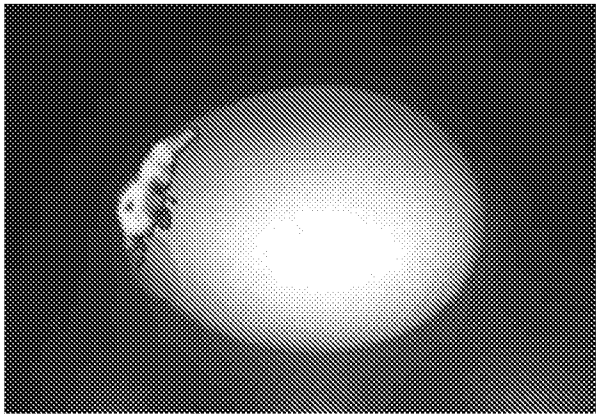


Fig. 1B

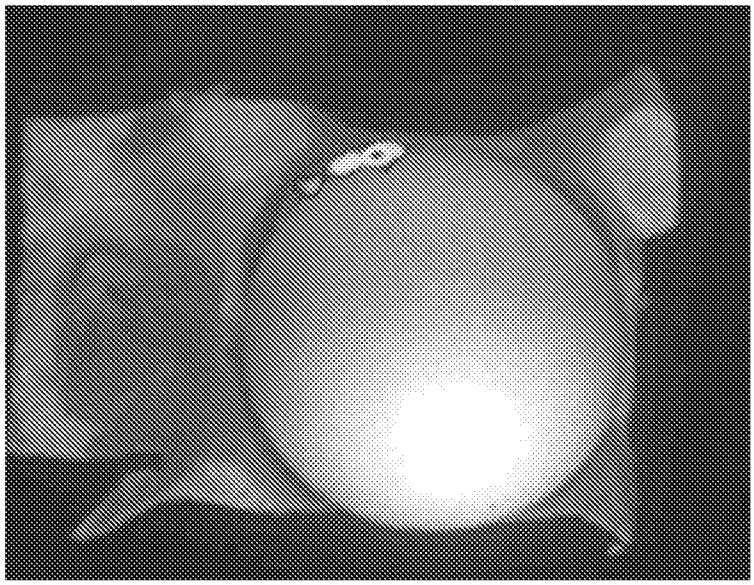


Fig. 2A

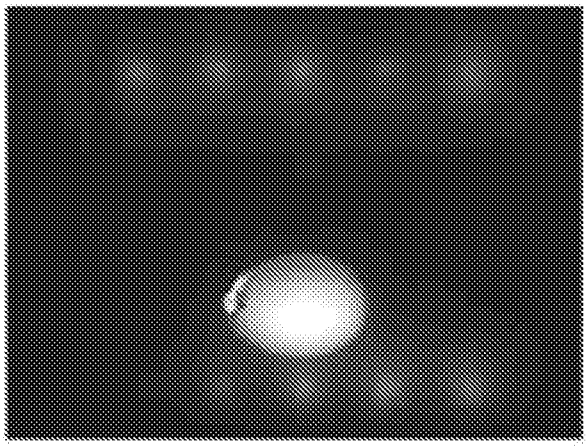


Fig. 2B



Fig. 3A

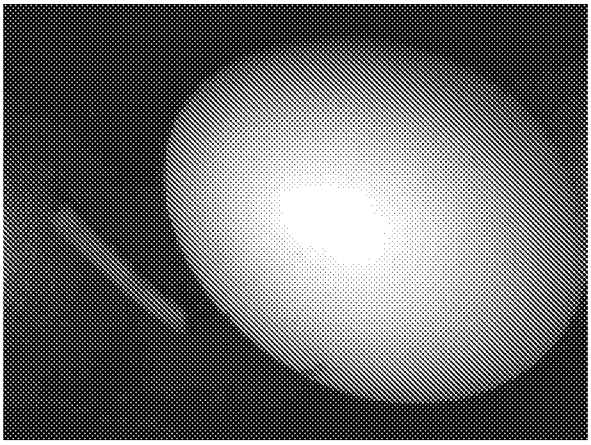


Fig. 3B

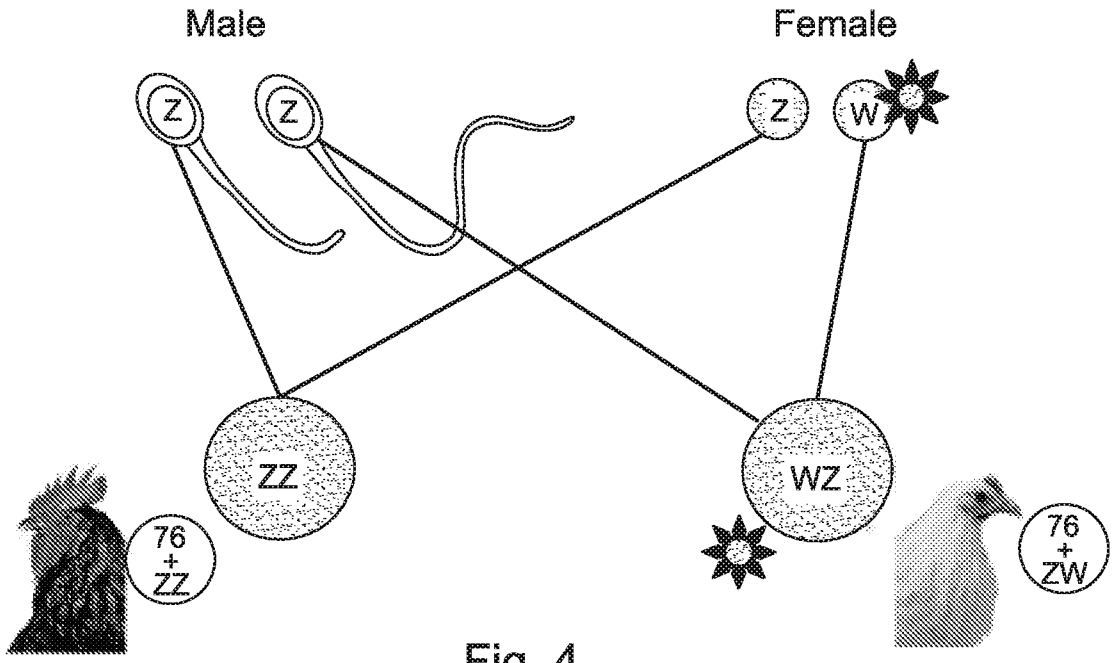

Male
Female
Z
Z
Z
W
ZZ
WZ
76
+
ZZ
76
+
ZW

Fig. 4

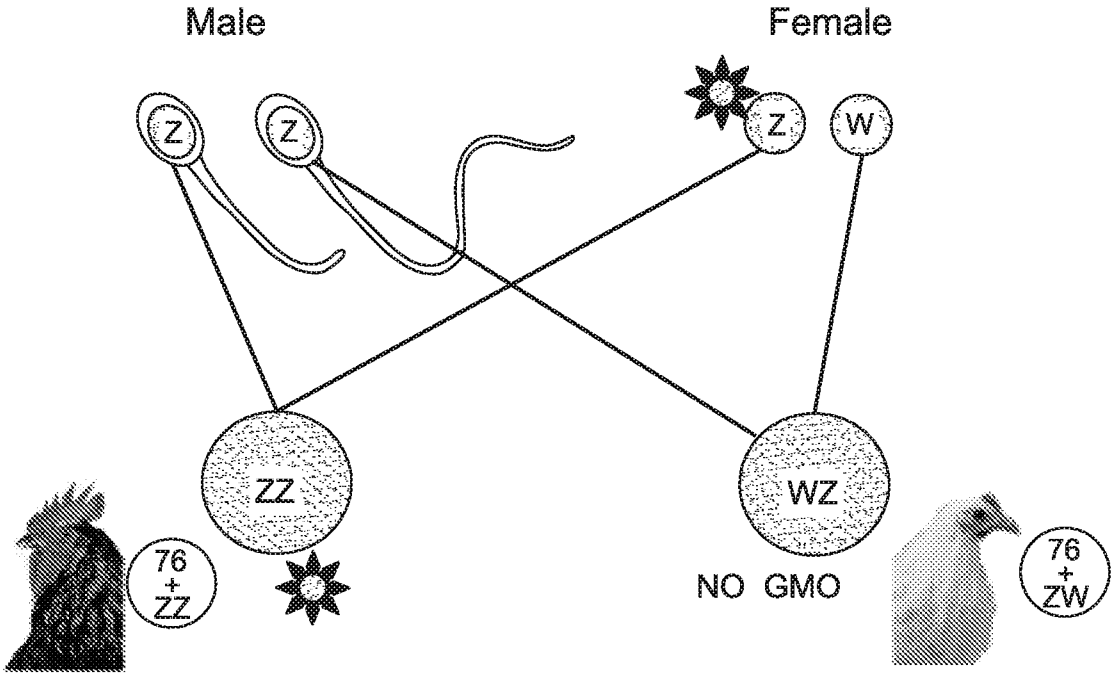

Male
Female
Z
Z
Z
W
ZZ
WZ
NO GMO
76
+
ZZ
76
+
ZW

Fig. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL2016/051291

A. CLASSIFICATION OF SUBJECT MATTER
IPC (2017.01) C12Q 1/68, C12N 15/85

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC (2017.01) C12Q 1/68, C12N 15/85

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Databases consulted: PATENTSCOPE, Google Patents, Google Scholar, FamPat database

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CA 2264450 A1 HADJANTONAKIS ANNA-KATERINA 05 Sep 1999 (1999/09/05) the whole document	1,2,5-20,23-48,50
A		3,4,21,22,49
A	US 2006160080 A1 ROSLIN INSTITUTE (EDINBURGH) 20 Jul 2006 (2006/07/20) ¶92	3,4,21,22,49
P,A	Cooper, Caitlin A., et al. "Generation of gene edited birds in one generation using sperm transfection assisted gene editing (STAGE)." Transgenic Research (2016): 1-17 28 Nov 2016 (2016/11/28) the whole document	1-50
A	Urwin, Nigel. "Would you prefer to eat genetically modified eggs, or see day-old chicks destroyed? | Nigel Urwin." The Guardian. January 16, 2014. Accessed March 23, 2017. https://www.theguardian.com/commentisfree/2014/jan/17/would-you-prefer-to-eat-genetically-modified-eggs-or-see-day-old-chicks-destroyed. 16 Jan 2014 (2014/01/16) the whole document	1-50

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 Mar 2017	23 Mar 2017

Name and mailing address of the ISA:
Israel Patent Office
Technology Park, Bldg.5, Malcha, Jerusalem, 9695101, Israel
Facsimile No. 972-2-5651616

Authorized officer
VILSKER Olga
OlgaV@justice.gov.il
Telephone No. 972-2-5651780

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL2016/051291

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L	Tizard, Mark. "Precision genome engineering in the chicken: The gap between science and market place." Lecture, IWRAB?II, Brasil, August 2014. Accessed March 23, 2017. http://www.isb.vt.edu/documents/animal-biotech/2nd-Intl-Workshop-docs/Day-1/Tizard%20-%20PGE%20in%20chicken.pdf. 21 Aug 2014 (2014/08/21) the whole document, especially p.17	1-50

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IL2016/051291

Patent document cited search report	Publication date	Patent family member(s)	Publication Date
CA 2264450 A1	05 Sep 1999	CA 2264450 A1	05 Sep 1999
US 2006160080 A1	20 Jul 2006	US 2006160080 A1	20 Jul 2006
		AU 2003251058 A1	03 Mar 2004
		CA 2495730 A1	26 Feb 2004
		EP 1529118 A1	11 May 2005
		GB 0218955 D0	25 Sep 2002
		WO 2004016812 A1	26 Feb 2004

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108474034 A
(43)申请公布日 2018.08.31

(21)申请号 201680078336.7

(22)申请日 2016.12.01

(30)优先权数据
62/262,409 2015.12.03 US

(85)PCT国际申请进入国家阶段日
2018.07.09

(86)PCT国际申请的申请数据
PCT/IL2016/051291 2016.12.01

(87)PCT国际申请的公布数据
WO2017/094015 EN 2017.06.08

(71)申请人 艾吉伊特有限公司
地址 以色列耶路撒冷

(72)发明人 D·奥芬

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 庞东成 张培源

(51)Int.Cl.
C12Q 1/6879(2018.01)
A01K 67/027(2006.01)
C12N 5/10(2006.01)
C12N 15/85(2006.01)
C12N 15/90(2006.01)

权利要求书4页 说明书30页
序列表39页 附图4页

(54)发明名称

对未孵化卵中的禽类胚胎进行性别确定的方法及其用具

(57)摘要

本发明涉及在禽类受试对象中进行性别确定和鉴定的方法。更具体而言，本发明提供了使用转基因禽类动物的非侵入性方法，所述转基因禽类动物包含整合在至少一个性别染色体Z或W中的至少一个报道基因。本发明的转基因禽类动物用于对未孵化的禽类卵中的胚胎进行性别确定和选择。

CN 108474034 A

1.一种对禽类受精的未孵化卵进行性别确定的方法,所述方法包括以下步骤:

(a)提供/获得至少一个转基因禽类动物,其包含整合在性别染色体Z和W中至少其一内的至少一个位置或部位上的至少一种外源报道基因;

(b)从所述转基因禽类受试对象、特别是动物或其任何细胞获得至少一个受精卵;

(c)确定在所述卵中是否检测到至少一个可检测信号,其中检测到所述至少一个可检测信号表明所述至少一种报道基因的表达,从而表明在所述禽类胚胎中存在所述W染色体或Z染色体。

2.如权利要求1所述的方法,其中,所述至少一个转基因禽类受试对象包含至少两种不同的报道基因,各报道基因整合在性别染色体Z或W中的一个内的至少一个基因座上。

3.如权利要求1和权利要求2中任一项所述的方法,其中,所述报道基因是至少一种生物发光报道基因。

4.如权利要求3所述的方法,其中,所述报道基因是荧光素酶。

5.如权利要求1、3和4中任一项所述的方法,其中,所述方法还包括向步骤(b)的所述卵提供与所述生物发光报道基因相容的底物和酶中的至少一种的步骤,以形成在步骤(c)检测的所述可检测信号。

6.如权利要求1所述的方法,其中,所述至少一个转基因禽类受试对象、特别是动物是雌性禽类动物,并且其中,所述至少一种报道基因整合在雌性染色体Z的至少一个位置上,从而检测到可检测信号表明所述未孵化卵中的所述胚胎是雄性。

7.如权利要求1所述的方法,其中,所述至少一个转基因禽类受试对象、特别是动物是雌性禽类动物,并且其中,所述至少一种报道基因整合在雌性染色体W的至少一个位置上,从而检测到可检测信号表明所述未孵化卵中的所述胚胎是雌性。

8.如权利要求1所述的方法,其中,使用至少一种可编程的工程化核酸酶(PEN)将所述至少一个报道基因整合在所述转基因禽类受试对象、特别是动物的所述性别染色体中。

9.如权利要求8所述的方法,其中,所述PEN是成簇的规则间隔的短回文重复(CRISPR)II型系统。

10.如权利要求9所述的方法,其中,通过由至少一个CRISPR/CRISPR相关内切核酸酶9(Cas9)系统介导的同源性定向修复(HDR),将所述至少一个报道基因整合在所述转基因禽类动物的所述性别染色体内。

11.如权利要求10所述的方法,其中,通过用下述(a)和(b)对所述禽类动物的至少一种细胞或者引入所述禽类动物中的至少一种细胞进行共转染,而将所述至少一个报道基因整合在所述转基因禽类动物的性别染色体内:

(a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种向导RNA(gRNA)的至少一种核酸序列,和

(b)至少一种第二核酸序列,其包含至少一个所述报道基因。

12.如权利要求11所述的方法,其中,所述第二核酸序列中的所述至少一个报道基因在其5'和3'处侧接有用于在整合位点处HDR的同源臂。

13.如权利要求11和12中任一项所述的方法,其中,所述第二核酸序列中的所述至少一个报道基因与性别特异性启动子、胚胎特异性启动子和诱导型启动子中的任一种可操作地连接。

14.如权利要求11至13中任一项所述的方法,其中,所述至少一个报道基因整合在所述性别染色体的至少一个非编码区内。

15.如权利要求14所述的方法,其中,所述至少一个报道基因整合在性别W染色体基因座1022859-1024215处的至少一个位点。

16.如权利要求11和15所述的方法,其中,所述至少一种gRNA包含如SEQ ID NO.1和2中任一个所示的核酸序列。

17.如权利要求16所述的方法,其中,所述gRNA包含如SEQ ID NO.1所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

18.如权利要求16所述的方法,其中,所述gRNA包含如SEQ ID NO.2所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

19.一种禽类转基因动物,其包含至少一个外源报道基因,所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个基因座处。

20.如权利要求19所述的转基因动物,其中,所述至少一个转基因动物包含至少两个不同的报道基因,各个报道基因整合在性别染色体Z或W中的一个中的至少一个基因座处。

21.如权利要求19和权利要求20中任一项所述的转基因动物,其中,所述报道基因是至少一种生物发光报道基因。

22.如权利要求21所述的转基因动物,其中,所述报道基因是荧光素酶。

23.如权利要求19所述的转基因动物,其中,所述至少一个转基因禽类动物是雌性,并且其中所述至少一个报道基因整合在雌性染色体Z的至少一个位置处。

24.如权利要求19所述的转基因动物,其中,所述至少一个转基因禽类动物是雌性,并且其中所述至少一个报道基因整合在雌性染色体W的至少一个位置处。

25.如权利要求19所述的转基因动物,其中,使用至少一种PEN将所述至少一个报道基因整合在所述转基因动物的所述性别染色体内。

26.如权利要求25所述的转基因动物,其中,所述PEN是CRISPR II型系统。

27.如权利要求26所述的转基因动物,通过由至少一个CRISPR/Cas9系统介导的HDR,将所述至少一个报道基因整合在所述转基因禽类动物的所述性别染色体内。

28.如权利要求27所述的转基因动物,其中,通过用下述(a)和(b)对所述禽类动物的至少一种细胞或者引入所述禽类动物中的至少一种细胞进行共转染,而将所述至少一个报道基因整合在所述转基因禽类受试对象、特别是动物的性别染色体内:

(a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种gRNA的至少一种核酸序列,和

(b)至少一种第二核酸序列,其包含至少一个所述报道基因。

29.如权利要求28所述的转基因动物,其中,所述第二核酸序列中的所述至少一个报道基因在其5'和3'处侧接有用于在整合位点处HDR的同源臂。

30.如权利要求28和29中任一项所述的转基因动物,其中,所述第二核酸序列中的所述至少一个报道基因与性别特异性启动子、胚胎特异性启动子和诱导型启动子中的任一种可操作地连接。

31.如权利要求28至30中任一项所述的转基因动物,其中,所述至少一个报道基因整合在所述性别染色体的至少一个非编码区内。

32.如权利要求31所述的转基因动物,其中,所述至少一个报道基因整合在性别W染色体基因座1022859-1024215处的至少一个位点。

33.如权利要求28和32所述的转基因动物,其中,所述至少一种gRNA包含如SEQ ID NO.1和2中任一个所示的核酸序列。

34.如权利要求33所述的转基因动物,其中,所述gRNA包含如SEQ ID NO.1所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

35.如权利要求33所述的转基因动物,其中,所述gRNA包含如SEQ ID NO.2所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

36.一种细胞,其包含至少一个外源报道基因,所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个基因座处。

37.如权利要求36所述的细胞,其中,所述细胞是禽类细胞。

38.如权利要求37所述的细胞,其中,所述禽类细胞是原始生殖细胞(PGC)。

39.如权利要求36至38中任一项所述的细胞,其中,通过用下述(a)和(b)对所述细胞进行共转染而将所述至少一个报道基因整合在所述细胞的性别染色体内:

(a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种gRNA的至少一种核酸序列,和

(b)至少一种第二核酸序列,其包含至少一种所述报道基因。

40.如权利要求39所述的细胞,其中,所述第二核酸序列中的所述至少一个报道基因在其5'和3'处侧接有用于在整合位点处HDR的同源臂。

41.如权利要求39和40中任一项所述的细胞,其中,所述gRNA包含如SEQ ID NO.1所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

42.如权利要求39和40中任一项所述的细胞,其中,所述gRNA包含如SEQ ID NO.2所示的核酸序列,并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

43.一种试剂盒,其包括:

(a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种向导RNA(gRNA)的至少一种核酸序列,

(b)至少一种第二核酸序列,其包含至少一个所述报道基因。

44.如权利要求43所述的试剂盒,其中,所述第二核酸序列中的所述至少一个报道基因在其5'和3'处侧接有用于在整合位点处HDR的同源臂。

45.如权利要求43和44中任一项所述的试剂盒,其中,所述第二核酸序列中的所述至少一个报道基因与性别特异性启动子、胚胎特异性启动子和诱导型启动子中的任一种可操作地连接。

46.如权利要求43至45中任一项所述的试剂盒,其中,所述至少一个报道基因整合在所

述性别染色体的至少一个非编码区中，并且其中所述至少一种gRNA包含如SEQ ID NO.1和2中任一个所示的核酸序列。

47.如权利要求46所述的试剂盒，其中，所述gRNA包含如SEQ ID NO.1所示的核酸序列，并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

48.如权利要求46所述的试剂盒，其中，所述gRNA包含如SEQ ID NO.2所示的核酸序列，并且其中包含在所述第二核酸序列内的所述至少一个报道基因在其5'和3'处分别侧接有包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

49.如权利要求43至48中任一项所述的试剂盒，其中，所述报道基因是至少一种生物发光报道基因。

50.如权利要求43至49中任一项所述的试剂盒，其用于制备转基因禽类动物，所述转基因禽类动物包含至少一个外源报道基因，所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个基因座处。

对未孵化卵中的禽类胚胎进行性别确定的方法及其用具

技术领域

[0001] 本发明涉及在禽类受试对象中进行性别确定和鉴定的方法。更具体而言，本发明提供用于对未孵化的禽类卵中的胚胎进行性别确定和选择的非侵入性方法和转基因禽类动物。

背景技术

[0002] 以下列出了被认为与本公开的主题背景相关的参考文献：

[0003] • WO 2010/103111

[0004] • WO 2014/0296707

[0005] • US 6244214

[0006] • 06124456A2

[0007] • US2014069336A

[0008] • WO16005539

[0009] • WO 96/39505

[0010] • WO 97/49806

[0011] • Quansah,E.,Long,J.A.,Donovan,D.M.,Becker,S.C.,Telugu,B.,Foster Frey,J.A.,Urwin,N.2014.Sperm-mediated transgenesis in chicken using a PiggyBac transposon system.Poultry Science Association Meeting Abstract.BARC Poster Day.

[0012] • Jinek,M.,Chylinski,K.,Fonfara,I.,Hauer,M.,Doudna,J.A.,&Charpentier,E.(2012).A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.Science,337(6096),816-821.

[0013] • Cong,L.,&Zhang,F.(2015).Genome engineering using CRISPR-Cas9 system.Chromosomal Mutagenesis,197-217.

[0014] 本文对上述参考文献的引用不应被推断为意味着它们以任何方式与当前公开的主题的可专利性相关。

[0015] 本发明的背景

[0016] 在食品工业中，每天通过窒息或磨碎剔除数十亿只鸡。雄性被终结因为它们不用于产卵或培育用作肉类，并且弱或不健康的雌性也被终止。由于伦理和经济的考虑，因此非常期望在孵化之前进行卵内或胚胎性别确定的方法。

[0017] 具体而言，视觉鉴定家禽的受精卵对于以下而言是重要的：移除未受精卵以(通过防止孵化未受精卵而)节约孵化成本，以及降低这些易污染的未受精卵继续与受精卵一起孵化的生物安全风险。

[0018] 目视鉴定当在未孵化卵内时处于胚胎早期的卵的生育力(fertility)包括外部光源验卵，并且可能很难，在早期胚胎阶段几乎不可能。一个更大的挑战是鉴定胚胎的性别，目前还没有可用的方法来区分发现受精的未孵化卵中的雄性和雌性。然而，在早期胚胎阶

段鉴定生育力以及确定家禽的性别对于养禽业、科学研究和保护而言至关重要。通过形态特征确定幼鸟的性别对大多数物种来说极具有挑战性。性别可以通过个体排泄孔性别鉴定来确定，其涉及手动挤出鸡的粪便，这样轻微地打开鸡的肛门排泄孔，允许查看鸡是否具有小的“凸起”，凸起将表明鸡是雄性。然而，该方法有高风险的鸟类伤害和性别确定错误，并且需要受过训练的人员手动进行繁琐的工作。

[0019] 排泄孔性别鉴定或鸡性别鉴定是区分鸡和其它幼体性别的方法，通常由被称为鸡雌雄鉴别师或小鸡雌雄鉴别师的受过培训的人进行。鸡的性别鉴定主要是由大型商业孵化场实行，它们必须了解两性之间的差异，以便将它们分成各性别群体，并将其纳入不同的计划，其中包括使一个群体生长而剔除另一群体，这是因为它们是不满足商业需要的性别(例如，从来自商业饲养线的产卵者的卵孵化出的雄性)。该雄性不会有良好的肉产量，并且不会产卵；因此它会在性别鉴定后被剔除。性别鉴定后，相关的性别将继续其生涯，以用于他/她的目的，而其它性别或其大部分由于与产卵无关而将在孵化后几天内被剔除。

[0020] 在产卵的农场中，雄鸡是不想要的，并且性别不合需要的鸡几乎立即被杀死以降低饲养者的成本。鸡沿着传送带移动，其中鸡性别鉴定师将雄性分离出，并将它们扔进一个通常在绞肉机中被活着磨碎的滑槽。

[0021] 在其孵化之前鉴定和确定卵的生育力和卵中胚胎的性别，可以消除未受精卵和卵中不想要的胚胎类型，从而极大地降低孵化成本(这包括能源和效率成本以及空气污染和能源消耗)。此外，鸡的痛苦将会停止，并且防止剔除导致的污染。自动化的性别鉴定装置另外将通过消除对鸡性别鉴定师的需求来降低卵生产成本，同时减少所需孵化场的规模，因为在早期阶段，50％的鸡卵会从工艺中扣除，从而减少孵化这些卵的费用，并减少之后对任何精心设计的杀戮程序的需要。

[0022] 在用于饲养、产卵或肉类生产的所有商业类型的鸟类中，都需要确定生育力和胚胎的性别。在下述方面有很大的经济回报：节能，减少生物安全风险，垃圾处置，性别鉴定劳动力成本和性别鉴定错误，剔除成本和处置以及动物福利方面。

[0023] WO 2010/103111公开了一种侵入性方法，其包括一系列步骤，其中向卵中引入经标记的抗体，其特别设计用于匹配胚胎上的性别特异性抗原。

[0024] WO 2014/0296707公开了设计用作定量或评价注射到鸟卵中的接种效率的生物标志物的发光组合物。在该公开内容中没有公开甚至暗示性别确定。美国专利6244214公开了卵内注射设备和检测方法。

[0025] WO 06124456A2公开了通过确定来自禽卵的胚液(例如，尿囊液或血液)样品中雌激素类固醇化合物的存在来进行对禽类胚胎进行卵内性别确定的侵入性方法。利用荧光显微镜通过竞争性免疫测定法测量从所述禽类卵获得的样品中的分析物来确定化合物的存在。

[0026] 还公开了光谱方法，其中US2014069336A基于筛选禽类胚胎羽毛颜色(预孵化)并基于羽毛颜色确定禽类胚胎的性别，或者WO16005539公开了对入射光信号获得壳特异性光谱响应的装置。

[0027] 针对该问题的其它遗传方法包括对从受精卵获得的DNA样品进行DNA测序以检测位于鸟类的Z和W染色体上的两个特定基因(WO 96/39505)，或使用与雌性W染色体的特异性序列杂交的寡核苷酸探针(WO 97/49806)。这些方法是侵入性的，因此未提供安全策略。

[0028] 成簇的规则间隔的短回文重复(CRISPR)/CRISPR相关(Cas)系统是现有技术的基因编辑系统,允许简单的构建体设计,成功率高(J.Doudna,2012)。

[0029] Niu等(2014)将向导RNA(gRNA)和Cas9 RNA注入猴卵母细胞以修饰三个靶基因,并且Hwang等(2013)修改了斑马鱼胚胎中的drd3和gsk3b基因以获得双基因座突变体。Cong和Zhang(2015)修改了CRISPR系统以编辑活细胞中的任何基因。

[0030] Veron和同事们(2015年)证明,通过针对PAX7转录因子的CRISPR gRNA质粒的电穿孔而修饰鸡胚中体细胞的表达水平(Nadège等2015),Bai和同事编辑了PPAR-g、ATP合成酶ε亚基(ATP5E)

[0031] Quansah,E.等公开了使用PiggyBac转座子系统在鸡中进行精子介导的转基因。具体而言,他们公开了aGFP质粒和Lipofectamine LTXTM 9LPX)组合对鸡精子的生活力、运动性或生育力没有影响。

[0032] 因此,目前没有可用的在孵化鸡之前的卵阶段进行性别鉴定的有效和非侵入性方法。所以,长期以来需要一种能够对未孵化卵中的胚胎进行准确和安全的性别鉴定的方法。

发明内容

[0033] 本发明的第一方面涉及一种对未孵化卵(特别是受精的未孵化卵)中的禽类或者禽类胚胎进行性别确定的方法。在一些具体实施方式中,所述方法可以包括以下步骤:

[0034] 首先,在步骤(a)中,提供或获得至少一个转基因禽类受试对象或动物,其包含整合在性别染色体Z和W中至少之一中的至少一个位置或部位(本文也称为基因座)上的至少一个外源报道基因。在第二步(b)中,从所述转基因禽类受试对象、特别是动物或其任何细胞获得至少一个受精卵。

[0035] 下一步骤(c)包括确定在所述卵中是否检测到至少一个可检测信号。在更具体的实施方式中,检测到至少一个可检测信号表明所述至少一个报道基因的表达,从而表明在禽类胚胎中存在W染色体或Z染色体。

[0036] 在第二方面,本发明涉及一种禽类转基因动物,在其至少一个细胞中,其包含至少一个外源报道基因,所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个位置或部位(本文也称为基因座)处。

[0037] 在又一方面,本发明涉及一种细胞,其包含至少一个外源报道基因,所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个位置或基因座处。

[0038] 在还又一方面,本发明提供一种试剂盒,其包含:

[0039] (a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种向导RNA(gRNA)的至少一种核酸序列,和(b)至少一种第二核酸序列,其包含至少一个所述报道基因。

附图说明

[0040] 为了更好地理解本文公开的主题并且举例说明其在实践中可以如何实施,现在将仅通过非限制性实例并参考附图来描述实施方式,其中:

[0041] 图1A-1B.荧光素酶报道基因信号穿透卵壳

[0042] 用萤光素皮下注射表达萤光素酶的转基因小鼠。此后10分钟切除耳(图1A)和尾

(图1B)并将其引入未受精卵中。使用生物空间光子成像仪(Bio space lab,USA)对卵进行成像。

[0043] 图2A-2B.荧光素酶报道基因信号在受精卵中形成并穿透卵壳

[0044] 将从表达萤光素酶的转基因小鼠中切下的耳(图2A)和尾(图2B)引入携带10日龄鸡胚的受精卵中。随后注射荧光素以诱导生物发光。使用生物空间光子成像仪(Bio space lab,USA)在10分钟后拍摄图像。

[0045] 图3A-3B.GFP报道基因信号不能通过卵壳检测到

[0046] 将来自表达GFP的转基因小鼠的尾引入鸡胚(10天)或置于壳外。只有放置在卵壳外的尾(图3A)可以用GFP荧光观察到,而当放置在卵内时没有检测到信号(图3B)。此后5分钟使用Maestro 2.2Imager(Cambridge Research&Instrumentation,Inc.USA)拍摄图像。

[0047] 图4.雌性禽类胚胎的检测

[0048] 将荧光素酶报道基因(星号)引入雌性转基因鸡(母鸡)的W染色体中。只有携带W和Z染色体的雌性卵提供报道基因、具体为萤光素酶的信号。

[0049] 图5.雄性禽类胚胎的检测

[0050] 将荧光素酶报道基因(星号)引入雌性转基因鸡(母鸡)的Z染色体中。雄性胚胎通过荧光素酶信号检测到,雌性不具有外源DNA。

具体实施方式

[0051] 每天数十亿只雄性小鸡通过窒息或磨碎而被终结,因为它们不能用于产卵或培育用作肉类。在孵化前确定胚胎的性别的能力在伦理和经济上都是非常重要的。

[0052] 在鸡中,性染色体的遗传构成是雄性为ZZ,雌性为ZW。这意味着W染色体决定了雌性性别。这与人类不同,其中来自父亲的Y决定了男性性别。

[0053] 本发明提供一种用于性别确定的非侵入性有效方法,其使用整合在转基因禽类受试对象的性别特异性染色体中的报道基因。该报道基因在未孵化卵的胚胎中的表达清楚且准确地鉴定了所述胚胎的性别。

[0054] 因此,本发明的第一方面涉及一种性别确定以及可选地选择未孵化卵(特别是受精的未孵化卵)中的禽类或禽类胚胎的方法。在一些具体实施方式中,该方法可以包括以下步骤:

[0055] 第一,在步骤(a)中,提供或获得至少一个转基因禽类受试对象或动物,其包含整合在性别染色体Z和W中至少之一中的至少一个位置或部位(location)处的至少一个外源报道基因。在第二步骤(b)中,从所述转基因禽类受试对象、特别是动物或其任何细胞获得至少一个受精卵。

[0056] 下一步骤(c)包括确定在所述卵中是否检测到至少一个可检测信号。在更具体的实施方式中,检测到至少一个可检测信号表明所述至少一个报道基因的表达,从而表明在所述禽类胚胎中存在所述W染色体或Z染色体。因此,在报道基因已整合在雌性转基因禽类的Z染色体中的情况下,在所检查的卵中鉴定出可检测信号表明该胚胎具有其中整合有报道基因的母本Z染色体,并且胚胎由此被鉴定为雄性。作为选择,在报道基因已整合在雌性转基因禽类的W染色体中的情况下,在所检查的卵中鉴定出可检测信号表明该胚胎携带母本W染色体并因此确定为雌性,从而提供其性别确定。

[0057] 应该意识到，如下文中更详细描述的，本发明提供的转基因禽类可以是雌性或雄性。在更具体的实施方式中，在转基因禽类受试对象是雌性时，通过本发明的方法鉴定的卵由本发明提供的转基因雌性禽类产下。在更具体的实施方式中，转基因雌性可以通过转基因雄性或通过野生型禽类雄性受精。更进一步，受精可以通过交配或通过将转基因禽类雌性用获自转基因或野生型禽类雄性的精子授精而发生。在其它实施方式中，在转基因禽类是雄性时，通过本发明的方法鉴定的卵可以由与本发明提供的转基因雄性交配的野生型或转基因雌性产下，或者通过其任何细胞(具体而言，包含整合在其性别染色体中的本发明的外源报道基因的精子细胞)授精。

[0058] 因此，本发明提供了用于检测未孵化的受精卵内禽类胚胎的性别的方法。应该意识到，本发明的方法可适用于禽类胚胎的任何胚胎期的未孵化卵。

[0059] 应该注意的是，本文所用的“禽类胚胎的胚胎发育阶段或步骤”是指：第1天的阶段，其中胚盘处于囊胚阶段并且分裂腔呈现暗环的形状；第2天的阶段，其中第一凹槽出现在囊胚的中心并且出现卵黄膜；第3天的阶段，其中血液循环开始，可以辨别头部和躯干，以及脑和开始搏动的心脏结构；第4天的阶段，其中羊膜腔正在发育以围绕胚胎并出现尿囊；第5天的阶段，其中胚胎呈C形并且四肢延伸；第6天的阶段，其中上肢和下肢的手指变得不同；第7天的阶段，其中颈部清楚地将头部与身体分开，形成喙并且大脑逐渐进入头部区域；第8天的阶段，其中容易看见眼睛色素沉着，能区分翅膀和腿并且外耳道打开；第9天的阶段，其中出现爪，并且第一羽囊出芽；第10天的阶段，其中存在鼻孔，眼睑长出并且出现卵齿；第11天的阶段，其中眼睑孔具有椭圆形并且胚胎具有鸡的形态；第12天的阶段，其中羽囊围绕外耳道并覆盖上眼睑，而下眼睑覆盖角膜的大部分；第13天的阶段，其中尿囊变成绒毛膜尿囊膜，而爪和腿鳞变得明显；第14—16天的阶段，其中全身生长迅速，卵黄收缩加速，卵白逐渐消失；第17天的阶段，其中肾系统产生尿酸盐，喙点指向气囊并且卵白被完全吸收；第18天的阶段，其中卵黄内化和羊水量减少；第19天的阶段，其中卵黄质再吸收加速并且喙准备刺穿内壳膜；在第20天的阶段中，卵黄被完全再吸收，脐带闭合，鸡刺穿内壳膜，呼吸进入气囊并准备孵化；第21天的阶段，其中鸡通过其卵齿以圆形方式刺穿卵壳，在12到18小时内自己从卵壳中脱离出来，并使其下垂干燥。

[0060] 更具体而言，本发明的方法可适用于在蛋内在胚胎发育过程的每个阶段在卵内确定禽类胚胎的性别。更具体而言，从第1天、第2天、第3天、第4天、第5天、第6天、第7天、第8天、第9天、第10天、第11天、第12天、从第13天、从第14天、从第15天、从第16天、从第17天、从第18天、从第19天、从第20天和从第21天起。更具体而言，本发明的方法可适用于胚胎性别的早期检测，特别是从第1天至第10天，更具体地说，第1天至第5天的胚胎性别的早期检测。

[0061] 如上所述，本发明的方法可适用于受精的未孵化的卵。术语“受精的卵”在下文中指母鸡产的卵，其中母鸡在两周内与公鸡交配过，允许雄性精子沉积到雌性输卵管漏斗中并且在卵子从卵巢释放时发生受精事件。如本文所用，“未孵化的卵”涉及在结构上完整(而非破碎)的壳内的含胚的卵(本文中也称为可育卵)。

[0062] 本发明的方法基于确定由报道基因形成的可检测信号，所述报道基因整合在产下所检查的卵的转基因禽类雌性或雄性的特定基因座。

[0063] 如本文所用的“将外来或外源DNA/基因整合在染色体中”，在下文中是指生物体染色体的核苷酸序列的永久修饰。该修饰在细胞分裂过程中进一步传递，并且如果发生在生

殖细胞系中，它也会传递给后代。在这种情况下，整合的报道基因可以转移到未孵化卵内的胚胎中。如本文所用的术语“外源”是指从生物体外部利用特异性操作的载体、病毒或任何其它载质(vehicle)通过转化或转染引入生物体内。根据某些实施方式的整合的外源基因可以是报道基因。术语“报道基因”是指编码多肽的基因，所述多肽表达可以以多种已知测定检测到，并且其中所检测的信号的水平指示所述报道基因的存在。

[0064] 如上所述，外源报道基因可以整合在禽类性别染色体Z或W中。如本文所用的禽类“性别染色体Z或W”是指鸡中决定后代的性别的染色体系统，其中雄性是同配性别(ZZ)，而磁性是异配性别(ZW)。卵子中W染色体的存在决定了后代的性别，而已知Z染色体是更大的，并拥有更多的基因。

[0065] 本发明的方法基于可检测信号的检测，该可检测信号指示并反映了报道基因的存在并由此指示并反映了特定的性别染色体的存在。“可检测信号”在下文中是指通过观察或仪器可察觉的改变。不旨在限制，信号可以直接检测或仅在存在试剂的情况下检测。在一些实施方式中，可检测的响应是光学信号，包括但不限于化学发光基团。

[0066] 应该意识到，在一些具体实施方式中，由本发明的方法提供的至少一个转基因禽类受试对象可以包含至少两个不同的报道基因，每个报道基因可以整合在性别染色体Z或W的至少一个位置或部位。在至少两个不同的报道基因的情况下，每个性别染色体可以被不同地标记。形成的可检测信号的评估可以指示所检查胚胎的性别。

[0067] 在又一些具体实施方式中，本发明的转基因禽类体内包含的报道基因可以是至少一种生物发光报道基因。因此，在一些实施方式中，所表达的多肽是生物发光蛋白，因此该测定测量从生物发光反应发射的光的水平。

[0068] 术语“生物发光”是指生物分子如蛋白质发光。生物发光涉及分子氧、加氧酶和作用于底物荧光素的萤光素酶，将在下文中更详细描述。

[0069] 在更具体的实施方式中，报道基因可以是萤光素酶。术语“荧光素酶”在下文中是指产生生物发光(光子发射)的一类氧化性酶。发射的光子可以通过光敏装置如光度计或改造的光学显微镜来检测。萤光素酶可以通过基因工程在各种生物体中产生，主要用作报道基因。荧光素酶天然存在于细菌、藻类、真菌、水母、昆虫、虾和乌贼中。在细菌中，负责发光反应的基因(编码进lux操纵子的lux基因)已经被分离，并广泛用于构建发射在490nm具有最大强度的蓝绿光的生物报道子。可利用lux的三种变体，一种在<30℃下起作用，另一种在<37℃下起作用，第三种在<45℃下起作用。lux遗传系统由五个基因luxA、luxB、luxC、luxD和luxE组成。取决于所使用的这些基因的组合，可以构建数种不同类型的生物发光生物报道子。萤光素酶蛋白是由luxA和luxB基因产物形成的异二聚体。luxC、luxD和luxE基因产物分别编码还原酶、转移酶和合成酶，它们在单一复合物中一起作用以生成用于生物发光反应的醛底物。luxAB生物报道子只含有能够产生光信号的luxA和luxB基因。但是，为了充分完成发光反应，必须对细胞供应底物(长链醛)。

[0070] 另一方面，luxCDABE生物报道子含有lux盒的所有五个基因，因此允许完全独立的光生成系统，其不需要额外添加底物，也不需要外部光源的任何激发。由于其快速和容易使用，以及实时和在线重复执行生物测定的能力，使得luxCDABE生物报道子极具吸引力。因此，在某些实施方式中，本发明的方法可使用luxCDABE生物报道子作为报道基因。

[0071] 在又一些进一步的实施方式中，本发明的方法可以使用luc基因作为报道基因。萤

火虫萤光素酶(luc基因)催化产生550-575nm范围内的可见光的反应。也可利用叩头虫荧光素酶,其产生更接近595nm峰值的光。两种萤光素酶都需要添加外源底物(荧光素)以发生光反应。

[0072] 应该意识到,本领域已知的任何来源的本文所述的任何萤光素酶都可以适用于本发明的方法和试剂盒。

[0073] 在又一些具体的实施方式中,本发明的方法可以使用的萤光素酶可以是桡脚虫(Gaussia princes)萤光素酶。在更具体的实施方式中,本发明使用的萤光素酶可以是由GenBank:AY015993.1公开的核酸序列编码的萤光素酶,其具有GenBank:AAG54095.1公开的氨基酸序列。在又一些进一步的具体实施方式中,本发明的方法和试剂盒所使用的萤光素酶可以由包含如SEQ ID NO.22所示的序列的核酸序列编码。在又一些进一步的实施方式中,这种萤光素酶可以包含如SEQ ID NO.23所示的氨基酸序列,或其任何同源物、突变体或衍生物。

[0074] 在又一些进一步的实施方式中,本发明使用的萤光素酶可以是北美萤火虫(P.pyralis)(萤火虫)萤光素酶。在一些具体实施方式中,这种荧光素酶可以是如GenBank:M15077.1所公开的核酸序列编码的萤光素酶,其具有如GenBank:AAA29795.1所公开的氨基酸序列。在又一些进一步的具体实施方式中,本发明的方法和试剂盒所使用的萤光素酶可以由包含如SEQ ID NO.20所示的序列的核酸序列编码。在又一些进一步的实施方式中,这种萤光素酶可以包含如SEQ ID NO.21所示的氨基酸序列,或其任何同源物、突变体或衍生物。

[0075] 如上所述,本发明的方法使用的萤光素酶可能需要补充额外的试剂,特别是底物。

[0076] 因此,在又一些进一步的实施方式中,所述方法可以进一步包括向步骤(b)的所述卵提供与生物发光报道基因相容的至少一种底物和酶的步骤。应该注意的是,可能需要这种底物或酶来形成步骤(c)中检测到的可检测信号。更具体而言,本发明的方法可以包括向步骤(b)的卵提供(例如通过注射提供)用于萤光素酶的底物的步骤。在一些具体的实施方式中,这种底物可以是萤光素。如本文所用的荧光素是产生生物发光的生物体中发现的发光化合物的通用术语。荧光素通常经历酶促催化氧化,并且所得到的激发态中间体在衰减到其基态时发光。在又一些进一步的实施方式中,将作为所述可检测信号形成中的必需要素的底物荧光素,注入所述卵,特别是在如步骤(c)中所进行的测定和确定所述信号之前注入所述卵。在一些实施方式中,所述底物可以在所述禽类受试对象(特别是动物)的胚胎发育的第1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21日注入。在又一些进一步的实施方式中,形成可检测信号所需的底物和/或其它酶可作为编码所述底物和/或酶的核酸序列(其可操作地与所述报道基因连接)提供给受精卵。此类具体实施方式可以涉及使用如上所述的luxCDABE生物报道子。luxCDABE系统包含lux盒的五个基因,因此允许完全独立的光生成系统,不需要额外添加底物或不需要外部光源的任何激发。

[0077] 在一些实施方式中,应该注意的是可检测信号,具体而言,生物发光信号可以使用合适的生物发光手段来检测。在一些实施方式中,由荧光素酶报道基因形成的可检测信号可以通过光敏装置例如发光计或改造的光学显微镜或电荷耦合器件(CCD)(高度敏感性光子检测器)来检测。

[0078] 在更进一步的实施方式中,通过本发明的方法提供的至少一种转基因禽类受试对

象或动物可以是雌性禽类受试对象或动物。在更具体的实施方式中，当至少一个报道基因整合在雌性染色体Z的至少一个位置中时，检测到可检测信号表明未孵化卵中的胚胎是雄性。

[0079] 在又一些进一步的实施方式中，通过本发明的方法提供的至少一种转基因禽类受试对象或动物可以是雌性禽类受试对象或动物。在更具体的实施方式中，当至少一个报道基因整合在雌性染色体W的至少一个位置中时，检测到可检测信号表明未孵化卵中的胚胎是雌性。

[0080] 在又一些进一步的实施方式中，通过本发明的方法提供的转基因动物可以是具有整合在其Z染色体中的报道基因的雄性受试对象。在这种情况下，在由这种转基因雄性或其任何精子受精的卵中确定的可检测信号，表明胚胎携带包含转基因报道基因的亲本Z染色体，因此是雄性。在更进一步的实施方式中，检测由转基因雄性禽类受精的转基因雌性禽类(均携带整合在其Z染色体中的本发明的报道基因)产下的卵中的可检测信号，在强烈信号的情况下，可以表明胚胎携带整合在其雌性和雄性Z染色体中的报道基因的两个拷贝。在信号不太强烈的情况下，卵可能会被确定为雌性。

[0081] 如前所述，本发明的方法涉及提供转基因禽类动物。转基因禽类动物的制备需要使用可能利用特定核酸酶的基因工程方法。

[0082] 因此，在更具体的实施方式中，可以使用至少一种可编程工程化核酸酶(PEN)将所述至少一个报道基因整合在由本发明的方法提供的转基因禽类受试对象或动物的性别染色体中。如本文所用的术语“可编程工程化核酸酶(PEN)”是指切割特定DNA序列的合成型酶，其源自涉及双链DNA损伤的DNA修复的天然存在的核酸酶并且能够进行直接的基因组编辑。

[0083] 成簇的规则间隔的短回文重复(CRISPR)II型系统是一种细菌免疫系统，其已被修改用于基因组工程化。

[0084] 然而，应该意识到，本文也可以适用其它基因组工程化方法，例如锌指核酸酶(ZFN)或转录激活子样效应核酸酶(TALEN)，其依赖于使用可定制的DNA结合蛋白核酸酶，该核酸酶需要设计和产生特定的核酸酶-对，以用于每个基因组靶。

[0085] 如本文所用，也称为SPIDR(间隔物散布直接重复序列)的CRISPR阵列构成最近描述的DNA基因座家族，其通常对特定细菌物种具有特异性。CRISPR阵列是首次在大肠杆菌中识别的独特类型的散布短序列重复序列(SSR)。在随后的几年中，在结核分枝杆菌(Mycobacterium tuberculosis)、地中海富盐菌(Haloferax mediterranei)、詹氏甲烷球菌(Methanocaldococcus jannaschii)、海栖热袍菌(Thermotoga maritima)和其它细菌和古细菌中发现了类似的CRISPR阵列。应该理解，本发明考虑使用任何已知的CRISPR系统，特别是本文公开的CRISPR系统。CRISPR-Cas系统在原核生物中进化，以通过靶向外来DNA或RNA来防止噬菌体攻击和不希望的质粒复制。CRISPR-Cas系统基于短同源DNA序列(存在于重复序列之间，称为间隔物)靶向DNA分子。这些间隔物指导CRISPR相关(Cas)蛋白与被称为原间隔物(proto-spacer)的外来DNA内的序列匹配(和/或互补)，原间隔物(proto-spacer)随后被切割。间隔物可以合理设计以靶向任何DNA序列。而且，该识别元件可以被分开设计以识别和靶向任何期望的目标。对于CRISPR系统，如本领域技术人员将认识到的，天然存在的CRISPR基因座的结构包括通常被称为“重复序列”的许多短重复序列。重复序列成簇出

现，并且通常由被称为“间隔物”的独特间插序列间隔开。通常，CRISPR重复序列长度在约24至47个碱基对(bp)之间变化并且部分回文。间隔物位于两个重复序列之间，并且通常每个间隔物具有长度为约20以下至72以上bp的独特序列。因此，在某些实施方式中，在编码本发明方法和试剂盒中的至少一种gRNA的序列中使用的CRISPR间隔物可以各自包含10至75个核苷酸(nt)。更具体而言，包含约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75或更多个核苷酸(nt)。在一些具体实施方式中，间隔物包含约20至25个核苷酸，更具体而言约20个核碱基。

[0086] 除了至少一个重复序列和至少一个间隔物之外，CRISPR基因座还包括前导序列，以及可选的编码至少一个tracrRNA的序列。前导序列通常是与第一重复的5’末端直接连接的长达550bp的富含AT的序列。

[0087] 在更具体的实施方式中，PEN可以是成簇的规则间隔的短回文重复(CRISPR)II型系统。

[0088] 更具体而言，描绘了三种主要类型的CRISPR-Cas系统：I型、II型和III型。

[0089] II型CRISPR-Cas系统包括“HNH”型系统(链球菌样；也被称为Nmeni亚型，用于脑膜炎奈瑟氏菌(Neisseria meningitidis)血清群A品系Z2491或CASS4)，其中除了普遍存在的Cas1和Cas2外，Cas9(一种单一的非常大的蛋白质)似乎足以产生crRNA并切割靶DNA。Cas9含有至少两个核酸酶结构域，靠近氨基末端的RuvC样核酸酶结构域和在蛋白质中间的HNH(或McrA样)核酸酶结构域，但是这些结构域的功能仍有待阐明。然而，由于HNH核酸酶结构域在限制性酶中是丰富的并且具有负责靶标切割的内切核酸酶活性。

[0090] II型系统通过不寻常的机制切割pre-crRNA，该机制涉及tracrRNA和pre-crRNA中一部分重复之间的双链体形成；pre-crRNA加工途径中的第一次切割随后发生在该重复区域中。更进一步，应该注意的是，II型系统包含cas9、cas1、cas2、csn2和cas4基因中的至少一个。应该意识到，任何II型CRISPR-Cas系统可以适用于本发明，特别是II-A或B型中的任何一种。

[0091] 因此，在又一些进一步的替代性实施方式中，本发明的方法和试剂盒中使用的至少一种cas基因可以是II型CRISPR系统(II-A型或II-B型)的至少一种cas基因。在更具体的实施方式中，本发明的方法和试剂盒所使用的至少一种II型CRISPR系统的cas基因可以是cas9基因。应该意识到，这种系统可以还包含cas1、cas2、csn2和cas4基因中的至少一个。

[0092] 双链DNA(dsDNA)被Cas9切割是“II型CRISPR-Cas”免疫系统的标志。CRISPR相关蛋白Cas9是RNA导向的DNA内切核酸酶，其使用RNA：DNA互补性来鉴定序列特异性双链DNA(dsDNA)切割的靶位点，产生HDR所需的双链制动(DSB)，这引起报道基因在特定的靶序列(例如在禽类性别染色体W和Z内的特定靶标)中的整合。靶向DNA序列由CRISPR阵列指定，CRISPR阵列是由短回文重复序列分隔的一系列约30至40bp间隔物。该阵列被转录为pre-crRNA，并被加工成与Cas蛋白复合物相结合的较短的crRNA，从而靶向被称为原间隔物的互补DNA序列。这些原间隔物靶标还必须具有额外的邻近序列，其被称为靶标识别所需的原间隔物邻接基序(PAM)。结合之后，Cas蛋白复合物充当DNA内切核酸酶以切割靶标上的两条链，随后通过外切核酸酶活性发生DNA降解。

[0093] 如本文所用的CRISPR II型系统需要包含两个基本组分:“向导”RNA(gRNA)和非特异性CRISPR相关内切核酸酶(Cas9)。gRNA是由Cas9结合所需的“支架”序列和定义待修饰的基因组靶标的约20个核苷酸长的“间隔”或“靶向”序列组成的短合成RNA。因此,可以通过简单地改变gRNA中存在的靶向序列来改变Cas9的基因组靶标。如本文所用的向导RNA(gRNA)是指内源性细菌crRNA和tracrRNA的合成融合体,为Cas9核酸酶提供靶向特异性和支架/结合能力。其也被称为“单一向导RNA(single guide RNA)”或“sgRNA”。CRISPR最初用于“敲除”各种细胞类型和生物体中的靶基因,但对Cas9酶的修饰已经将CRISPR的应用扩展到“敲入”靶基因,选择性激活或抑制靶基因,纯化DNA的特定区域,甚至使用荧光显微镜观察活细胞中的DNA图像。此外,产生gRNA的容易程度使得CRISPR成为最可扩展的基因组编辑技术之一,并且最近已被用于全基因组筛选。

[0094] 其中将整合本发明的报道基因的待编辑基因组内的靶标,特别是性别染色体Z或W内的特定靶标基因座,应该紧邻原间隔物邻接基序(Protospacer Adjacent Motif,PAM)的上游存在。

[0095] PAM序列对于靶标结合是绝对必需的,并且确切的序列取决于Cas9的物种(对于酿脓链球菌(Streptococcus pyogenes)Cas9而言为5'GNG 3')。在某些实施方式中,来自酿脓链球菌的Cas9用于本发明的方法和试剂盒中。尽管如此,应该意识到,任何已知的Cas9都可以适用。本发明中有用的Cas9的非限制性实例包括但不限于:酿脓链球菌(SP),本文也表示为SpCas9;金黄色葡萄球菌(Staphylococcus aureus,SA),本文也表示为SaCas9;脑膜炎奈瑟氏菌(Neisseria meningitidis,NM),本文也表示为NmCas9;嗜热链球菌(Streptococcus thermophilus,ST),本文也表示为StCas9;和齿垢密螺旋体(Treponema denticola,TD),本文也表示为TdCas9。在一些具体实施方式中,酿脓链球菌M1GAS的Cas9,特别是蛋白质ID:AAK33936.1的Cas9可以适用于本发明的方法和试剂盒中。在一些实施方式中,Cas9蛋白可由如SEQ ID NO.24所示的核酸序列编码。在进一步的具体实施方式中,Cas9蛋白可以包含如SEQ ID NO.25所示的氨基酸序列,或其任何衍生物,突变体或变体。一旦表达,Cas9蛋白和gRNA通过gRNA“支架”结构域和Cas9上表面暴露的带正电荷的沟槽之间的相互作用形成核糖蛋白复合物。Cas9在gRNA结合时发生构象变化,使分子从无活性的非DNA结合构象转变为活性的DNA结合构象。重要的是,gRNA的“间隔物”序列仍然自由地与靶DNA相互作用。Cas9-gRNA复合物与结合具有PAM的任何基因组序列,但gRNA间隔物与靶DNA匹配的程度决定了Cas9是否会切割。一旦Cas9-gRNA复合物结合推定的DNA靶标,gRNA靶向序列3'末端的“种子”序列开始与靶DNA退火。如果种子和目标DNA序列匹配,则gRNA继续以3'至5'方向与靶DNA退火。

[0096] 如果gRNA间隔物和靶序列之间存在足够的同源性,则Cas9将只切割靶标。此外,Cas9核酸酶具有两个功能性内切核酸酶结构域:RuvC和HNH。Cas9在靶标结合后发生第二次构象变化,其定位核酸酶结构域以切割靶DNA的相反链。Cas9介导的DNA切割的最终结果是在靶DNA内的双链断裂(DSB),其发生在PAM序列上游约3至4个核苷酸处。

[0097] 然后所得到的DSB可以通过两种一般修复途径之一修复,所述两种一般修复途径是有效但容易出错的非同源末端连接(NHEJ)途径和效率较低但高保真度的同源性定向修复(HDR)途径。在一些实施方式中,导致本发明的报道基因与性别染色体W或Z内的特定靶基因座的特异性整合的插入,是修复由Cas9引起的DSB的结果。在一些具体实施方式中,本发

明的报道基因通过HDR整合或敲入靶基因座。

[0098] 如本文所用的术语“同源性定向修复(HDR)”是指在细胞中修复双链DNA损伤的机制。HDR最常见的形式是同源重组。HDR修复机制只能在细胞核中存在DNA的同源碎片(homologue piece)时(主要在细胞周期的G2和S期)被细胞利用。当同源DNA碎片不存在时，可以替代地发生被称为非同源末端连接(NHEJ)的另一过程。用于基因组编辑的可编程工程化核酸酶(PEN)策略基于特定双链DNA切割后HDR机制的细胞激活。

[0099] 如前所讨论的，Cas9通过两个核酸酶结构域RuvC和HNH的组合活性而产生双链断裂(DSB)。每个核酸酶结构域内对于内切核酸酶活性至关重要的确切氨基酸残基是已知的(在酿脓链球菌Cas9中对于HNH而言为D10A，对于RuvC而言为H840A)，并且已经生成仅含有一个活性催化结构域的Cas9酶的修饰形式(称为“Cas9切口酶”)。Cas9切口酶仍然基于gRNA特异性结合DNA，但切口酶仅能够切割DNA链之一，产生“缺口”或单链断裂而不是DSB。使用完整的互补DNA链作为模板，通过HDR(同源性定向修复)快速修复DNA切口。因此，需要靶向相反链的两个切口酶在靶DNA内产生DSB(通常称为“双切口”或“双切口酶”CRISPR系统)。该要求显著地提高了靶标特异性，因为不太可能两个脱靶切口会产生地足够近以致于导致DSB。因此应该理解，本发明进一步包括使用双切口酶方法来产生双切口诱导的DSB以增加特异性并减少脱靶效应。

[0100] 因此，在某些实施方式中，所述至少一个报道基因可通过至少一种CRISPR/CRISPR相关内切核酸酶9(Cas9)系统介导的同源性定向修复(HDR)而整合在转基因禽类受试对象(特别是动物)的性别染色体中。

[0101] 在一些进一步的实施方式中，本发明试剂盒的gRNA可以包含至少一种CRISPR RNA(crRNA)和至少一种反式激活crRNA(tracrRNA)。

[0102] 在一些替代性实施方式中，本发明的试剂盒可以包含编码至少一种gRNA的核酸序列。此类核酸序列可以包含CRISPR阵列，其包含至少一个间隔物序列，所述间隔物序列靶向目标基因组DNA序列中的至少一个原间隔物并因此与其同一。应该注意，核酸序列还包含编码至少一种tracrRNA的序列。

[0103] 在一些实施方式中，根据本公开内容的CRISPR阵列包含至少一个间隔物和至少一个重复序列。在又一个实施方式中，本发明还涵盖了提供pre-crRNA的选择，该pre-crRNA能够加工成可以靶向相同或不同靶标的数种最终gRNA产物。

[0104] 在又一些更具体的实施方式中，包含在本发明的gRNA内的crRNA可以是与靶基因组DNA序列互补的单链核糖核酸(ssRNA)序列。在一些具体实施方式中，靶基因组DNA序列可以位于紧邻原间隔物邻接基序(PAM)序列的上游或更上游。

[0105] 如本文所指出的，本发明试剂盒的gRNA可以与靶基因组DNA至少部分地互补。在某些实施方式中，“互补性”是指各自遵循锁钥原理的两个结构之间的关系。实质上互补性是DNA复制和转录的基本原理，因为它是两个DNA或RNA序列之间共有的性质，从而当它们彼此反向平行比对时，序列中每个位置上的核苷酸碱基将是互补的(例如，A和T，或者U、C和G)。

[0106] 如上所述，由本发明试剂盒的gRNA靶向的基因组DNA序列位于紧邻PAM序列的上游。在一些实施方式中，此类PAM序列可以是核酸序列NGG。

[0107] 在某些实施方式中，本发明涉及的PAM序列可以包含N，即任何核苷酸，具体地，腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)或胸腺嘧啶(T)中的任一种。在另一些实施方式中，根据本

发明的PAM序列由A、G、C或T和两个鸟嘌呤组成。

[0108] 根据一个实施方式，编码本发明的gRNA的多核苷酸可以包含至少一个间隔物并可选地包含至少一个重复序列。在又一些进一步的实施方式中，编码本发明的gRNA的DNA可以包含至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或更多个，特别是110、120、130、140、150、160、170、180、190、200个或更多个间隔物。在一些实施方式中，各间隔物位于两个重复序列之间。应该进一步理解，编码本发明gRNA的核酸序列的间隔物可以是相同的或不同的间隔物。在更多实施方式中，这些间隔物可以靶向相同或不同的靶基因组DNA。在又一些进一步的实施方式中，此类间隔物可以靶向至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或更多个靶基因组DNA序列。这些靶序列可以来自单个基因座或可选地来自数个靶基因座。

[0109] 如本文所用，术语“间隔物”是指被设计为靶向特定序列且位于CRISPR阵列的多个短直接重复序列(即CRISPR重复序列)之间的非重复间隔物序列。在一些具体的实施方式中，间隔物可包含约15至约30个核苷酸，具体地约15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30或更多个核苷酸。更具体而言，约20-25个核苷酸。

[0110] 由本发明的CRISPR系统编码的向导或靶向RNA可以包含CRISPR RNA(crRNA)和反式激活RNA(tracrRNA)。由CRISPR间隔物编码的靶向RNA的序列没有特别限制，除了要求其针对(即具有与以下相同或互补的区段)的禽类基因组DNA中在本文中也被称为“原间隔物”的靶序列。此类原间隔物包含与由编码本发明方法和试剂盒的gRNA的核酸序列内包含的CRISPR间隔物编码的靶向RNA具有充分互补性的核酸序列。

[0111] 在一些实施方式中，crRNA包含20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或35或40nt的间隔物(靶向)序列或由其组成，然后是19-36nt的重复序列。在具体的非限制性实施方式中，靶向间隔物可以包含靶向任何一种基因组DNA序列的区段或由其组成，针对该基因组DNA序列的代表性间隔物序列在本文中指出。

[0112] 应该注意的是，在一些具体实施方式中，本发明的CRISPR系统的间隔物可以编码靶向向导RNA(gRNA)。“gRNA”或“靶向RNA”是这样的RNA：当从编码它的CRISPR系统的部分转录时，包含RNA序列的至少一个区段，其与靶基因组DNA中的DNA序列相同(除了在RNA的情况下用U代替T之外)，或者互补(从而“靶向”)。本公开内容的CRISPR系统可以可选地编码多于一种靶向RNA，并且靶向RNA针对基因组DNA中的一种或多种靶序列。

[0113] 再进一步地，在一些实施方式中，可以用以下(a)和(b)共转染禽类受试对象(特别是动物)的至少一种细胞或者共转染引入到禽类受试对象(特别是动物)中的至少一种细胞中，从而将所述至少一个报道基因整合在转基因禽类受试对象(特别是动物)的性别染色体中：(a)至少一个第一核酸序列，其包含编码至少一个Cas9蛋白的至少一个核酸序列和编码至少一个向导RNA(gRNA)的至少一个核酸序列；和(b)包含至少一个报道基因的至少一个第

二核酸序列。

[0114] 因此，为了制备本发明方法所使用的转基因禽类动物，应提供至少两种核酸分子。

[0115] 如本文所用，“核酸或核酸分子”与术语“多核苷酸”可互换，且其通常指任何多核糖核苷酸或多脱氧核糖核苷酸，它们可为未修饰的RNA或DNA或经修饰的RNA或DNA或其任何组合。“核酸”包括但不限于单链和双链核酸。如本文所用，术语“核酸”还包括如上所述的含有一个或多个经修饰碱基的DNA或RNA。如本文所用，术语“寡核苷酸”定义为由两个或更多个脱氧核糖核苷酸和/或核糖核苷酸组成的分子，并且优选多于三个。其确切的大小取决于许多因素，而这进而取决于寡核苷酸的最终功能和用途。寡核苷酸长度可以是约8至约1,000个核苷酸。更具体而言，本发明的试剂盒所使用的寡核苷酸分子可以包含长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、200、300、400、500、600、700、800、900、1000个或更多个碱基中的任一种。

[0116] 核酸分子可以由天然存在的核苷酸(例如DNA和RNA)的单体或天然存在的核苷酸的类似物(例如天然存在的核苷酸的α-对映体形式)或经修饰的核苷酸或其任何组合组成。在本文中该术语还包括cDNA，即通过逆转录酶(RNA-依赖性DNA聚合成酶)的作用从RNA模板产生的互补或复制DNA。

[0117] 就此而言，“分离的多核苷酸”是从生物体的基因组中分离出的核酸分子。例如，已经从细胞的基因组DNA分离出的、编码本发明方法和试剂盒所使用的报道基因或其任何衍生物或同系物的DNA分子，以及编码本发明方法和试剂盒的CRISPR/Cas9和gRNA的序列，是分离的DNA分子。分离的核酸分子的另一实例是化学合成的核酸分子，其未整合在生物体的基因组中。从特定物种分离出的核酸分子小于该物种染色体的完整DNA分子。

[0118] 在一些实施方式中，本发明的方法和试剂盒所使用的核酸序列，具体而言，包含编码Cas9和gRNA的序列的核酸序列，或者本发明的报道基因，可以构建在载体内提供。因此本发明进一步涉及包含本发明的多核苷酸的重组DNA构建体，以及可选地与本发明的核酸序列可操作连接的另外的附加元件，例如启动子、调节控制元件，翻译、表达和其它信号。

[0119] 如本文所用，术语“重组DNA”、“重组核酸序列”或“重组基因”是指包含编码本发明CRISPR系统之一(具体而言CRISPR/Cas9II型)的开放阅读框的核酸，以及将Cas9靶向至禽类染色体Z和/或W内的特定基因座的本发明的gRNA。在又一个实施方式中，本文使用的重组DNA进一步指包含开放阅读框的核酸，所述开放阅读框编码本发明的报道基因，特别是转基因。

[0120] 如本文所提到的，术语“基因”或“转基因”是指编码蛋白质产物的核酸，其是天然存在的或合成的。术语“核酸”旨在表示天然和/或合成的线性、环状的有序阵列，例如cDNA、基因组DNA(gDNA)、mRNA和RNA、寡核苷酸、寡核苷及其衍生物。

[0121] 短语“可操作地连接”意在表示以允许转基因转录的方式连接。术语“编码”旨在表示，例如当主题核酸与合适的控制序列(例如启动子和增强子元件)在合适的载体(例如表达载体)中连接时以及当将载体引入合适的系统或细胞中时，主题核酸可以在合适的表达系统中转录并翻译成期望的多肽或主题蛋白质。

[0122] 应该理解的是，在一些实施方式中，由本发明的方法和试剂盒所提供和使用的第一和第二核酸序列中的至少一个可以被构建并包含在载体内。如本文所用，“载体”或“载

质”包括载体，例如质粒、噬菌粒、病毒，可整合的DNA片段和其它载质，它们能够使DNA片段整合在宿主的基因组中，或者能够表达未整合的遗传元件。载体通常是自我复制的DNA或RNA构建体，其含有所需核酸序列以及可操作连接的遗传控制元件，所述遗传控制元件在合适的宿主细胞中被识别并影响所需间隔物的翻译。通常，遗传控制元件可以包括原核启动子系统或真核启动子表达控制系统。此类系统通常包括转录启动子、转录增强子以提高RNA表达水平。载体通常含有允许载体独立于宿主细胞复制的复制起点。在又一些替代性实施方式中，本发明所使用的表达载体可以包含将本发明的所需报道基因整合在禽类性别特异性染色体W和/或Z中所需的元件。

[0123] 因此，术语“控制和调节元件”包括启动子、终止子和其它表达控制元件。此类调节元件描述于Goeddel；[Goeddel.，等，Gene Expression Technology:Methods in Enzymology 185,Academic Press,San Diego,Calif.(1990)]。例如，在这些载体中可以使用当与DNA序列可操作连接时控制DNA序列的表达的各种表达控制序列中的任一种，以使用本发明的方法表达编码任何所需蛋白质的DNA序列。

[0124] 载体可另外包含适当的限制性位点、抗生素抗性或用于选择含载体细胞的其它标记。质粒是最常用的载体形式，但是本领域已知或者将已知的具有等效功能的其它形式的载体，适合在本文使用。参见例如Pouwels等，Cloning Vectors:a Laboratory Manual (1985 and supplements),Elsevier,N.Y.;and Rodriquez,等(eds.)Vectors:a Survey of Molecular Cloning Vectors and their Uses,Buttersworth,Boston,Mass(1988，将其通过引用并入本文。

[0125] 为了产生由本发明方法所使用的转基因禽类动物，必须制备包含整合在性别染色体Z或W内的特定基因座中的报道基因的禽类细胞。此类细胞可以通过用本发明的方法和试剂盒提供的第一和第二核酸序列或与包含其的任何构建体共转染细胞来制备。本文使用的“转染”是指将遗传物质如DNA和双链RNA插入哺乳动物细胞中的过程。将DNA插入细胞能够使用细胞自己的机制来表达或生产蛋白质。因此，本文所用的共转染是指将至少两种不同的核酸分子或包含其的任何载体同时转染至各个单细胞。更进一步而言，待转染的核酸序列可以在短时间内瞬时表达，或者引入到基因组DNA中，其中变化在细胞分裂时从细胞传递到细胞。

[0126] 因此，本发明提供了基于报道基因(特别是萤光素酶)的表达在卵内对禽类胚胎进行卵内性别确定的方法。“表达”通常是指将基因编码信息转化为细胞中存在和操作的结构的过程。因此，根据本发明，报道基因的“表达”具体地可以指转录成多核苷酸，翻译成蛋白质，或者甚至蛋白质的翻译后修饰。

[0127] 在又一些进一步的具体实施方式中，第二核酸序列中的至少一个报道基因可以在其5’和3’侧接有同源臂。应该意识到，在一些实施方式中，这些臂是必需的，因此有利于报道基因在整合位点处的HDR。

[0128] 在更具体的实施方式中，由本发明方法所使用的第二核酸序列中的报道基因可以侧接有两个臂，所述两个臂是同源的，或者与包含在性别染色体Z或W内的目标基因座内的至少一个核酸序列显示约70％、71％、72％、73％、74％、75％、76％、77％、78％、79％、80％、81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的同源性或同一性，其充当整合位点以促进经由HDR的特异

性整合。在某些实施方式中，靶序列在本文中也称为至少一个由“间隔物”序列识别的“原间隔物”，所述“间隔物”序列是由本发明使用的并由第一核酸序列提供的gRNA的一部分。

[0129] 本文所用的术语“同源臂”是指引入到特定载体或病毒中的HDR模板，其用于产生特定突变或将新元件插入到基因中，其具有一定量的围绕待修饰靶序列的同源性(取决于使用哪种PEN)。在又一些具体的实施方式中，在CRISPR用作PEN时，臂序列(左侧，上游和右侧，下游)可以包含约10-5000bp，特别约50-1000bp，100-500bp，特别50、100、200、300、400、500、600、700、800、900、1000bp。

[0130] 在又一些进一步的实施方式中，由本发明的方法和试剂盒所提供的第一核酸序列编码的gRNA内的靶向序列(在本文中也称为“间隔物”序列)与包含在性别染色体Z或W内的靶基因座内的至少一个核酸序列(在本文中称为“原间隔物”)显示约70％、71％、72％、73％、74％、75％、76％、77％、78％、79％、80％、81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的同源性或同一性。

[0131] 在一些实施方式中，第二核酸序列中的至少一个报道基因可以与性别特异性启动子、胚胎特异性启动子(例如Mason等，1996中提及的α-珠蛋白启动子)和诱导型启动子(例如衍生自美国专利5750385中要求保护的大豆SSU基因的光诱导型启动子，或者如Weisshaar等，1991所述衍生自欧芹查耳酮合成酶CHS启动子的光诱导型启动子，或EL222的工程化版本，Metta-Mena等，2014引用的当用蓝光照射时激活表达的细菌光-氧-电压蛋白)可操作地连接。在更具体的实施方式中，报道基因处于胚启动子的控制下，由此限制了转基因报道基因在胚胎期表达，而在成体鸡中不表达。在此类实施方式中，为了诊断目的，使用报道转基因并且其仅在胚胎期表达。

[0132] 更具体而言，本文使用的“启动子”是指具有控制置于下游的基因表达的能力的DNA的特定区域。

[0133] 因此，“鸡中的性别特异性启动子”在下文中是指仅在特定鸡性别(即雄性或雌性)中才激活基因表达的启动子。更进一步而言，“鸡中的发育特异性启动子”是指仅在鸡发育的特定阶段激活基因表达的启动子。

[0134] 在一些具体实施方式中，可以将所述至少一个报道基因插入并由此整合在目标性别染色体的至少一个非编码区中。此类方法避免了未孵化胚胎发育和成熟可能需要的基因的破坏。

[0135] 如本文所用，“非编码区”是指不编码蛋白质序列的生物体DNA的组分。一些非编码DNA区域被转录成功能性非编码RNA分子，非编码DNA区的其它功能包括蛋白编码序列的转录和翻译调节，支架附着区域，DNA复制起点，着丝粒和端粒。假设的非功能部分(或未知功能的DNA)通常被称为“垃圾DNA”。

[0136] 在一些具体实施方式中，所述至少一个报道基因可以整合在性别W染色体上的至少一个位点。在更具体的实施方式中，W染色体中的特定基因座可以是位置1022859-1024215。在一些具体实施方式中，靶基因座可以包含如SEQ ID NO.3所示的核酸序列。

[0137] 在更具体的实施方式中，将报道基因靶向W染色体内的此类特定位置所需的至少一种gRNA可以包含如SEQ ID NO.1和2中任一个所示的核酸序列，在本文将这些gRNA分别命名为gRNA1和gRNA2。

[0138] 在又一些更具体的实施方式中，本发明用于制备转基因禽类雌性的方法采用的gRNA可以包含如SEQ ID NO.1所示的核酸序列(gRNA1)。在这种情况下，包含在所述第二核酸序列内的所述至少一个报道基因可以在其5'和3'侧接有包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂，它们分别促进其整合在W染色体中的所述特定基因座。应该意识到，这些臂在本文中也分别称为左臂和右臂。

[0139] 在又一些替代性实施方式中，用于制备本发明转基因禽类雌性的gRNA可以包含如SEQ ID NO.2所示的核酸序列(gRNA2)。在这种情况下，包含在所述第二核酸序列内的至少一个报道基因可以在其5'和3'侧接有分别如SEQ ID NO.6和7所示的氨基酸序列的同源臂。应该意识到，这些臂在本文中分别也称为左臂和右臂。

[0140] 在又一些替代性实施方式中，本发明用于制备转基因禽类动物的方法采用的至少一个报道基因可以整合在性别Z染色体上的至少一个位点。在更具体的实施方式中，Z染色体中的特定基因座可以是雌性鸡的染色体Z中如SEQ ID NO:15所示的9156874-9161874；如SEQ ID NO:16所示的27764943-27769943；如SEQ ID NO:17所示的42172748-42177748；如SEQ ID NO:18所示的63363656-63368656；和如SEQ ID NO:19所示的78777477-78782477区域中的任一个。

[0141] 在更具体的实施方式中，将报道基因靶向Z染色体内地此类特定位置所需的所述至少一个gRNA可以包含以下任一种所示的核酸序列：如SEQ ID NO.11所示的gRNA3：ACAGACCTATGATATGT；如SEQ ID NO.12所示的gRNA4:CGATTATCACTCACAAG；如SEQ ID NO.13所示的gRNA5:CTGGTTAGCATGGGGAC；如SEQ ID NO.14所示的gRNA6:GTAAAGAGTCAGATACA。

[0142] 在又一些实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.41所示的核酸序列的左臂和包含如SEQ ID NO.42所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:11的gRNA3指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.43所示的核酸序列的左臂和包含如SEQ ID NO.44所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:12的gRNA4指导的特定基因座。在更进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.45所示的核酸序列的左臂和包含如SEQ ID NO.46所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:13的gRNA5指导的特定基因座。在一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.47所示的核酸序列的左臂和包含如SEQ ID NO.48所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:14的gRNA6指导的特定基因座。

[0143] 适合于整合在Z染色体内的特定基因座中的gRNA序列的其它非限制性实例可以包括但不限于Z染色体基因座chrZ_42174515_-1的gRNA7，其包含核酸序列GTAATACAGAGCTAAACCAG，也如SEQ ID NO:26所示；Z染色体基因座chrZ_9157091_1的gRNA8，其包含核酸序列ACAGACCTATGATATGTGAG，也如SEQ ID NO:27所示；Z染色体基因座chrZ_27767602_-1的gRNA9，其包含核酸序列GAGCTTGTGAGTGATAATCG，也如SEQ ID NO:28所示；Z染色体基因座chrZ_78779927_1的gRNA10，其包含核酸序列GTAAAGAGTCAGATACACAG，也如SEQ ID NO:29所示；和Z染色体基因座chrZ_63364946_-1的gRNA11，其包含核酸序列CAGTGGGTACTGAAGCTGTG，也如SEQ ID NO:30所示。

[0144] 在又一些进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.31所示的核酸序列的左臂和包含如SEQ ID NO.32所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:26的gRNA7指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.33所示的核酸序列的左臂和包含如SEQ ID NO.34所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:27的gRNA8指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.35所示的核酸序列的左臂和包含如SEQ ID NO.36所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:28的gRNA9指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.37所示的核酸序列的左臂和包含如SEQ ID NO.38所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:29的gRNA10指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.39所示的核酸序列的左臂和包含如SEQ ID NO.40所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:30的gRNA11指导的特定基因座。

[0145] 当已经用外源序列,特别是本发明使用的报道基因靶向和/或转染禽类宿主的合子细胞的基因座时,可能需要使用此类细胞来产生转基因动物。对于这样的程序,在将靶向构建体引入胚胎干(ES)细胞中之后,可以将细胞铺在适当的培养基中的饲养层上,例如补充有生长因子和细胞因子、胎牛血清和抗生素的DMEM。胚胎干细胞可以具有单个靶向基因座(杂合子)或两个靶向基因座(纯合子)。可以通过使用选择性培养基来检测含有构建体的细胞,并且在集落生长足够的时间后,可以挑选集落并分析基因靶向的发生。在一些具体实施方式中,可以使用构建体序列内外的引物,应用PCR来验证所需外源序列在靶基因座中的整合。显示基因靶向的集落然后可以用于注射到禽类胚胎中。然后ES细胞可以被胰蛋白酶消化,并且经修饰的细胞可以通过卵侧面形成的开口注射。将卵密封后,可以在适当的条件下孵育卵,直到孵化。可以例如通过检查其生物学样品,例如血液样品来测试新孵化的禽类是否存在靶构建序列。在禽类成熟后,繁育它们并检查其子代以确定外源整合序列是否通过种系传递。

[0146] 产生嵌合禽类,其部分源自能够将遗传修饰通过种系传递的经修饰胚胎干细胞或合子细胞。将含有外源序列,特别是本发明所用的报道基因或其部分的禽类品系与其中恢复了禽类野生型基因座或其部分的品系交配,应该产生表现出卵内可检测性别的后代。

[0147] 更进一步地,转基因禽类也可以通过其它方法产生,其中一些在下面讨论。在适合转化以产生转基因动物的禽类细胞中有原始生殖细胞(PGC)、精细胞和合子细胞(包括胚胎干细胞)。精子细胞可以通过任何合适的方法用DNA构建体转化,所述方法包括电穿孔、微粒轰击,脂质转染等。精子可以用于禽类的人工授精。如上所述,可以检查授精的禽类的子代的外源序列。

[0148] 作为选择,原始生殖细胞可以从禽类卵中分离,通过任何适当的方法用本发明的外源报道基因转染,并转移或插入到新胚胎中,在胚胎中它们可以引入发育中的性腺中。如本发明所述,可以检查孵化的禽类及其子代的外源报道基因序列。

[0149] 在另一种方法中,可以通过任何合适的手段将分离自卵的分散的胚盘细胞用外源

报道基因序列或其部分转染，整合在性别特异性染色体Z或W，然后将完整卵注射到胚盘下腔中。如上所述，可以检查孵化的禽类受试对象及其子代的外源报道基因。

[0150] 鸡原始生殖细胞(PGC)是卵子和精子的前体。因此，在其一些方面，本发明提供了通过种系传递系统生产转基因鸡，所述系统使用与报道基因构建体以及与指导报道基因整合在性别特异性染色体W和Z中的CRISPR/Cas9gRNA构建体共转染的PGC。将PGC分选并转移到2.5天受体胚胎的血流中以进行种系传递

[0151] 因此，在一些具体实施方式中，“转基因禽类动物的制备”是指涉及用于生产具有基因组修饰的鸡的基因工程技术的多步骤方法，其中a)从两日龄鸡胚的血液中分离原始生殖细胞(PGC)；b)通过使用慢病毒系统，Piggybac转座子载体，TALENS或CRISPR/Cas9技术将转基因构建体整合在培养的PGC中；(c)鉴定转基因PGCs并其注射到胚胎的循环系统并迁移到发育中的性腺；d)将受体胚胎在37℃下温育直至孵化；(d)孵化的雄性被饲养至性成熟并且与野生型母鸡杂交；(e)筛选后代以鉴定来自转基因PGC的那些。

[0152] 因此，在第二方面，本发明涉及一种禽类转基因动物，其在其至少一个细胞中包含整合在性别染色体Z和W中至少一个的至少一个位置或部位(本文中也称为基因座)中的至少一个外源报道基因。

[0153] 术语“禽类”涉及来源于鸟类的任何物种，其特征在于羽毛、无齿的喙、产硬壳卵、高代谢率、四腔心脏，以及轻量但强壮的骨骼。禽类包括但不限于鸡、鹌鹑、火鸡、鸭、雉(Gallinacea sp)、鹅、野鸡和其它家禽。术语“母鸡”包括禽类物种的所有雌性。“转基因禽类”通常是指下述禽类，其已经具有异源DNA序列，或者一个或多个另外的DNA序列，所述另外的序列对于禽类正常是内源的(本文统称为“转基因”)，但是染色体整合在禽类的生殖细胞中。作为这种转移和整合的结果，转移的序列可以通过生殖细胞传递给转基因禽类的后代。转基因禽类(包括其子代)也具有整合在体细胞性染色体中的转基因。

[0154] 在一些具体的实施方式中，本发明的至少一种转基因动物可以包含至少两种不同的报道基因。在这种情况下，每个报道基因可以被整合在性别染色体Z或W之一中的至少一个位置或部位中。

[0155] 在又一些另外的实施方式中，包含在本发明的转基因动物内的报道基因可以是至少一种生物发光报道基因。

[0156] 在更具体的实施方式中，此类生物发光报道基因可以包含或可以是荧光素酶。

[0157] 在某些实施方式中，本发明提供的至少一种转基因禽类动物可以是雌性。在更具体的实施方式中，此类转基因禽鸟雌性中的至少一个报道基因可以整合在雌性染色体Z的至少一个位置中。

[0158] 在又一些替代性实施方式中，所述至少一种转基因禽类动物可以是雌性，具有整合在雌性染色体W的至少一个位置中的至少一个报道基因。

[0159] 在一些具体的实施方式中，至少一个报道基因可以使用至少一个PEN整合在本发明的转基因动物的性别染色体中。

[0160] 更具体地说，在某些实施方式中此类PEN可以是CRISPR II型系统。

[0161] 在更具体的实施方式中，所述至少一个报道基因可以通过由至少一种CRISPR/Cas9系统介导的HDR整合在本发明的转基因禽类动物的性别染色体中。

[0162] 在更具体的实施方式中，可以通过用至少两个核酸序列共转染禽类动物的至少一

个细胞或待导入所述禽类动物的至少一个细胞，从而将所述至少一个报道基因整合在本发明的转基因禽类动物的性别染色体中。更具体而言，此类细胞可以共转染有(a)至少一个第一核酸序列，其包含：编码至少一个Cas9蛋白的至少一个核酸序列和编码至少一个gRNA的至少一个核酸序列，从而提供CRISPR介导的整合；和(b)至少一个第二核酸序列，其包含：至少一个报道基因。

[0163] 在更具体的实施方式中，第二核酸序列中的至少一个报道基因可以在其5'和3'侧接有同源臂。这些臂与靶性别染色体内的整合靶位点显示同源性，从而促进整合位点的HDR。

[0164] 在还更具体的实施方式中，第二核酸序列中的至少一个报道基因可以与性别特异性启动子、胚胎特异性启动子和诱导型启动子中的任何一个可操作地连接。此类启动子应该将本发明的报道基因的表达限制为特定的期望性别(在性别特异性启动子的情况下)、特定的胚胎期(胚胎特异性启动子)或特定条件(诱导性条件)。

[0165] 在又一些进一步的具体实施方式中，包含在本发明的转基因禽类动物内的至少一个报道基因可以整合在其性别染色体之一的至少一个非编码区中。

[0166] 在某些实施方式中，至少一个报道基因可以整合在性别W染色体上的至少一个位点。在一些具体实施方式中，整合位点可以位于W染色体上的基因座1022859-1024215处，具体而言，染色体W:1022859-1024215的galGal5_dna范围。在又一些进一步的具体实施方式中，此类基因座包含如SEQ ID NO.3所示的核酸序列。

[0167] 为了在上述基因座内的任何位置处特异整合本发明的报道基因，可能需要特定的gRNA。因此，在一些具体的非限制性实施方式中，用于制备本发明的转基因禽类动物的合适的gRNA可以包含如SEQ ID NO.1和2中任一个所示的核酸序列。在一些具体实施方式中，这些gRNA在本文中分别被称为gRNA1和gRNA2。

[0168] 在一些具体实施方式中，已经使用包含如SEQ ID NO.1所示的核酸序列的gRNA1制备本发明提供的转基因禽类动物。为了能够将本发明的报道基因整合在此类特定的位置上，应该被整合的报道基因在某些实施方式中必须携带特定的臂，所述臂促进将其整合在由所用的gRNA指导的靶整合位点中。因此，在一些具体实施方式中，所述至少一个报道基因可以包含在第二核酸序列内，其中该报道基因在其5'和3'侧接有分别包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

[0169] 在又一些替代性实施方式中，已经使用包含如SEQ ID NO.2所示的核酸序列的gRNA2制备本发明提供的转基因禽类动物。在此类情况下，为了能够将本发明的报道基因整合在由所述gRNA2识别的特定位点，包含在第二核酸序列内地至少一个报道基因根据具体实施方式可以在其5'和3'侧接有分别包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

[0170] 在又一些进一步的替代性实施方式中，本发明的转基因禽类动物可以包含整合在性别Z染色体上至少一个位点中的至少一个报道基因。在一些具体的非限制性实施方式中，此类禽类转基因动物可以是携带整合在Z染色体中的转基因报道基因的雌性。在更具体的实施方式中，Z染色体中的特定基因座可以是雌性鸡染色体Z的下述区域中的任一个：如SEQ ID NO:15所示的9156874-9161874；如SEQ ID NO:16所示的27764943-27769943；如SEQ ID NO:17所示的42172748-42177748；如SEQ ID NO:18所示的63363656-63368656；和如SEQ ID NO:19所示的78777477-78782477。

[0171] 在更具体的实施方式中，将报道基因靶向Z染色体内此类特定位置所需的至少一个gRNA可以包含由下述任一种所示的核酸序列：gRNA3:ACAGACCTATGATATGT，如SEQ ID NO.11所示；gRNA4:CGATTATCACTCACAAG，如SEQ ID NO.12所示；gRNA5:CTGGTTAGCATGGGGAC，如SEQ ID NO.13所示；gRNA6:GTAAAGAGTCAGATACA，如SEQ ID NO.14所示。

[0172] 在又一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.41所示的核酸序列的左臂和如SEQ ID NO.42所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:11的gRNA3指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.43所示的核酸序列的左臂和如SEQ ID NO.44所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:12的gRNA4指导的特定基因座。在更进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.45所示的核酸序列的左臂和如SEQ ID NO.46所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:13的gRNA5指导的特定基因座。在一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.47所示的核酸序列的左臂和如SEQ ID NO.48所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:14的gRNA6指导的特定基因座。

[0173] 适合于整合在Z染色体内的特定基因座的gRNA序列的进一步非限制性实例可以包括但不限于：Z染色体基因座chrZ_42174515_-1的gRNA7，其包含也如SEQ ID NO:26所示的核酸序列GTAATACAGAGCTAAACCAG；Z染色体基因座chrZ_9157091_1的gRNA8，其包含也如SEQ ID NO:27所示的核酸序列ACAGACCTATGATATGTGAG；Z染色体基因座chrZ_27767602_-1的gRNA9，其包含也如SEQ ID NO:28所示的核酸序列GAGCTTGTGAGTGATAATCG；Z染色体基因座chrZ_78779927_1的gRNA10，其包含也如SEQ ID NO:29所示的核酸序列GTAAAGAGTCAGATACACAG；和Z染色体基因座chrZ_63364946_-1的gRNA11，其包含也如SEQ ID NO:30所示的核酸序列CAGTGGGTACTGAAGCTGTG。

[0174] 在又一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.31所示的核酸序列的左臂和如SEQ ID NO.32所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:26的gRNA7指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.33所示的核酸序列的左臂和如SEQ ID NO.34所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:27的gRNA8指导的特定基因座。在更进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.35所示的核酸序列的左臂和如SEQ ID NO.36所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:28的gRNA9指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.37所示的核酸序列的左臂和如SEQ ID NO.38所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:29的gRNA10指导的特定基因座。在更进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.39所示的核酸序列的左臂和如SEQ ID NO.40所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:30的gRNA11指导的特定基因座。

[0175] 另一方面，本发明涉及包含至少一个外源报道基因的细胞，所述外源报道基因整合在性别染色体Z和W中的至少一个中的至少一个位置或部位。

[0176] 在一些具体实施方式中，本发明提供的细胞可以是禽类细胞。

[0177] 在一些具体实施方式中，本发明提供的禽类细胞可以是原始生殖细胞(PGC)。

[0178] 术语“生殖细胞”是指与其它生殖细胞结合后发育成配子的胚胎细胞。本文所用的“原始生殖细胞(PGC)”涉及充当配子祖细胞并产生多能胚胎干细胞的生殖系干细胞。在胚胎发生过程中，被发信号变成PGC的原肠胚胎中的细胞迁移到变为性腺的生殖脊中，并分化成成熟的配子。

[0179] 在又一些进一步的实施方式中，本发明提供的细胞可以包含整合在细胞的性别染色体中的至少一个报道基因。在更具体的实施方式中，报道基因的此类特异性整合可以通过用将细胞与以下共转染来实现：(a)至少一个第一核酸序列，其包含至少一个编码至少一个Cas9蛋白的核酸序列和至少一个编码至少一个向导RNA(gRNA)的核酸序列；和(b)至少一个包含至少一个所述报道基因的第二核酸序列。

[0180] 在某些实施方式中，共转染到本发明的细胞中的第二核酸序列中的至少一个报道基因可以在其5’和3’侧接有用于在整合位点处HDR的同源臂。

[0181] 在一些具体实施方式中，为了将报道基因整合在本发明提供的细胞中的染色体W上，应该使用特定的gRNA。在进一步的具体实施方式中，gRNA可以包含如SEQ ID NO.1所示的核酸序列，其在本文被称为gRNA1。在此类情况下，包含在第二核酸序列内的至少一个报道基因可以在其5’和3’侧接有分别包含如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

[0182] 在又一些替代性实施方式中，本发明提供的细胞可以通过使用本文称为gRNA2的gRNA来制备。在一些实施方式中，gRNA2可以包含如SEQ ID NO.2所示的核酸序列。在此类具体实施方式中，包含在第二核酸序列内的至少一个报道基因可以在其5’和3’侧接有分别包含如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

[0183] 在又一些替代实施方式中，本发明提供的细胞可以通过将本发明的至少一个报道基因整合在细胞的Z染色体中来制备。在某些实施方式中，为了制备本发明的细胞，至少一个报道基因可以整合在性别Z染色体上的至少一个基因座。在更具体的实施方式中，Z染色体中的特定基因座可以是雌性鸡染色体Z的下述区域中的任一个：如SEQ ID NO:15所示的9156874-9161874；如SEQ ID NO:16所示的27764943-27769943；如SEQ ID NO:17所示的42172748-42177748；如SEQ ID NO:18所示的63363656-63368656；和如SEQ ID NO:19所示的78777477-78782477。

[0184] 在更具体的实施方式中，将报道基因靶向本发明细胞Z染色体内地此类特定位置所需的至少一个gRNA可以包含由下述任一种所示的核酸序列：gRNA3:ACAGACCTATGATATGT，如SEQ ID NO.11所示；gRNA4:CGATTATCACTCACAAG，如SEQ ID NO.12所示；gRNA5:CTGGTTAGCATGGGGAC，如SEQ ID NO.13所示；gRNA6:GTAAAGAGTCAGATACA，如SEQ ID NO.14所示。

[0185] 在又一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.41所示的核酸序列的左臂和如SEQ ID NO.42所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:11的gRNA3指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座

中,可以使用包含如SEQ ID NO.43所示的核酸序列的左臂和如SEQ ID NO.44所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:12的gRNA4指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.45所示的核酸序列的左臂和如SEQ ID NO.46所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:13的gRNA5指导的特定基因座。在一些进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.47所示的核酸序列的左臂和如SEQ ID NO.48所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:14的gRNA6指导的特定基因座。

[0186] 适合于整合在Z染色体内的特定基因座的gRNA序列的进一步非限制性实例可以包括但不限于:Z染色体基因座chrZ_42174515_-1的gRNA7,其包含也如SEQ ID NO:26所示的核酸序列GTAATACAGAGCTAAACCAG;Z染色体基因座chrZ_9157091_1的gRNA8,其包含也如SEQ ID NO:27所示的核酸序列ACAGACCTATGATATGTGAG;Z染色体基因座chrZ_27767602_-1的gRNA9,其包含也如SEQ ID NO:28所示的核酸序列GAGCTTGTGAGTGATAATCG;Z染色体基因座chrZ_78779927_1的gRNA10,其包含也如SEQ ID NO:29所示的核酸序列GTAAAGAGTCAGATACACAG;和Z染色体基因座chrZ_63364946_-1的gRNA11,其包含也如SEQ ID NO:30所示的核酸序列CAGTGGGTACTGAAGCTGTG。

[0187] 在又一些进一步的实施方式中,为了将本发明的报道基因整合在本发明细胞的Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.31所示的核酸序列的左臂和如SEQ ID NO.32所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:26的gRNA7指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.33所示的核酸序列的左臂和如SEQ ID NO.34所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:27的gRNA8指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.35所示的核酸序列的左臂和如SEQ ID NO.36所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:28的gRNA9指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.37所示的核酸序列的左臂和如SEQ ID NO.38所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:29的gRNA10指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.39所示的核酸序列的左臂和如SEQ ID NO.40所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:30的gRNA11指导的特定基因座。

[0188] 在更进一步的方面,本发明提供一种试剂盒,其包含:

[0189] (a)至少一种第一核酸序列,其包含编码至少一种Cas9蛋白的至少一种核酸序列和编码至少一种向导RNA(gRNA)的至少一种核酸序列,和(b)至少一种第二核酸序列,其包含至少一种所述报道基因。

[0190] 在某些实施方式中,在本发明试剂盒内包含的第二核酸序列中的至少一个报道基因可以在其5'和3'侧接有用于在整合位点处HDR的同源臂。

[0191] 在又一些进一步的具体实施方式中,本发明试剂盒的第二核酸序列中的至少一个报道基因与性别特异性启动子、胚胎特异性启动子和诱导型启动子中的任一种可操作地连

接。

[0192] 在一些实施方式中，所述至少一个报道基因可以整合在性别染色体、具体而言整合在染色体W上的至少一个非编码区中。在该情况下，本发明试剂盒的第一核酸序列可以编码至少一个包含如SEQ ID NO.1和2中任一个所示的核酸序列的gRNA，本文中也分别称为gRNA1和gRNA2。

[0193] 在一些具体的实施方式中，本发明试剂盒的第一核酸序列可以包含gRNA，其为gRNA1。在一些实施方式中，此类gRNA1可以包含如SEQ ID NO.1所示的核酸序列。在此类情况下，包含在本发明试剂盒的所述第二核酸序列内的报道基因可以在其5'和3'侧接有包含分别如SEQ ID NO.4和5所示的氨基酸序列的同源臂。

[0194] 在又一些替代性实施方式中，本发明的试剂盒可以在其第一核酸序列中包含编码gRNA2的序列。在一些具体实施方式中，此类序列编码如SEQ ID NO.2所示的核酸序列。在又一些进一步的实施方式中，在本发明试剂盒的第二核酸序列内包含的至少一个报道基因可以在其5'和3'侧接有包含分别如SEQ ID NO.6和7所示的氨基酸序列的同源臂。

[0195] 在又一些进一步的替代性实施方式中，至少一个报道基因可以整合在性别Z染色体上的至少一个位点中。在更具体的实施方式中，Z染色体中的特定基因座可以是雌性鸡染色体Z的下述区域中的任一个：如SEQ ID NO:15所示的9156874-9161874；如SEQ ID NO:16所示的27764943-27769943；如SEQ ID NO:17所示的42172748-42177748；如SEQ ID NO:18所示的63363656-63368656；和如SEQ ID NO:19所示的78777477-78782477。

[0196] 因此，在更具体的实施方式中，本发明试剂盒的第一核酸序列可以包含gRNA，所述gRNA是以下中的任一个：gRNA3:ACAGACCTATGATATGT，如SEQ ID NO.11所示；gRNA4:CGATTATCACTCACAAG，如SEQ ID NO.12所示；gRNA5:CTGGTTAGCATGGGGAC，如SEQ ID NO.13所示；gRNA6:GTAAAGAGTCAGATACA，如SEQ ID NO.14所示。

[0197] 在进一步的实施方式中，在本发明试剂盒的第二核酸序列内包含的至少一个报道基因可以在其5'和3'侧接有包含如SEQ ID NO.41-48中任一个所示的氨基酸序列的同源臂。更具体而言，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.41所示的核酸序列的左臂和如SEQ ID NO.42所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:11的gRNA3指导的特定基因座。在进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.43所示的核酸序列的左臂和如SEQ ID NO.44所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:12的gRNA4指导的特定基因座。在更进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.45所示的核酸序列的左臂和如SEQ ID NO.46所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:13的gRNA5指导的特定基因座。在一些进一步的实施方式中，为了将本发明的报道基因整合在Z染色体内的特定基因座中，可以使用包含如SEQ ID NO.47所示的核酸序列的左臂和如SEQ ID NO.48所示的核酸序列的右臂，来将本发明的报道基因整合在如SEQ ID NO:14的gRNA6指导的特定基因座。

[0198] 本发明试剂盒的第一核酸序列的进一步非限制性实例可以包括gRNA，所述gRNA为以下中的任一种：Z染色体基因座chrZ_42174515_-1的gRNA7，其包含也如SEQ ID NO:26所示的核酸序列GTAATACAGAGCTAAACCAG；Z染色体基因座chrZ_9157091_1的gRNA8，其包含也

如SEQ ID NO:27所示的核酸序列ACAGACCTATGATATGTGAG;Z染色体基因座chrZ_27767602_-1的gRNA9,其包含也如SEQ ID NO:28所示的核酸序列GAGCTTGTGAGTGATAATCG;Z染色体基因座chrZ_78779927_1的gRNA10,其包含也如SEQ ID NO:29所示的核酸序列GTAAAGAGTCAGATACACAG;和Z染色体基因座chrZ_63364946_-1的gRNA11,其包含也如SEQ ID NO:30所示的核酸序列CAGTGGGTACTGAAGCTGTG。

[0199] 在进一步的实施方式中,在本发明试剂盒的第二核酸序列内包含的至少一个报道基因可以在其5'和3'侧接有包含如SEQ ID NO.31至40中任一个所示的氨基酸序列的同源臂。更具体而言,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.31所示的核酸序列的左臂和如SEQ ID NO.32所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:26的gRNA7指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.33所示的核酸序列的左臂和如SEQ ID NO.34所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:27的gRNA8指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.35所示的核酸序列的左臂和如SEQ ID NO.36所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:28的gRNA9指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.37所示的核酸序列的左臂和如SEQ ID NO.38所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:29的gRNA10指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.39所示的核酸序列的左臂和如SEQ ID NO.40所示的核酸序列的右臂,来将本发明的报道基因整合在如SEQ ID NO:30的gRNA11指导的特定基因座。

[0200] 在一些实施方式中,在本发明试剂盒的第二核酸序列内包含的报道基因可以是至少一种生物发光报道基因。

[0201] 在又一些进一步的实施方式中,本发明的试剂盒可适用于制备转基因禽类动物,所述转基因禽类动物包含整合在性别染色体Z和W中至少一个中的至少一个位置或部位中的至少一个外源报道基因。

[0202] 在一些实施方式中,本发明的方法可以使用如本文所述的本发明的任何试剂盒。

[0203] 更进一步,必须意识到,本发明的试剂盒可以进一步包含制备本发明的转基因禽类动物所需的任何试剂、缓冲液、培养基或材料。本发明的试剂盒可进一步包含说明书以及用于其不同组分的容器。

[0204] 应该意识到,在某些实施方式中,本发明的试剂盒和方法使用的寡核苷酸或多核苷酸是分离的和/或纯的分子。如本文所用,当用于指核酸时,"分离的"或"纯的"是指天然存在的序列已从其正常细胞(例如染色体)环境中移除或在非天然环境中合成(例如,人工合成)。因此,"分离的"或"纯的"序列可以在无细胞溶液中或置于不同的细胞环境中。术语"纯的"并不意味着该序列是唯一存在的核苷酸,而是它基本上不含(约90-95%纯度)与其天然相关的非核苷酸材料,因此与分离的染色体区分开。

[0205] 除非另有说明,否则本文使用的所有科技术语具有本领域通常使用的含义。本文提供的定义是为了便于理解本文经常使用的某些术语,并不意味着限制本公开的范围

[0206] 在详细描述本发明的具体方面和实施方式之前，应该理解的是，本发明不限于所描述的特定方法和实验条件，因为这些方法和条件可以变化。还应该理解的是，本文使用的术语仅用于描述特定实施方式的目的，而不是限制性的，因为本发明的范围将仅由所附权利要求限制。

[0207] 如在本说明书和所附权利要求中所使用的，除非上下文另外明确指出，否则单数形式“一种”，“一个”和“该”包括复数指代。因此，例如，提及“一种方法”包括本文所述类型的一种或多种方法和/或步骤，和/或在阅读本公开内容时对于本领域技术人员而言将变得显而易见的那些，等等。

[0208] 除非另外定义，否则本文使用的所有科技术语具有与本发明所属领域的普通技术人员通常理解的相同的含义。尽管在本发明的实践或测试时可以使用与本文所述相似或等同的任何方法和材料，但现在描述优选的方法和材料。

[0209] 在整个说明书和随后的权利要求书中，除非上下文另有要求，否则词语“包括”以及诸如“包含”和“含有”的变化将被理解为暗示包含所述整数或步骤或整数组或步骤组，但不排除任何其它整数或步骤或整数组或步骤组。更具体而言，术语“包括”、“包含”、“含有”、“带有”、“具有”及其缀合词意味着“包括但不限于”。该术语包括术语“由……组成”和“基本上由……组成”。短语“基本上由……组成”是指组合物或方法可以包括另外的成分和/或步骤，但条件是另外的成分和/或步骤不实质上改变所要求保护的组合物或方法的基本和新颖特征。

[0210] 如本文所使用的术语“约”表示可能偏离所提及的值至多1％，特别是5％，更特别是10％，更特别是15％，并且在一些情况下至多20％的值，偏离范围包括整数值，并且如果适用，非整数值也构成连续范围。如本文所用，术语“约”是指±10％。

[0211] 应该注意的是，本发明的各种实施方式可以以范围的形式呈现。应该理解的是，范围格式的描述仅仅是为了方便和简洁，并且不应该被解释为对本发明范围的不灵活的限制。因此，范围的描述应被认为已经具体公开了所有可能的子范围以及该范围内的单个数值。例如，诸如1至6的范围的描述应该被认为已经具体公开了子范围，例如1至3，1至4，1至5，2至4，2至6，3至6等，以及该范围内的单个数值，例如1、2、3、4、5和6。不管范围的宽度如何，这都适用。无论在本文何时指出数值范围，其意图包括在所指范围内的任何引用的数值(分数或整数)。第一指示数值和第二指示数值之间的“范围/范围之间”和“范围/从……起的范围”第一指示数值“至”第二指示数值在本文中可互换使用，并且意在包括第一和第二指示数值以及它们之间的所有分数和整数数值。

[0212] 提出以下实施例以便为本领域普通技术人员提供关于如何制造和使用本发明的方法和组合物的完整公开和描述，并且不旨在限制发明人所认为的其发明的范围。已经致力于确保所用数值(例如数量，温度等)的准确性，但是应该考虑到一些实验误差和偏差。除非另有说明，否则份数是重量份，分子量是平均分子量，温度是摄氏度，压力是在大气压或接近大气压。

[0213] 实例是在实施本发明的各方面时发明人所采用的技术的代表。应该理解的是，尽管这些技术是用于实践本发明的优选实施方式的实例，但根据本公开内容，本领域技术人员将意识到可以进行许多修改而不脱离本发明的精神和预期范围。

[0214] 应该意识到，为了清楚起见，在单独实施方式的上下文中描述的本发明的某些特

征也可以在单个实施方式中组合提供。相反,为了简洁起见,在单个实施方式的上下文中描述的本发明的各种特征也可以单独地或以任何合适的子组合提供,或者适合于在本发明的任何其它描述的实施方式中提供。在各种实施方式的上下文中描述的某些特征不被认为是那些实施方式的必要特征,除非该实施方式在没有那些元件的情况下不起作用。

[0215] 如上文所描述的以及如以下权利要求部分所要求保护的本发明的各种实施方式和方面在以下实例中找到实验支持。

[0216] 尽管已经公开和描述,但是应该理解,本发明不限于本文公开的具体实例、方法步骤和组合物,因为这样的方法步骤和组成可以稍微变化。还应该理解的是,这里使用的术语仅用于描述特定实施方式的目的,而不旨在限制,因为本发明的范围将仅由所附权利要求及其等同物限制。

[0217] 必须注意的是,如在本说明书和所附权利要求中所使用的,除非内容另有明确规定,否则单数形式“一种”、“一个”和“该”包括复数指代物。

[0218] 实施例

[0219] 无需进一步详述,相信本领域技术人员可以使用前面的描述最大限度地利用本发明。因此,以下优选的具体实施方式应被解释为仅仅是说明性的,而不是以任何方式限制所要求保护的发明。

[0220] 本文中未具体描述的本领域已知的标准分子生物学方案通常基本上按照Sambrook等,Molecular cloning:A laboratory manual,Cold Springs Harbor Laboratory,New-York(1989,1992)和Ausubel等,Current Protocols in Molecular Biology,John Wiley and Sons,Baltimore,Maryland(1988)进行。

[0221] 试剂

[0222] 动物:

[0223] 商业白色来亨鸡从Hendrix ISA和明尼苏达州获得。

[0224] 标记线鸡(Marker Line chickens)来自加拿大不列颠哥伦比亚省Agassiz的Pacific Agri-Food Research Centre。

[0225] 转基因小鼠CAG-luc-eGFP来自The Jackson Laboratory(目录号L2G85)。

[0226] 转基因小鼠C57BL/6-Tg(CAG-EGFP)1Osb/J来自The Jackson Laboratory(目录号00329)。

[0227] 严格按照IACUC批准的方案进行动物实验,并在Crystal Bioscience IACUC委员会的监督下进行,以确保动物在实验过程中不患病或死亡。

[0228] 载体:

[0229] Cas9 SmartNuclease™一体化标记载体订购自System Bioscience Inc.,目录号为CAS8/9xx系列。

[0230] pCMV-Gluc 2载体订购自New England Biolabs Inc.,目录号为N8081S。

[0231] 细胞系:

[0232] 雌性细胞是原鸡(Gallus gallus),鸡T淋巴细胞。

[0233] 雄性细胞是原鸡,鸡肝,订购自ATCC®Number:CRL-2118™。

[0234] 原始生殖细胞(PGC)系。

[0235] 实验步骤

[0236] 通过卵壳进行的报道基因生物发光检测

[0237] 将D-荧光素(Sigma-Aldrich Co.LLC,以色列,目录号2591-17-5)在室温下溶于DPBS至终浓度为15mg/mL。

[0238] 将量为0.1ml荧光素或盐水溶液(阴性对照)皮下注射到表达转基因萤光素酶的小鼠的颈部和肩部周围的松散皮肤中。10分钟后切下耳和尾,并将其引入鸡胚(10天)中。

[0239] 作为选择,在将0.1ml荧光素或盐水溶液直接注射到卵中之前,将从表达转基因萤光素酶的转基因小鼠切下的耳和尾引入鸡胚中。

[0240] 通过使用生物空间光子成像仪(Bio space lab,USA)观察生物发光。

[0241] 非限制性(RF)克隆

[0242] 通过应用非限制法(Peleg Y等,2010)将gRNA插入Cas9-SmartNuclease™载体中。引物订购自Sigma-Genosys(Rehovot,Israel),随后的RF反应使用Phusion聚合酶(Thermo Scientific,Hudson,NH,USA)进行。质粒纯化分别使用MEGAspin试剂盒和DNA-spin质粒DNA纯化试剂盒(Intron Biotechnology Biotechnology,Daejoen,South Korea)进行。

[0243] 细胞培养

[0244] PGO在KO-DMEM(Life Technologies)中生长,其中40%在布法罗大鼠(buffalo rat)肝细胞(BRL,ATCC)上预调理并补充有7.5%胎牛血清(Hyclone)、2.5%经照射的鸡血清、1X非必需氨基酸、2mM谷氨酰胺、1mM丙酮酸钠,0.1mMβ-巯基乙醇(全部来自Life Technologies)、4ng/ml重组人成纤维细胞生长因子、6ng/ml重组小鼠干细胞因子(均来自R&D Systems),并在BRL细胞的经照射的饲养层上生长。细胞每周传代3次到新鲜的饲养层上。

[0245] 转染

[0246] 为了对染色体W或Z的上述基因座进行稳定的转染子靶向,向5x10^6个细胞中加入15μg含有gRNA的载体和15μg含有环形荧光素酶的载体,并用V-缓冲液(lonza,Walkersville)使其至100μl体积。将细胞悬浮液转移到2mm比色皿中并经受8个350伏/100μs的方波脉冲(BTX 830电穿孔仪)。然后将细胞铺上耐新霉素的经辐照的BRL,并以每孔10^5个细胞的密度接种于48孔板中。3天后,加入40μg/ml新霉素以选择具有荧光素酶报道基因的稳定整合的细胞。

[0247] 转基因鸡的制备

[0248] 将浓缩的载体(可以是滴度约为10^7MOI的慢病毒)或质粒DNA)注射到新产卵中的25个胚胎。注射每周进行三次注射。注射后3周孵化所注射的胚胎。这些是G0鸟。孵化后立即从孵化鸡的CAM样品中提取DNA,并通过半定量PCR检测载体DNA的存在/不存在。对2-3周龄G0鸡取血并重复PCR筛选。饲养G0只鸟直至性成熟,雄性为16-20周,雌性为20-24周。公鸡从大约16周进行精液生产测试。

[0249] 对母鸡受精,每天收集受精卵。3周后G1鸡孵化,将每个单独的鸡翅膀绑扎并从壳取出鸡绒毛尿囊膜(CAM)样品。从CAM样品中提取DNA并进行PCR筛选转基因的存在,预测其为单拷贝水平。重复筛选以在2-3周后基于来自血液样品中的DNA确认并区分鸡的性别。

[0250] 在几周龄时,从G1鸟取血样以制备用于PCR分析的基因组DNA。使用G1鸟来繁殖G2。

[0251] 实施例1

[0252] 选择报道基因以在家禽中进行视觉性别识别

[0253] 为了证明视觉识别卵内家禽的性别的可行性，与荧光报道基因相比地评估了生物发光的应用。因此，首先使用表达报道基因的转基因小鼠，所述报道基因例如萤火虫萤光素酶(具有如SEQ ID NO:20所示的核酸序列;编码如SEQ ID NO:21所示的氨基酸序列)和绿色荧光蛋白(eGFP)。

[0254] 为了观察萤光素酶活性，在图1中将萤光素皮下注射到表达萤光素酶的转基因小鼠，然后切下尾和耳并通过未受精卵的卵壳中的5mm孔引入。如该图所示，在尾和耳样品(图1A,1B)中通过卵壳清楚地观察到萤光素酶可检测信号。因此，本发明人接下来检查了在卵内诱导荧光素酶反应的可行性。因此，将表达萤光素酶的转基因小鼠的耳和尾切下，通过孔引入带有10日龄鸡胚的受精卵中，随后注射荧光素。如在图2A和2B中清楚地显示，卵内萤光素酶反应成功产生了能够穿透卵壳的可检测信号。

[0255] 另一方面，使用GFP作为报道基因进行的类似实验清楚地表明，在将表达GFP的转基因小鼠的尾和耳引入鸡胚后，GFP信号不可检测到，如图3所示。

[0256] 因此进一步选择荧光素酶报道基因，特别是萤火虫萤光素酶(包含如SEQ ID NO.21所示的氨基酸序列，其由如SEQ ID NO:20所示的核酸序列编码)，以用于引入性染色体W和Z中。

[0257] 图4表示用于鉴定卵内胚胎性别的本发明方法的示意图。更具体地，提供了一种转基因雌性母鸡，其含有的性别特异性染色体(W)整合有萤光素酶报道基因。在由所述母鸡产的卵中，通过可检测信号观察到的报道基因的表达表明胚胎携带W性染色体，因此是雌性。这使得能够选择继续孵育雄性，而携带报道基因的雌性则被丢弃。这一选择可能与家禽更相关。

[0258] 图5示意性地提供了帮助卵内确定雄性胚胎的另一替代方式。更具体而言，提供携带性别特异性Z染色体且其中整合有报道基因的转基因雌性鸡，产生没有报道基因的雌性胚胎(接收母系野生型W染色体)或表达转基因萤光素酶基因的雄性胚胎(其接受母系经标记的Z染色体)。

[0259] 实施例2

[0260] 向导RNA载体的设计

[0261] 为了将萤光素酶报道基因整合在性别染色体W或Z中，选择CRISPR/Cas9介导的HDR方法。然后从两个性别染色体中寻找相关的gRNA位点。

[0262] 对包含如SEQ ID NO.3所示的核酸序列的雌性鸡的染色体W的区域1022859-1024215进行分析，以用于向导RNA设计。通过非限制克隆方案，选择、合成两个向导RNA，并将其分别克隆到含有野生型Cas9核酸酶(Horizon)的Cas9 SmartNuclease载体中，所述两个向导RNA是：gRNA1:GCACTAGGAACCAGCAGCAG，如SEQ ID NO.1所示，和gRNA2:GTAGCCCCAAGAGGGCTAGG，如SEQ ID NO.2所示。

[0263] 这两种gRNA的预测参数显示于表1中：

[0264] 表1

[0265]

gRNA参数	gRNA1	gRNA2
sgRNA设计器	0.506	0.63
sscore	0.8677	0.5323

sgRNA计分器	94.8	99.9

[0266] 为了向导RNA设计，分析雌性鸡的染色体Z的下述区域:9156874-9161874，如EQ ID NO:15所示;27764943-27769943，如SEQ ID NO:16所示;42172748-42177748，如SEQ ID NO:17所示;63363656-63368656，如SEQ ID NO:18所示;和78777477-78782477，如SEQ ID NO:19所示。通过非限制克隆方案，选择、合成四个向导RNA，并将其分别克隆到含有野生型Cas9核酸酶(Horizon)的Cas9 SmartNuclease载体中，所述四个向导RNA是:gRNA3:ACAGACCTATGATATGT，如SEQ ID NO.11所示;gRNA4:CGATTATCACTCACAAG，如SEQ ID NO.12所示;gRNA5:CTGGTTAGCATGGGGAC，如SEQ ID NO.13所示;gRNA6:GTAAAGAGTCAGATACA，如SEQ ID NO.14所示。

[0267] 适合于整合在Z染色体内的特定基因座的gRNA序列的进一步非限制性实例可以包括但不限于:Z染色体基因座chrZ_42174515_-1的gRNA7，其包含也如SEQ ID NO:26所示的核酸序列GTAATACAGAGCTAAACCAG;Z染色体基因座chrZ_9157091_1的gRNA8，其包含也如SEQ ID NO:27所示的核酸序列ACAGACCTATGATATGTGAG;Z染色体基因座chrZ_27767602_-1的gRNA9，其包含也如SEQ ID NO:28所示的核酸序列GAGCTTGTGAGTGATAATCG;Z染色体基因座chrZ_78779927_1的gRNA10，其包含也如SEQ ID NO:29所示的核酸序列GTAAAGAGTCAGATACACAG;和Z染色体基因座chrZ_63364946_-1的gRNA11，其包含也如SEQ ID NO:30所示的核酸序列CAGTGGGTACTGAAGCTGTG。

[0268] 这些gRNA具有很少的预测的脱靶位点，其中没有一个位于已知的编码序列中。

[0269] 实施例3

[0270] 萤光素酶靶向载体的设计

[0271] 将与雌性W染色体或雌性Z染色体基因座的上述适当侧接序列同源的侧接序列，引入表达萤光素酶的载体中的CMV启动子上游和新霉素抗性下游或者polyA位点下游(订制的合成DNA，Integrated DNA Technologies，Inc.，USA)。

[0272] 对于雌性W染色体，报道基因，特别是萤光素酶可以使用如SEQ ID NO.1所示的Guide 1(gRNA1)或如SEQ ID NO.2所示的Guide 2(gRNA2)进行克隆。为了使用gRNA1进行克隆，提供包含如SEQ ID NO.4所示的核酸序列的“左臂”和包含如SEQ ID NO.5所示的核酸序列的“右臂”。为了使用gRNA2进行克隆，提供包含如SEQ ID NO.6所示的核酸序列的“左臂”和包含如SEQ ID NO.7所示的核酸序列的“右臂”。

[0273] 更进一步，用于CMV启动子上游的区域的“左臂”包含如SEQ ID NO.8所示的核酸序列，用于新霉素抗性下游的区域的“右臂”可以包含如SEQ ID NO.9所示的核酸序列，或用于polyA位点下游的区域的“右臂”可以包含如SEQ ID NO.10所示的核酸序列。

[0274] 对于雌性Z染色体，荧光素酶报道基因可以使用如SEQ ID NO.11所示的gRNA3、如SEQ ID NO.12所示的gRNA4、如SEQ ID NO.13所示的gRNA5、如SEQ ID NO.14所示的gRNA6进行克隆。

[0275] 为了使用gRNA3进行克隆，提供包含如SEQ ID NO.41所示的核酸序列的“左臂”和包含如SEQ ID NO.42所示的核酸序列的“右臂”。为了使用gRNA4进行克隆，提供包含如SEQ ID NO.43所示的核酸序列的“左臂”和包含如SEQ ID NO.44所示的核酸序列的“右臂”。为了使用gRNA5进行克隆，提供包含如SEQ ID NO.45所示的核酸序列的“左臂”和包含如SEQ ID NO.46所示的核酸序列的“右臂”。为了使用gRNA6进行克隆，提供包含如SEQ ID NO.47所示

的核酸序列的“左臂”和包含如SEQ ID NO.48所示的核酸序列的“右臂”。

[0276] 在又一些进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.31所示的核酸序列的左臂和包含如SEQ ID NO.32所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:26的gRNA7指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.33所示的核酸序列的左臂和包含如SEQ ID NO.34所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:27的gRNA8指导的特定基因座。在更进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.35所示的核酸序列的左臂和包含如SEQ ID NO.36所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:28的gRNA9指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.37所示的核酸序列的左臂和包含如SEQ ID NO.38所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:29的gRNA10指导的特定基因座。在进一步的实施方式中,为了将本发明的报道基因整合在Z染色体内的特定基因座中,可以使用包含如SEQ ID NO.39所示的核酸序列的左臂和包含如SEQ ID NO.40所示的核酸序列的右臂来将本发明的报道基因整合在如SEQ ID NO:30的gRNA11指导的特定基因座。

[0277] 实施例4

[0278] CRISPR处理细胞的种系传递

[0279] 如实验程序中详述的,将上述两种载体,具体地说,gRNA/Cas9和报道基因载体共转染至PGC。在鉴定到稳定的克隆后,扩增细胞并通过PCR确认荧光素酶整合。已确认的克隆在14-16阶段(H&H)注入受体鸡胚。将注入的胚胎转移至替代壳(surrogate shell)并孵育直到在37℃下孵化。孵化后通过针对W染色体的PCR确定鸡的性别。

[0280] 雌性和雄性嵌合体生长至性成熟,并繁育成野生型雄性和雌性鸡。评估孵化鸡的萤光素酶表达,并通过PCR证实种系子代携带靶向萤光素酶。

序列表

<110> 艾吉伊特有限公司

<120> 对未孵化卵中的禽类胚胎进行性别确定的方法及其用具

<130> 2455179

<150> US 62/262,409

<151> 2015-12-03

<160> 48

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Guide 1 for W chromosome

<400> 1

gcactaggaa ccagcagcag 20

<210> 2

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Guide 2 for W chromosome

<400> 2

gtagccccaa gagggctagg 20

<210> 3

<211> 1460

<212> DNA

<213> Artificial Sequence

<220>

<223> galGal5_dna range=chrW:1022859-1024215 5'pad=0 3'pad=0 strand=+ repeat Masking=none

<400> 3

tcgggctaac attagacccg ttgtgtctta atcctgccgt cgggtataca tatactaaaa 60

gaacctcctt tccctcctat aaatcggagc gagacaaggc attctttttt tattattatt 120

ttggttgcct gcccatcaga tgtaatgaaa gttgcacagg gtagggctaa gaaggtaaag 180

gagttaaatt cccccgggag gttagagatt ccctgtattg tggttgcctg cccattagac 240

gtaacaaaag ttgtgaaggg taggactaag gatccaggta aagaggttag tcccccaggg 300

ttagagtccc tcctcagaag taaatgtaaa acataaccac ggtaattgtc gttgagtgcg 360

cttgtttgat tgttaattgt ttgtgctgtt attgtatgct gattgtttgt gttgttatta 420
tacaaagagc tcacatgcat cgttaaaatg ggcattggat cacctagggg tgagggaatt 480
ttgaaaaagt cacttttggg ctgtgtttta agtcattgga aagaaatagc tggatcccct 540
ggtggggttg cacagaagga tgatttaagt acgatggatg cctttgaggc aggcactagg 600
aaccagcagc agtggcagat ttgtgatcat ggaaagaatt ggtgggagca tacggcacca 660
ggactcagat aagatagcta gctaaggtga tgaattgtag atggtgtcag ataacgaaag 720
aagaaaaaca aatgtaataa agtgttaaat gccaatatgt tttttataat gaggcaatat 780
cctgaatggc attagtgtgg gatcaatttg gcccagatga ccctctgatt ttaatggtag 840
aaaatgataa gagagagagg gagaaaggaa gagagtgtaa aacgtgttgt tcagcatgta 900
gtattggact gagatgcttc aagcgaaaca agaatgagca ggaggaggat ttagagatgc 960
tagtagcccc aagagggcta gggggcctca aaatcatgga acccaactcc ctggagtaga 1020
agggtcaggt aatgagtcca aaagagggac tataggaacc atgccagtaa cagagaggac 1080
tcacagccag catagggtcg tgctgtaagc cctgttgagg cagttggaaa tgacagttcg 1140
gtagtagtta aggtaccctt ttaaattaca gatttaaaca tttggaaagc agctgctggc 1200
agctaccgtg atgatcctaa acgggtagct aatgcttttg aaatgatgat taaaactcag 1260
gaactggata ggaaagatat ggaatttatt atgcacatgt tgtttgatag tacagaaaaa 1320
taaatgattc actagaccac acggacccaa gtggaggatc aggtaatggc aggggttttg 1380
cttggatgta atgctggaaa ggacaagagt cctttcctgt cctggggagg atttagaggt 1440
ggttagacct tggggaggag 1460

<210> 4
<211> 592
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA1, W chromosome, "Left arm"
<400> 4
tcgggctaac attagacccg ttgtgtctta atcctgccgt cgggtataca tatactaaaa 60
gaacctcctt tccctcctat aaatcggagc gagacaaggc attctttttt tattattatt 120
ttggttgcct gcccatcaga tgtaatgaaa gttgcacagg gtagggctaa gaaggtaaag 180
gagttaaatt cccccgggag gttagagatt ccctgtattg tggttgcctg cccattagac 240
gtaacaaaag ttgtgaaggg taggactaag gatccaggta aagaggttag tcccccaggg 300
ttagagtccc tcctcagaag taaatgtaaa acataaccac ggtaattgtc gttgagtgcg 360
cttgtttgat tgttaattgt ttgtgctgtt attgtatgct gattgtttgt gttgttatta 420
tacaaagagc tcacatgcat cgttaaaatg ggcattggat cacctagggg tgagggaatt 480
ttgaaaaagt cacttttggg ctgtgtttta agtcattgga aagaaatagc tggatcccct 540
ggtggggttg cacagaagga tgatttaagt acgatggatg cctttgaggc ag 592

<210> 5
<211> 848
<212> DNA

<213> Artificial Sequence
<220>
<223> For cloning using the gRNA1, W chromosome, "Right arm"
<400> 5
tggcagattt gtgatcatgg aaagaattgg tgggagcata cggcaccagg actcagataa 60
gatagctagc taaggtgatg aattgtagat ggtgtcagat aacgaaagaa gaaaaacaaa 120
tgtaataaag tgttaaatgc caatatgttt tttataatga ggcaatatcc tgaatggcat 180
tagtgtggga tcaatttggc ccagatgacc ctctgatttt aatggtagaa aatgataaga 240
gagagaggga gaaaggaaga gagtgtaaaa cgtgttgttc agcatgtagt attggactga 300
gatgcttcaa gcgaaacaag aatgagcagg aggaggattt agagatgcta gtagccccaa 360
gagggctagg gggcctcaaa atcatggaac ccaactccct ggagtagaag ggtcaggtaa 420
tgagtccaaa agagggacta taggaaccat gccagtaaca gagaggactc acagccagca 480
tagggtcgtg ctgtaagccc tgttgaggca gttggaaatg acagttcggt agtagttaag 540
gtaccctttt aaattacaga tttaaacatt tggaaagcag ctgctggcag ctaccgtgat 600
gatcctaaac gggtagctaa tgcttttgaa atgatgatta aaactcagga actggatagg 660
aaagatatgg aatttattat gcacatgttg tttgatagta cagaaaaata aatgattcac 720
tagaccacac ggacccaagt ggaggatcag gtaatggcag gggttttgct tggatgtaat 780
gctggaaagg acaagagtcc tttcctgtcc tggggaggat ttagaggtgg ttagaccttg 840
gggaggag 848
<210> 6
<211> 962
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA2, W chromosome, "Left arm"
<400> 6
tcgggctaac attagacccg ttgtgtctta atcctgccgt cgggtataca tatactaaaa 60
gaacctcctt tccctcctat aaatcggagc gagacaaggc attctttttt tattattatt 120
ttggttgcct gcccatcaga tgtaatgaaa gttgcacagg gtagggctaa gaaggtaaag 180
gagttaaatt cccccgggag gttagagatt ccctgtattg tggttgcctg cccattagac 240
gtaacaaaag ttgtgaaggg taggactaag gatccaggta aagaggttag tcccccaggg 300
ttagagtccc tcctcagaag taaatgtaaa acataaccac ggtaattgtc gttgagtgcg 360
cttgtttgat tgttaattgt ttgtgctgtt attgtatgct gattgtttgt gttgttatta 420
tacaaagagc tcacatgcat cgttaaaatg ggcattggat cacctagggg tgagggaatt 480
ttgaaaaagt cacttttggg ctgtgtttta agtcattgga aagaaatagc tggatcccct 540
ggtggggttg cacagaagga tgatttaagt acgatggatg cctttgaggc aggcactagg 600
aaccagcagc agtggcagat ttgtgatcat ggaaagaatt ggtgggagca tacggcacca 660
ggactcagat aagatagcta gctaaggtga tgaattgtag atggtgtcag ataacgaaag 720
aagaaaaaca aatgtaataa agtgttaaat gccaatatgt tttttataat gaggcaatat 780

cctgaatggc attagtgtgg gatcaatttg gcccagatga ccctctgatt ttaatggtag 840
aaaatgataa gagagagagg gagaaaggaa gagagtgtaa aacgtgttgt tcagcatgta 900
gtattggact gagatgcttc aagcgaaaca agaatgagca ggaggaggat ttagagatgc 960
ta 962

<210> 7
<211> 478
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA2, W chromosome, "Right arm"
<400> 7
gggcctcaaa atcatggaac ccaactccct ggagtagaag ggtcaggtaa tgagtccaaa 60
agagggacta taggaaccat gccagtaaca gagaggactc acagccagca tagggtcgtg 120
ctgtaagccc tgttgaggca gttggaaatg acagttcggt agtagttaag gtacccttt 180
aaattacaga tttaaacatt tggaaagcag ctgctggcag ctaccgtgat gatcctaaac 240
gggtagctaa tgcttttgaa atgatgatta aaactcagga actggatagg aaagatatgg 300
aatttattat gcacatgttg tttgatagta cagaaaaata aatgattcac tagaccacac 360
ggacccaagt ggaggatcag gtaatggcag gggttttgct tggatgtaat gctggaaagg 420
acaagagtcc tttcctgtcc tggggaggat ttagaggtgg ttagaccttg gggaggag 478

<210> 8
<211> 669
<212> DNA
<213> Artificial Sequence
<220>
<223> A "left arm" for the region upstream to the CMV-promoter
<400> 8
aggcaaggct tgaccgacaa ttgcatgaag tcgggctaac attagacccg ttgtgtctta 60
atcctgccgt cgggtataca tatactaaaa gaacctcctt tccctcctat aaatcggagc 120
gagacaaggc attctttttt tattattatt ttggttgcct gcccatcaga tgtaatgaaa 180
gttgcacagg gtagggctaa gaaggtaaag gagttaaatt ccccccgggag gttagagatt 240
ccctgtattg tggttgcctg cccattagac gtaacaaaag ttgtgaaggg taggactaag 300
gatccaggta aagaggttag tcccccaggg ttagagtccc tcctcagaag taaatgtaaa 360
acataaccac ggtaattgtc gttgagtgcg cttgtttgat tgttaattgt ttgtgctgtt 420
attgtatgct gattgtttgt gttgttatta tacaaagagc tcacatgcat cgttaaaatg 480
ggcattggat cacctagggg tgagggaatt ttgaaaaagt cacttttggg ctgtgtttta 540
agtcattgga aagaaatagc tggatcccct ggtggggttg cacagaagga tgatttaagt 600
acgatggatg cctttgaggc aggcactagg aaccagcagc gatgtacggg ccagatatac 660
gcgttgaca 669

<210> 9

<211> 911
<212> DNA
<213> Artificial Sequence
<220>
<223> A "right arm" for the region downstream the Neomycin-resistance
<400> 9
ttctatcgcc ttcttgacga gttcttctga cagtggcaga tttgtgatca tggaaagaat 60
tggtgggagc atacggcacc aggactcaga taagatagct agctaaggtg atgaattgta 120
gatggtgtca gataacgaaa gaagaaaaac aaatgtaata aagtgttaaa tgccaatatg 180
ttttttataa tgaggcaata tcctgaatgg cattagtgtg ggatcaattt ggcccagatg 240
accctctgat tttaatggta gaaaatgata agagagagag ggagaaagga agagagtgta 300
aaacgtgttg ttcagcatgt agtattggac tgagatgctt caagcgaaac aagaatgagc 360
aggaggagga tttagagatg ctagtagccc caagagggct agggggcctc aaaatcatgg 420
aacccaactc cctggagtag aagggtcagg taatgagtcc aaaagaggga ctataggaac 480
catgccagta acagagagga ctcacagcca gcatagggtc gtgctgtaag ccctgttgag 540
gcagttggaa atgacagttc ggtagtagtt aaggtaccct tttaaattac agatttaaac 600
atttggaaag cagctgctgg cagctaccgt gatgatccta aacgggtagc taatgctttt 660
gaaatgatga ttaaaactca ggaactggat aggaaagata tggaatttat tatgcacatg 720
ttgtttgata gtacagaaaa ataaatgatt cactagacca cacggaccca agtggaggat 780
caggtaatgg caggggtttt gcttggatgt aatgctggaa aggacaagag tcctttcctg 840
tcctggggag gatttagagg tggttagacc ttggggagga gggttcgaaa tgaccgacca 900
agcgacgccc a 911
<210> 10
<211> 911
<212> DNA
<213> Artificial Sequence
<220>
<223> A "right arm" for the region downstream the polyA site
<400> 10
atctgtgtgt tggttttttg tgtgtctaga cagtggcaga tttgtgatca tggaaagaat 60
tggtgggagc atacggcacc aggactcaga taagatagct agctaaggtg atgaattgta 120
gatggtgtca gataacgaaa gaagaaaaac aaatgtaata aagtgttaaa tgccaatatg 180
ttttttataa tgaggcaata tcctgaatgg cattagtgtg ggatcaattt ggcccagatg 240
accctctgat tttaatggta gaaaatgata agagagagag ggagaaagga agagagtgta 300
aaacgtgttg ttcagcatgt agtattggac tgagatgctt caagcgaaac aagaatgagc 360
aggaggagga tttagagatg ctagtagccc caagagggct agggggcctc aaaatcatgg 420
aacccaactc cctggagtag aagggtcagg taatgagtcc aaaagaggga ctataggaac 480
catgccagta acagagagga ctcacagcca gcatagggtc gtgctgtaag ccctgttgag 540
gcagttggaa atgacagttc ggtagtagtt aaggtaccct tttaaattac agatttaaac 600

atttggaaag cagctgctgg cagctaccgt gatgatccta aacgggtagc taatgctttt 660
gaaatgatga ttaaaactca ggaactggat aggaaagata tggaatttat tatgcacatg 720
ttgtttgata gtacagaaaa ataaatgatt cactagacca cacggaccca agtggaggat 780
caggtaatgg caggggtttt gcttggatgt aatgctggaa aggacaagag tcctttcctg 840
tcctggggag gatttagagg tggttagacc ttggggagga gcaagtaaga tgcttttctg 900
tgctgcaata g 911

<210> 11
<211> 17
<212> DNA
<213> Artificial Sequence
<220>
<223> Guide 3 for Z chromosome
<400> 11
acagacctat gatatgt 17

<210> 12
<211> 17
<212> DNA
<213> Artificial Sequence
<220>
<223> Guide 4 for Z chromosome
<400> 12
cgattatcac tcacaag 17

<210> 13
<211> 17
<212> DNA
<213> Artificial Sequence
<220>
<223> Guide 5 for Z chromosome
<400> 13
ctggttagca tggggac 17

<210> 14
<211> 17
<212> DNA
<213> Artificial Sequence
<220>
<223> Guide 6 for Z chromosome
<400> 14
gtaaagagtc agataca 17

<210> 15

<211> 4980
<212> DNA
<213> Artificial Sequence
<220>
<223> ENSGALG00000025945_ENSGALG00000023622|chrZ:9156874-9161874
<400> 15

cctgcaggac atgtggacac tctgccaggt actgggggag gatcccctcc actgctccct 60
gggagatggc aaacatctag agagggacac tgcagggatg ctgcatgaac acctccagcc 120
ttctgcaggc tgtgatgggg tcagggctga aggtcaacac aagcaatgac tgcttgctga 180
caatgtgggc attgccgacc acagacctat gatatgtgag tggtgagtgg ggccaggaga 240
gcaggacagg agtgtggact cggaggcgcg ggcagaggag gtagctcaga gcatgagata 300
tattcgccag gtgtcagtgg acttctggca gtgcatactg cagagagctt tggttgctgg 360
ccatctggat gatgacaaaa tctcagccac agaccatgtg ggttggaaga gtcctcagga 420
gatcaccttt ccacctcctg ctaaagcaca ttccctacag tatgtttcac atgacaccgt 480
gctggcagat ttttaatatc atcagagaat gacactccat atcctctctg ggcagccttt 540
ttcagtgctc tgtcacccac aaagtaaagt atttgctcat gttctaatgg aacttcctat 600
gttctagttt gtgaccattg aaccttgttc tgtcactgga cagtgctgaa ggagcctggc 660
cctatccact tgactcccac actttacata tttacaagca tcgataagac ccccctcagt 720
cttctccaga ctagatagcc ccaggtctcc caaccattcc ttttatggga gatgcttcag 780
gcccctcatc acctctgtgg ccctctgctg gactgcctgc agtagttcca ggttttttt 840
tttgaactgg ggagcccaga acttgacacg gtattccaga tgtggcctca ccagcgcaga 900
ggagagggga aggctcacct ccctcgacct ggtggccatg ttttttttaa cagacctcag 960
gataccactg gccttcttgg cccaagggca cactgctggc tcatggccca tgggttggct 1020
tccaggactc tcaggtcctt ctccgtagag ctcctctcaa gcaggacaac cctcagcctg 1080
tactgatgtg tgcggtgcgg ttattcctcc ccacgtgcaa gactctacat ctgccattgt 1140
taaatctcat aagcttcctc tctgcccaac tctccactgt gtacgggcct tagtgaatgg 1200
cagcacagac ttctggtgtg aggatagcat tcagaacaag ctgctgtacc ttcccagtca 1260
cagaggtgag ggtgactggc ttgtactttc ccagctttgc cttcttgccc ttttggaaga 1320
ctggagcgac attggccttc tggaggaaac ggtcctcagg cacccctcct gttctccatg 1380
acctttgaaa gatgatcgag agcaccttgg cagtcacttc tgccagctcc ttcagcacgc 1440
atgagtgcat cccgccagtg cccatggatt tgtgttcatt gaattagcct aggtgacctc 1500
tgatcagatc cttcttgacc aaggggaaga cttcctttcc ccaggttctt atctccaggg 1560
cctgggattc ccaaggggca atcttagcag gaaaggatga agtaaaggag gtgtgcagtt 1620
acagtgcttt catggtgtcc tccattactc caatgaggtc acccacctca ttcagcagca 1680
gacatacatt ctcccagtct tccttttttct actgccatac taaagtagtc cttcatgttg 1740
tccttgactg aaggtcaccc aacttgattt ccagtgagcc ttagcattcc ttgtagcatt 1800
catgcatgtc ctgacaacat tcctattttc ttcccaagtg gccagagacc ttttccacac 1860
tccagaggct ttgttcttcc attactgcct ttccatgagc ttcttgctca tctctgcagg 1920
tctcatgcca cctttgcccg atttcttagt cttagagatg caccgatctt gaatttggaa 1980

gaagtggtgc ttgaatgaca accagctctc ttggtccccc ttgctttcca aagctctaac 2040
ccatggaatg cctccatgta tgtctgtgaa gaggtcaaag tcagctctcc tgaagtcgtg 2100
gttgcaatcc tacaaactgc ctgatttctc ccatgggaga tcccgaactc cactatctca 2160
cagtcactgc actcaaggct tccccccagcc ttcacatccc cagctacagg aagagaagag 2220
gctcgaggct acttgcagct gtctgcggaa ggcttctcca gcttcctcct cttcatcagg 2280
cggcctgttg taaacaccca caacagcgtc acccatcata gcctgtccct tagtcctcac 2340
ccataaactt tctactcatt catcatgcat cccttggcag agctctgtac attccagctg 2400
ttccctcaca taaagagcaa ctccaccacc tcacaccgac gtctgtctgt cctaaaaagt 2460
acagagccat ccctgagagc attccagtca tgcgagctgc cccaccacgt ccctctgctg 2520
atcgcactga gatcgcggcc ctacgaccat ccatggatct atgactcttc ctgtttactc 2580
cccgcactgt gtgtgatgtt acacaggcac ttcagagaag aaaagagagc acaggatgcc 2640
gatttccctt ggcaggaccc caggggccgc acgcagcccc cacctctatc agcaggtctg 2700
agcaggactg gagctgcagg cagagcacag ctcagccctg agacacctcc atgctgcctc 2760
ctgcacagcc tcggccccac agccccccggg aaggagagca cgggtggggt gccgggagcc 2820
acgctgcgct ctgacacact gcgtgtcccc tgctcccagc gcgagctgca gggcccgtgc 2880
aggaacgagc tgggccccaa cgtgaccctc ggccccggtc actgctgaac gacagcgttc 2940
agctctccag gctgcaaaga gcgagccgtg aaatggagca acacgcggga ttcccgggtg 3000
tgaattccta aacggattcg aaggtgccca acaggcagcc actgcgtccg cagctgtgct 3060
tcggtacctg gagacccgcg agcccgtctg ttcccaccag cctccagcag ccggtgctcg 3120
cggatcaaaa gaccgccgtg aaggccgcag ggccgtccct gggacggagc gccgcccgct 3180
tccgccgcca gccgcggctc cgccctcccc gccgagctga cgctggcagc gcgcagagcg 3240
cggagcggag ccgcggcgct ttccggcggc agcgcccgca cggggctcgc cgagggccgc 3300
gcagagcgct gtgcccggcg ggccgcgtcg ccgctccggc ggcaccgcga ttccggcacg 3360
gagcgcaacg agcgcccgct gcccggctgc aggacgcggt cggggccaca ccgcggcggg 3420
gccgcagcct gccttccgcg ggaccgcacg tccggcctcc ccgaatttgg ccccgcagcc 3480
cccgggctgt caccgtgctg tcaccgtgcc gttcacggtt tacgggctgg ttcggcccgg 3540
ttcggcccgg ttccgtgact ggcgggacag agggattcgt ggatccgcgc ctccgggaag 3600
ggaaacggga ccccgaataa ttaaaaacag cggccacgat ctgaggggtt acaaatgtac 3660
taaaacgtac taaaaacaaa cagctgctta acacctccta ttccaaaggc ccatcgacac 3720
cccttggcac cctcaaagta tccccttctg aggtaatggg taccaagaat gcacggagcc 3780
cctgggccag tcacagcagg gtgcttttgc cagtctgtac cggtgaggcc tatttcggcc 3840
acaaacacag tcaattcttg ggaaccccca ttcaccccat gaatataagt tgattctgtc 3900
ccttggtgac tcgagggcgt gacagtgcac tgtgcaccag ggtccaccag ggcctcgtat 3960
tcctgaggtt ctgatgtgcc acaccacgaa cccacacggt tcaatacact ccatcatcct 4020
ttcccccccat ggtgggggca gggtccctct atttgttacc tctgtcgttc ttagagcggt 4080
tgccatgtcc tgcagcaatg ggagcgatca cctctttggg gattaatctt ggtggttaat 4140
ctgccttgta attattttac ctgtgcttga aaatgcccct ccaggtgaaa gcaaactgca 4200
gcctgcattc tgtggccaaa ggaatggaaa agaaggcgtt agcaatgcca gtggtggcat 4260
accacttggc ttcttttgac tccagttcat actggagttc taacatgtcc agaacagcag 4320

cgctcagtgg gggtgtgact tcattctggc catggtagtc tactgtcagc ctccattctg 4380
cactggcttt atgcactgga gcacagggct tatggggagc ggcggaggga gctgggatgg 4440
ttcagtctgc agatgaggag gctcagggga gaccttattg ctctctgtaa ctccatgaag 4500
ggaggttgta gtgagctggg ggtcggtctc ttctctcatg tgactagtgg taggactaga 4560
gggaatggcc tccagttgca gcaggggaga ttcaggctgg atgttaggaa atgggctgga 4620
gtgggctgcc cggggaggtg gtggagtcac cgaccctgga ggtgctcaag gaatgtttag 4680
acaccgtgtt tagggacatg gtatagtgag aactgttggt gatgggtgga cagttgggct 4740
ggatgatctt gtaggtcttt ttaaagcttg gtgattcttt gattcgataa gtgcaaagtg 4800
ttgcatttgg gtcagagcaa tctcacgggt gtgcacaggc tgggagaaga actcctggga 4860
gaagaagaga gcagccctgc agagaaggac ttgacatgaa agacttaacg tcagccagac 4920
tccctctgct ctgccctcat gagcctctcc tactctgtaa aggtttgggc ccccagcaca 4980

<210> 16
<211> 4980
<212> DNA
<213> Artificial Sequence
<220>
<223> >ENSGALG00000015036_ENSGALG00000000438|chrZ:27764943-27769943
<400> 16

aggagtggct tctgaggctg gaaacaaggt gctcccctct tgtcaaacag ccaattccat 60
ctagctccaa cccatcctcc actggccaaa gctgagccca tcgctgatgg tgatagcatc 120
tgtaatagca tctcagaaag gctaaaagca ctgcgcagca gctgagagaa aagagtgagg 180
gaaaatggga cagaaacagc tctgcagccc tcagaggcaa tgcagaggag aggcaggagg 240
tgctccaggc atggagcatt agtttcttgc agtccaggag ttacaggtga tggtgcaggc 300
tctgtgcagc ccagtgatga ctgtttcagg atccatgctg cagcctgtgc cagagccagg 360
ggataggctt ggctgcttag tgcaggacgt ggagagccca tgctgagtgc aagctcccaa 420
aaggctagga cagaacaggt gtagtggagt agatcttgta ctgtggagca gtcttttcct 480
ggaggactaa cctgtgaaga ggaggactgc gctgaagcgg ttcatgaaga atgaaacctt 540
gtgggaggga ccccacactg aaaaaggaga aaaccacgat gaggaaggag cagcagagag 600
ctccccagcc ctgtgtgcac tgccaggaga gaaggaagta gaaaacgtgg gaatgtgcag 660
gtgtaaagct gagtctggga aaaaggtggg ggtggaataa aggtattttt agtcctatac 720
ctttctcatc aacctacttt aattgtagat tggccataaa ttaaatgttt tccccaggtc 780
aagtctgttt cttgcttgtg atggtggatg atctccctgt cctcatccca atcctgaagc 840
tatttcatca tatattctct ccctgccctg tttcaggaca gggagtgcaa gcagtttggc 900
agccagacga ggttaaccca gcaccagagg tttagtcaca tgcaagtgag agctattgcc 960
tcaacttaaa tgggagaaaa cattttagca ggttgagtct caccttgttt atccagatta 1020
tacagccatg ccagcacatg ggctgcatgc agcctggccc aacccatcct ctagccaccc 1080
cctgccccat actgaaaatg cagcggctct gccctgccaa gctgccaagc ctggctccag 1140
gcactcaccc attgcttaca aacaacaaaa cgtcagcgtg tgagcgcgcc ctgtgagctg 1200
caactggcat ggggagggca cagtcagcac aaatacatga cctgcaaatt ttcacatgga 1260

atgtatggag ttgcaatcag atactcagct caaaaaattt aatcatggct ccttacctgt 1320
ctttttttta acctttctct tagctcagag aggtatacct cacttcacaa tcatgccgta 1380
ttcatgtcat tgctttctgg cagtacgtac gtttgtgaac agctgttttc aagggtgatg 1440
cacaggaaga gtaaaatttc attaaaaaat ctccaaggag cacatggaga gctcactcag 1500
aagcctcacc gctgccactg aacaagactg atgcattagt ttcatgaaat caaggctatg 1560
tgtcccactg gtttcattat tttgttactc tttctttaaa acaaaaaaat tgttttattg 1620
cttatataca tcaactatgt tgttatatat tttccaaagg ccgctccaaa agtaatgcct 1680
cctgtttcat tctgttggcc caccgtgtca gaggcagatg gtggtacggc agtagagatg 1740
gaaccttccc accaatattc cattacatgt tgccatgtga tagacagcag cagatgggca 1800
ctctgacaga atggtatctg acatggaagc gtggatggag caaagttgtg tcactgaatt 1860
cctccataag gaaatgagca cccactaaca ttcactgatg cttgctgaac atttctggag 1920
accaaacagt ggatgtgagc acagtgaggt ggtgggtggt gcatttcagc agtggtgaca 1980
atgacgttgg gtcacctctg ctggtgcagg ttttgatgag tgcagcctgt gaggtcttgt 2040
tcatcaatgg agaaaatgca caactaatgg tggtgactgc tgaaaaatag tgttttgtag 2100
ctgagaattt tctctgtcaa acagtgctaa ttgtgctctt tgtatctgtt gtagtttcca 2160
tgggaataaa taggaggcat tacttgcaga ctgacctaca tatatgtggc cccaggcaag 2220
tcaacagctg tacgaaaagt ccatgcttcc tcattataaa cggaggaaaa aaagttgttt 2280
acagctgtaa tgggattata aaaggcaaat ggggatagta cagtggtgag aaccagatct 2340
atgaaggaaa agccctaaga aaaagagagt gcacagatac tcttaaccat ctaataactg 2400
tttcctccat cctacagctc agagttaaga cttacagagg actctagtac ttagtaagat 2460
gaatacgagc taatagtggc aaaaataatc ccagtgcctc aacactgacc tgggaaaaag 2520
gggcatgtat agaccttctg atattgtgat gctgtgtttg tacacttatt atctacattt 2580
tcagaaatta ggttaaactt cagaaaactg aagatctcca gggcttgtag cagaccctga 2640
ccaccagact ggtccccgat tatcactcac aagctctcaa accgattgtg gctgtttctc 2700
agaggcaact ctttgcgctc tagcccctct ataacatggg gcaacttccc ctgccccacc 2760
ttcccttcct gtatcttctg aaaagcttgt agccctcaat tgtcacgctc cagccatatg 2820
aatcatccca ccacgtccct gtgttcacaa ttaccttcta gtcttctaat tgcaccatgg 2880
cttccaattc ctcctgttta tcacccacgt tgcatccatt gatgtagagg cacttcagtc 2940
gggctatcag caatgttacc acctgtgagg agccctcttg agatctttta ggacaattta 3000
gaggttttcc ccactgtttc ctaaaattac agcattccct gtcccttctt agtgatatag 3060
aactccatcc ccttccacca tcaaacctag cttaagctct ggtaaatgat ccagccagtc 3120
tgcttcccaa aactcttgct cagttgcatt ctatctgatg cctacatgcc caacaccttc 3180
aaggcctgtc caagatcata gaacacaaag gcttgatcat gacaacatcc atgcagccaa 3240
tcattcacat gacccattcc tctcctcctt tccaggtccc aacggccaac tgagagttgt 3300
gctctgagct cttcagtgtc cttccgaggg atgtaaagtc ccttttaatg ttctgcattt 3360
tctttgtcac agccttctga gacccaactt gaatgtggag gagtggagag taatcatccc 3420
gctctgatta tcagatacat tagcctcttc ttggcacctc tggcagacag aagacttccc 3480
tggagagatt acctggatga taactggggg cctcagtgcc tcacagcagc aagtccccag 3540
ttactaagac tctacacctc tttggtggcg ctagttctga tgtagttctt cagttgctta 3600

acatgcttgt tgtttcccaa gatttgactg ctatctgcac cctccctttc ttccagcccc 3660
agagccttat atcatgagct gagaacattc tgtgttcaaa atacaagaca catttatgca 3720
atttcgtatt ctctacccat atttttgtga tagaaattag ggctacttat ttcccatagt 3780
gatttccatg ctgcctacca ggttctaagt caggtcaccc tgccagcttc ttccaaacca 3840
tgctcaggag tgcatcaaat cacacgcaga cccaaggaaa aaatgtacca aacaatatta 3900
ccacgcatgc aaactcctca gtagctctgt cccatatgct acaattaggt ctgtataggc 3960
ttgttagatc ttgtctactt gattagtgtt ttccagcagg cctacaatcc aacctaaact 4020
gaacagacat taaagacatg tttcactaat acaaaacaac tccacctagt tcaacagtct 4080
gctcttttgt gtatggtacc tgaaactctt ctgttcactg aacttatccc aacatttttc 4140
ctcctaatgc atacttttttc tagacgtgat gtgaaagtac agtgttatac cctggtgctg 4200
cattaatttg cataaccaca gcacacaaag cataatcata acacacaaag gcctcctatg 4260
aaaacatctg cttaaagtca taggtcactt atttattctc cctctagtta tgtaccagaa 4320
tgatgtatgt taataaataa tagattatta aatccatcaa aaattaggga agagctctag 4380
cacaaaggca aagcaaagaa gctttgaaga agtgcaagta aaatgaatta tttccaaaca 4440
actagcttct cactgtccca ctaagaacaa agactattct tatgcactgc tgtccttaaa 4500
attatcagct gctttgtgtt tatttcatgg agtaacacat gcaatatttt ttacaagaaa 4560
taagaaacca cagcagcaga agacgaatct gcttttattg aacgcagaaa tagatggaca 4620
ctgaaacaac agaagtttct accatgtttt gctgctgtga gagcaaatct gctgttgatc 4680
aactattagt tttgaaaagt atatgcaact ctgctatcaa ctccttcccc acatcacatc 4740
tcgtttcaag gagaaccttc ctgtccaggc aactagccag ggattttaat tttaacattt 4800
gctatctcag ataatctttt aatatttcta ggccgtgcac ataacccaca tagccctaac 4860
tccagatgtt cccagcatca cttagcagta gttccacttg aagcataaat acagcactgg 4920
cacaacaaat gaagacaatt caagcacgcc tttaacaagc atacagaaca tacagtttta 4980

<210> 17
<211> 4980
<212> DNA
<213> Artificial Sequence
<220>
<223> >ENSGALG00000012629_ENSGALG00000012628|chrZ:42172748-42177748
<400> 17

acaaattaaa tgtattttct cttgatacta ttttgattat aagctattaa cttgatttgt 60
aaaagcaccg atattatggt taatgtatgt atggtaactg ttgagtcaca gcctgaacct 120
ctgattgacc acctgcggaa aggtctgggt gagccctggg agcacaggca aaaacaattc 180
agttgtgcaa ctggaaggcc tctcttagac ctcatttaag ggctgactgt cactggggaa 240
ggatctgatt ctggagatcc cttattggtg aagctttccc ctgcaaacct agaatcttct 300
gatgcaagtg agcaatcttc ttcccttccc tagcagctct ctgttgcacc agtccttcca 360
tccttgcacc tttgttgtaa tgcttttccc atggcattga tctgtccagt tgctacagtg 420
tagaaacaag aaagtgtttt ctcaaaacat tatacatgct tactatcaga tcttacttaa 480
cattgataag aagtacacct ttcattacaa cagtgaacgg ataggcagca agcattcaga 540

acaatgaaaa agatctgcgt tctattgtaa agcaaacaaa ggtgatgacg gaacgtttat 600
ttgagataca ggcaccatag tacaagatat caaagaggaa ttaattcaac cctcccaaca 660
acagttggga aaaaaaaagg ataagctgag gaatgcagcc ttctccgatt tttactctgg 720
tgccggaagg aggaggagga gcgggggaaa acccagcgga ctcagactgc atacgcactc 780
tcggaagtac ctctgaggcg gggtgagagg aagggtggat ctaggctccc cccgctccac 840
acactcacgt gctcgtgtac tggttcaggc tgtgacttgg caactcctct acatctatcc 900
atctagttag actcaacaag accatcgtga ttcaggagtc tgaagatgct taactctctc 960
agaacgtcta gttctcctgg cctggtaaat gctatgtttc tcatactgcc tctctgaaga 1020
atgcttcgaa tgcttctagt gatgctctaa agttctaaac agaaaaatct ggagacagtt 1080
ctggtcttta gatagaaaaa atgccaacat gccaaaggat ggttacatcc ttcaagcaac 1140
cttgttgcat gctgtacaat agactcatgt aataacttag ccgtagtcat cgtatctctt 1200
atttacctgt tcgttattac attttcctgg tactgcttta tatttagtca gttgtccttt 1260
taagacaaat tttttggtgt gtgctaatag gcagttacca aatgttctag agggagggaa 1320
tataatcagt gagtatgtag atgtatatag atgtataacc agtaagtata cagagaagtt 1380
agggtggtct tttgagagca tatagctgga aagattatca catgggaaaa ggtaatcaga 1440
aataaatgga aaaatgctca ccagtgtcat aggctcacaa gaaaactccc caaggagcaa 1500
tctccagtaa cagacccttc ctcagtggca gtcagccctt aaatggcatc taggagaggt 1560
gtagccaggc accacccctt ctgtcacata ggtgaattgc cttcagctgt gctcctgcag 1620
ctgactcagt gttcacctca ggtgatcaat cagaggtcca gaccatgatt caacatttcc 1680
catgcggagg gagatgaaac aagactgtca tctcatcttc tgaagctgat aggagaatgg 1740
ttacaataag taggtcccct tccactggtt tagctctgta ttactcagtt taaaacttta 1800
tgtagttaat ggagttccat agaatcacat cgctggaaaa tgtatgtgat agctataccc 1860
aaaaagtgac acagcatcac agcctctacc aggtcctctg cctccagaca ggaccttgag 1920
ctgtgtgatc cctttgggat gtttctgttt aagacagggg acacttacac attaggaaac 1980
tggagtcatg tcggtggagg agaaagtggc tgtagaaaag gtaacggaca agaaatccct 2040
cactggctta gctgtaaagg tcagaggaaa atcatcttct gttctccata cacatctcct 2100
gtaattaggc acctgtgccc tttaaaaaag taaagtaatc aaagagtcat ttggtgaaaa 2160
tagaagccaa acagttacaa aatttagtta atattcaaag acaaaacaac tctggttcca 2220
aatataaaac cctttgtgta aacactgcgg gaaggtgcca aagagccacc tacacaaaga 2280
gatgccacag agtaatttgc gtaccccaac aataggatca tatatgggca aaccaaaaga 2340
ccaacaaaga ccccttgcaa taatcccttg attggcagaa gcaagtgcag cgttagattt 2400
agtatgtgtg acaattctga gaataggaag gaatttcact gaaactgtgt gaagatttgt 2460
atgttaaata catataatga gggttattta gctctggggt gtgcatgcta tgtggagaga 2520
tccccatgtg cccagcactg caatagacta atgtcagctt tttaaaccat catttggttt 2580
gaagagttta ttatgatttt cagaaacaga attatgtatt cagtgtctgt acattcttac 2640
aagcctgctg ttctgtacac atgaaaaagc tacttatggt gcaaatcagt atagagaagc 2700
agttttgaca tacagtggtc atgactgggt gatctgctgt cagaaggtaa gacactggta 2760
tacagaattg caaacgtgag atgaagacct caagtgcaac tctgtcatca taccacttct 2820
cttgctatta tctattcagt gtttgttcta aatattcagg cagaaggctt ggtttcacat 2880

caggtgttac atttaagtat tctttgccca ttttttgctg tttgtatgtg taacttcaga 2940
ttctcacatt gactgtgtag tgtaaatttg caaatagatg taacagcttc tctttaacat 3000
cccatgccaa acttttaatc actttcaatc aatgttggaa atctgcagta cataaatttt 3060
attttgattc attattcaga gaatcagaga aataaggact tttttagagg tttatgttta 3120
ttttagaaaa tataaataaa tcttgtatac atagatttcc aactagtaaa atttttagct 3180
atcgctgtag catttctatg aggagtttta ctttttactg tctcttatga aacttagtgg 3240
aaacaaagct aggtttcata aaacgtctaa tctcatattc atctctctcc aaaatatggt 3300
gaccttcaaa acaagaacag tttgcttatg taaacatttc ttacattgta ttctattaca 3360
ttatcatatt catgaagata ctttagaatt ttgttgctgt tagttctgtt gtgactggct 3420
gcttttctca gcttaagagc tctgcagaag tgatgataat atatattctc tttgctgagt 3480
tacaggaatt tcctgctttc ctctactgta cgttggagcc caggagatca gatccaacta 3540
ccagaatttt atatagctct ttgtgatcag tccttcctta gttgctgtta ttcctattga 3600
tgcaaaccta tttatggtgt attaaatgaa gaagatataa tgaggacagt ctaataatca 3660
aagttttttt ctgcagccag tagtacaggt agggtgacat gccttgcata aacagatcca 3720
gatgaaaggt gtgggtgagg caataagtct tcaggccctg ggctaaggca ggacctgttt 3780
tgtttcatga gatatgccta ggcctctcta tagacagggg gagatgtcag agaattctga 3840
agcatcagca gaggctatct tggttttcag gccctgtgct aagtatgaga ggaagttgag 3900
aacagataat gtgtgcaaca aagttagcaa catgcttttt caaaacagag gaaacgatcc 3960
ataaatactg gaacagaaca acctcacaga aatcactgac tgtccctttg ctggccatgg 4020
catcctgaag gttgtgctgg tagcactcat tgggttgctg tagctgataa ggcagtgcaa 4080
ggacattggg tctgcacctg ctatgaacag gcactggggc ctccaggagc tgctggtatt 4140
agtcatgggc atcagaggtg aagaaggcac ttgtagttct cacagctctg gcactgaccg 4200
atatttctgc tgtagttaat gggattattt gcatgtagtc aacagtttac aggctgattt 4260
ggctctgctc cgcaactagc aggacagagg gacccacaga tccatgtcct ggaaggggat 4320
aagggaaaag gatagagaga tgggcctaaa acaaaacaga agcaacgatc tgaggataaa 4380
caaagtaatt tactaaataa gatagtggaa tgcatgataa cacactatag tacaacataa 4440
tgcaatgtaa ttagaactgg agctaataaa tcaaatataa agagagtgtc tgaaagccaa 4500
aggccttact ctaatgctgt gatgagatgg cacccacaaa agagacgagc cagatgattg 4560
aaagggatgt atggtggtct accaaggcct gttatctgtt tgccctgagc agaaatgata 4620
gcagaacagc aaacagtcct ctggggaatg tagtagttct tctcttctgg gacaggtagc 4680
tggacctaga gcattaactc tttaaccccc agtacactac atgatgttat gatgtggaat 4740
accaataacc aaacatcata aaaccatgac atagtcttat gtcgctgtgt tgtaactgga 4800
agctgaggat ctccactcag tatttctgct gaatgaagca gatctttcag ttgaaggaga 4860
ccaacttaat catccacttc caacccctta ctatggacag gttcaccgtt cactagatca 4920
agctgcccag ggccccatcc aacctagcct tgaatgcctc cagggatgag gcatccacca 4980

<210> 18
<211> 4980
<212> DNA
<213> Artificial Sequence

<220>

<223> >ENSGALG00000015617_ENSGALG00000027714|chrZ:63363656-63368656

<400> 18

gaaaatctct gccactgatt gttttgcagc ttagttgtct agttattcgt gagatgttga 60
ttcaaacctt tgctcagatc aggcacagaa gcagtgctga actttgtttt tccgtagcct 120
tgctagtgtt gtaactgctg tcttagtgaa aacctgctta tttgcatccc cttcccaact 180
ttgggaaagt tgctgtctgt gattccagaa ttaggactcc ctcggttgca acatgactca 240
gattttcaga ccttggcttc agaggcacag tggtctttgg tgcagagaaa aatgaggctg 300
agaagaccag ggccatgatg agtgaggaag atgcaacaga acaagataga gaagagggaa 360
gagtttttta ggagctccac cagacatact gtagaagcag tgtagatgca cagtaccagt 420
aatgtactcc agccaaccac tcggtgaaga cccatgaaaa cctaatgagg ttcagtaagg 480
ccaagtgcag ggtgctgtac ttgggtcggg gcaatcccag gtatttatac aaactagggg 540
aagatctcct tgagagcagc cctgcagaga aggacttggg ggtcctggtg gatgagaagc 600
tggacgtgag ccagcagtgt gcacctgcag cccggaaggc caactgtgtt ctgggctgca 660
ttaaagaagg ggtgaccagc agggagaggg aggtgattgt ccccctctac tcagctcttg 720
tgaggcccca tctggagtac tgcatccagg cctggagccc ccagtacagg aaggacgtgg 780
agctcttgga gcgggtccag aggaagacca ctaagatgat cagagggctg gagcacctct 840
cctatgagga aaggttgagg gaactgtcct tgtttagctt agagaagaga aggcttcagg 900
gaaacctcag tgcagctttc cagtacttgg aaggagcaca taaacaggaa ggggaatggc 960
tgtttacgag ggtggatagt gacaggataa gggggaatgg ttttaaactg agacgggaga 1020
gttttaggtt agatattagg aggaagtttt tcacacagag ggtggtgacg cactggaaca 1080
ggttgcccaa ggaggttgtg gatgccccat ccctggaggc attcaaggcc aggctggatg 1140
tggctctggg cagcctgctc tggtggctgg caaccctcca catagcaggg gggttgaaac 1200
tgcgtgatca ttgtagtctt tttcaaccca ggccattcta tgattccatg aagactctgt 1260
aaggacccca atgccatcag ctgtccccac agcttcagta cccactgcca ccacattttt 1320
actgagatag gtagtttggt agccttcctc ccatcttgga taggggaggg ttaactgctg 1380
gcatttttgt tctttcctgc aatatgtgtt ttctgcatgc attctctttt cacccaaaat 1440
tttaatgtgg acggactttg aggatattct gcatctgccc aatatttctt acagcctcac 1500
atctactgat ttatgcacag tatcttatat ataatgtata ttatattatg tatattatgc 1560
actgaatcac atctactgtg tatgcacagt aacttctcct atggtactgt aagcccagaa 1620
atcccagata tgtcactaca acgtgtctgc tatgtattgc ttctgtgaga cacagatgtg 1680
ataatcagag gctgtacaca gtagaaatgc atacatatcc tactgtgaaa tctctgaaat 1740
ctagccttaa tcttggaaca gaaatgaata tggtgacatc ttgatctgat agaattggtt 1800
gccagtagca cagctgtaat catccatgat atgaatcaaa ccaagcacag gtaaacaggt 1860
gagagaaatc atgaacaatt acatgcaaac ggaccagcta aaatgtgttt gtttgttgtt 1920
ttttttttt caggtgattc ttgattacag taagatcaga agctgctaca ttagcagacc 1980
agccactgca ctcaaggctg tgattcacag cttgcagacc tgacagcagt tctgtggaag 2040
aggcaggtcc ctgtcaacca gtttaatcaa taaatcagtc tcgtgtacac aaataatgtt 2100
atcctgcacc actgctggtg ttacactatt tcacccaagt ttatcaccag caaactgagt 2160

cttatcgttc ctactgtgct ttgcttttct tgcttagcaa aacatatcag aaaggttttc 2220
agttaaaaaa aaaaaaaaaa aaaaaggatt atactgcagc catcaaagag ttctgttaaa 2280
gaagttgtgg ttcatattct tttcaagttc tgaagacaat ctgacaatag aggagatgag 2340
ggtttagtgg aggtaaatgg agctgtttct gaaagtaggg caccgtcgtg aggtaatctc 2400
accactagat gttgctgtaa gcatgatgag gggagaagct ggagcacaaa aggtgtaaac 2460
ccaatggttg atcatgcgca ctcggacaga ggaaatcttc aaaatgggaa agttagcttt 2520
gaactgattt ggtgggttcc cattaaaaat gggggactta ctgccttatt gtctttatgt 2580
acccatttgg gggatagatg ccatacagaa acaccagagc ttacggtgtg aattcattgc 2640
acatcttatt agatcaacaa aatggaataa aagaatacag aagattacta ctccattggg 2700
catgtggacg ttttacaggc ctggataaat tagatcttaa aaacaaacaa acaaacaaac 2760
aaaaaaactc cttgcaaaga taatgttatg taatattagt tgcaagaagt aagcaaacac 2820
acagaactgg gagcagaagc aaagcactaa gttattaaag caagttgcac attttgagtt 2880
gcattttgcc actggtttta taaacatgtt tagcatgtct ggtcagaatt tgggcaccag 2940
gatgctttta agatgtctgt ctatggaacc tgtcagtgct caagaataac ttctgttatt 3000
tggatgctgc accaaagaat tcagaggaag acgagccaag ccagacgtta tcatagtcac 3060
tagtaaagtg gttctaagcc taattaagac atgtcagaac tatgtgttgt gcaaccaaat 3120
cctccaaaag agaaatcaga ggtgaacttg tgcaataaat atagaagaca cgtaaatcct 3180
gaggcagtta gctaaccata tgaagccaat catacctgac tgctggacgc aggagactga 3240
accttacaga ccctggagaa tcactgtttg gcctagttag gcctgaatgg aatttaccaa 3300
gattcatgac tttaacatgg atcaggtgca aagaaaagaa ggctcagtta gttcctacag 3360
gctaccagat ctttttccac ctctgctcag cctggagctg tgggctctac ctgcccctga 3420
agtgaagtgg catcaactgc aacacttttt gcagaggcaa aaatcatcaa gtcgtgcctc 3480
tgcatttaa ggtgatggat tccaagagat agattccaaa cccaacactg agattcctct 3540
ggtgatgcag atgactgctt actgggatcc ttctctctct atagcctaat cccatgcagc 3600
accaaaagtt gggcacctga atactatcta tcttcctcca tctgacctct gctgctggct 3660
catcacgcat ggctcctgag cctgacagag ccatggggaa agacatggcc tgacctttc 3720
agacagcacg acttcctcca gtgtgctgca ctaacgacaa tattaaaaat aacactgcag 3780
gaaaacatgc tgttttctag caggctggag gaaaaaaaag gtgttcaagg tgagactcta 3840
taagatcact ctttccaaac acagcagtta actcttttct cacttgaagc tgaaattcat 3900
cttcagtagg tgacggttac aagcagctct ccaccggtaa gtgaaagttt gctgggtttt 3960
tttcaggcac ttcaagacac gaaggtaatt gctgcaacac cttcctagct cttcctgagt 4020
actcagccca aagctacact cttcagggct actctcctag cacggctctc tgtaattagc 4080
ctggctttct aacttatgaa ttaacgctgc agctactcaa cagggcagct ttgcagctat 4140
gaactttatg catacacttt ttattctggg tggtacaaat tcattgctaa gaaatcttat 4200
aaactctatg gaacactttt aaaagcctac atgtgattaa aaacaaaaaa acaaaaaaca 4260
ctgatatcac atgctatatt ctctagaaaa gagtattttg gaaacaattg taattcacag 4320
ctcaggagaa agagcaattg ttttcctttg tgctggagct ggaattctat gcagtataaa 4380
tacagatgtc caaacaaact cattaaatgg tagtcatagc aaatggagga aaaaaaaagg 4440
gggggggggg gggaactgaa gacactgaag aaattctgct gaaattcatg gagctgcaac 4500

aagaattgat attgattact tcaccatatt ccgctctaat taactagcac aacactgact 4560
gttgtgaaaa caagcatgtg atttgaccca ataagtgttt cttttagtaa ccaaaaccaa 4620
acaacaacaa accttctggc tgaccagatg gaggataaat tctaacctga tcattctgac 4680
acaaatgagg aattatgtag gttggaactg cattccaaag caaagcatgt ttttggctta 4740
gttgaacaca cagaattcat ttttcttgct gattcttgta actaagggca ggtatgaatt 4800
ggttccctta aaagaagtcc ccagaagcca tggtcctcaa caagtcttct cagattagca 4860
aggatgaacg actgagaaat caaaacccaa attcagatct aagagccttc aacagatttg 4920
ctaaatgaaa ctagcaaata ctgtagaagt taacttgatt ttgaaataaa acaggatttc 4980

<210> 19
<211> 4980
<212> DNA
<213> Artificial Sequence
<220>
<223> >ENSGALG00000008188_ENSGALG00000008197|chrZ:78777477-78782477
<400> 19

aaaggagttt cttgcctgaa acccccttaag attttgtgta ttgttgttgc cctggtttca 60
gctgggacag gactgatatc cttcacagta tctggtatga tgctgtgttt tggttttagg 120
agaaaaccag tgctgatggc acaccgaggt tgtagttgcg actgagcagt gctgcacaga 180
gccaaggatg tttcagctgc ccacactggc cttcctctaa gggggtgtga gggcacatgg 240
agctggaggg aaccaaaggg gtattccaca ccatatgtca ctgagcaaag acattgctgc 300
ccaggcagag aatacagtgg tagaaatacg tcacataggt gcccatctac cccagaactg 360
ggccagcgaa gaacatcaga acaacaagca tgcgaatcag gctgctcaga ttgatgtggc 420
tcaggtagat ctcaactgtc aactgtcacc cataaaatgg atgagacagg agcaaggcct 480
gcctgtccaa actgaccttg gggccccggg cacagctgct cccgagcaag gagaggcaga 540
gaattgtgga ggaaaccctc ttccacctag ctccaagggt gcacgctcaa tgcttggatg 600
aagattcaag ttctcaatga agaactgcat tcaaagacat ctttggaggt gacccagact 660
gcctctggga tgaatccaga acaaacagaa tcttcccagc aacacctttg tttatcatac 720
acgttcagaa atgcagtgcc tgcctggtat ttttttaaac tcatcacaga ggaatctctg 780
agtggagcag aggagtgttt cctcttgctt ttttcttccc cctttctgta agaaatgcat 840
atgccagttt ccccctaaat gttttcaaac tccaaattgt ttcctgctgg tgacattctc 900
ctctccagca ctgcctacac cccagctgcg tctgatagga aaagcaccca gtaccatgct 960
ggcatggcat ggtcagtgac acccactggt tagcatgggg acaaactgcc agagcagagg 1020
gctgctgaaa cgccgagtgc caactgacac tgttcagctg acagcctcac gagagtgctg 1080
cgagggttaa gtgaggcaac aaaatacaag tactcaaaaa tagaatggaa tggaatggaa 1140
taacagaggg aatggggaac aggctgcagt gagaaggaaa cccccgtgca ccgaccaacc 1200
cgctcccagt agcctggtcc ctaccttgcc ccactccagc agctccacga ggtactcgaa 1260
gtgcccatgt ccggccatcc ccagcatggc acagctgtca cctgggagca gcgggggtcc 1320
ggcaggcatc gccggcaccc acagcctggg ctgggctgag gatgggcttg gtgggcagct 1380
gggtgctgcc gtggcagctg agtgccgtcg tgcaggtggg ctcggtgctc tctgcagccc 1440

acggctcggg acccgtggtg ggcatactgc tgggcaaggg gacggctcgg atgtgcatcc 1500
cagagcaggg tgcatgatgt gtgcgcagcg gcacacagca gcgcacttgt ccccgaggca 1560
ccacccggtc tgatgaaatt ccaggttttg tttaagatga aaatgcatgt gggcatacat 1620
ttgcacatga gaacacctgc taatgaaaca cacagagttg tatgtgcagt ctccctggct 1680
ggtggcatgt ctgtgggggc agttaggcag gaagaaaagc tccaccatct gtttctgtgt 1740
cccgggcttg gtgctgccac tagagatggg atgcacacag catgcctgcg tgctgccgag 1800
gcagctcagt acctgctgct gggggagagc agccctcagg gagctgctgc aaagaccccg 1860
tgggatggct ggtgaacagt gctggagaca tggcagatgt cgcacccctg ctccacacca 1920
ccctccgctg cagctgcctg cacgatgcca tttcagcgtg ccacggcttg tttcttcctc 1980
ttgcagtctc aggttgccgc agcagcagct gcctgaggtc tggtggcaaa gggcagagca 2040
cccagcccac tgctctccat gagtgagcgg gtagggggca cgccgtgctt cgcttctggt 2100
atcgggtggc ctgtggagca catcacccct ggacagtgaa aacatgtcaa agggtgtttt 2160
ataacagtat agcatatccc tttgaggctg aatttcctga gcttatggat atggtagtaa 2220
tgctatacca ggctctctct gcaggttgct cactgaaaca atataccctt tctttctcaa 2280
gaaatgggac tcatgaatag ctaccaggcc tttctgctac tagatactgt agaattacat 2340
atatcagatg gctcattaat cacatagctt gttattggga cagaggcagg ttttcagcat 2400
ttcccacact gcttttctgc aaatggatta ggagtaaaga gtcagataca caggggacca 2460
tcactcactt cctaggccat gcttcgtagt ttaactccat gcaaggtatt ttcttgctct 2520
gatctaactc tagatcacta aatggcactt gtgcaggaca cttttccact gtcattttgg 2580
ggggtagaag tgtggtcaga gccagggagg cttgcagggc ccacaccagc tgtggcccaa 2640
gcagctgctg tacaatggct gtatggcagt atcgatgtga acttcctgat aataaacaga 2700
actcagcagc aaataaaccc agattgttct gatcagtaac gagaggctgt gcaaaaagct 2760
gagaaatgta cagcccttcc tgctcaccac tacagccctt catgcagcgg cctgctgcac 2820
tagggcctgc ttggccagag gcagggctgc agcttgagat acgcagcacc cagtacccca 2880
gcaggctgcc atgtgcttct gtcaggccat gaattgcaca agttggtata gttttataca 2940
gtgtcatgca gaagcacaac tcttaccttg gcacactgat atggaaagca aagtgtaaag 3000
gaactgtatg tttcttggct tgtgagttcc attggttctc ccaggacagc agtgacagct 3060
ctagattttc tggttgatct cttccatcta agtttttact tctgaataca tctgtgcagc 3120
aaagcctgat gctttttttt ttccccctcc cctgggatcg tgaaatatta ttttcccatc 3180
tccataatac agatgatcca ctcaaacatg catcctcagt cgcagtcctc agaagggcct 3240
tttcccacca tacacctaca cgctacacca gactttattg gtcaccatgc cttgagggat 3300
tctgcctaac acaggcacac aggaagggct gataccagta attcattgct gaggagagaa 3360
ggggcacttc ttttatagta gagccacaga atcaccagga ttggaaaaca cctccaagat 3420
catccaatcc aaccacccac ctaccaccag tatttcccac taaaccatgc cccttggtac 3480
cacatctaaa agcatcttct gcaccttcac agcttcattg cctttctcag gacgtgctcc 3540
agggccttgg tgtctttctt gtattgagtg gcccaaaatt gaacacaata ctcgaggtgc 3600
agcttcacca aagctgagtg cagaggggtg atcacctcta ctaggccttg ttttacccct 3660
tccccctgca tacctagttt aaagcccttt cagcgagccc taccagctcc tgggctagga 3720
tccatttccc cctttaagag aggtgaaacc cacccgcagg catcaggcca ggcactgagt 3780

taaactgctc catcagctaa ccaccccaat gccctgatgt cccttctgat agtcctcagg 3840
cttctctcag cgacttcatc actgccagcc tgagctatcc gtaatggata acaaccagag 3900
ggctgaacca gactagggag tctcctggta atgtccctaa gtaatgaatt catgggggaa 3960
atggatctgt gcagatgcac tgccatctca cagggctcaa ctgtcacaga cagttgtcac 4020
aaaccagagc agcagagaga gcagcaaagg agatatcata aatagtggaa aagggtgcca 4080
gtgtttgtta ttgactgctt cactgtgtgt ttggaaatca caaagaaaac aacaaaaaaa 4140
ccccacagaa tacacatttc caagcatgca tgcttcaggt aaggactgcg gcccgcttta 4200
attcccctgg aagcgtccag aggcctacac acaaatcagg ccacaacaaa caggttccgt 4260
gcaagcagtt tgtgagttgt atgaaaatag ctttcttcca taaagcgtgg cattgatggg 4320
aactgttact agcagatgaa aatctgccat gcacgacacc cttaatggtt ggagtggttc 4380
gactgagttt tgcattcctc attctgatgg atgctggcaa aaccaagatc acccaacacg 4440
taacactgaa tatagctcac caacagcaac gagccacacc agatactgag gcatctccaa 4500
ggctacgtct ggtggaaaga agagacagtc tcaccctgca gtgcaaaagc atcagggctt 4560
ctgcgtaagg tttattgatc ccaaagatag gcaagttaat attcatcagt ttaatagaca 4620
ggtactcaca ttactggaga tggcatgcca ggagagcctc tgccaagtgc tcactgccag 4680
tggcagccgg ccagtgcaca gcttcttggt gctgtggcca tctatataac agaaggaagc 4740
aattaagcat cacctatcaa gataagcaag atcattcaaa gagaataagg gtacttggtt 4800
agagctgcag agcacttcct cacgaatgag tttcttccag ctacatgcag atgttacagg 4860
cagtttgctt gtgttctagg gataacacaa acccctcctc atgggctagc tgcctgctca 4920
ggctgtggtc tgtatttttg cacgtgcact gctagccttt acttcacatt tccctttact 4980

<210> 20
<211> 1653
<212> DNA
<213> P.pyralis (firefly) luciferase gene (cds) GenBank: M15077.1
<400> 20

atggaagacg ccaaaaacat aaagaaaggc ccggcgccat tctatcctct agaggatgga 60
accgctggag agcaactgca taaggctatg aagagatacg ccctggttcc tggaacaatt 120
gcttttacag atgcacatat cgaggtgaac atcacgtacg cggaatactt cgaaatgtcc 180
gttcggttgg cagaagctat gaaacgatat gggctgaata caaatcacag aatcgtcgta 240
tgcagtgaaa actctcttca attctttatg ccggtgttgg gcgcgttatt tatcggagtt 300
gcagttgcgc ccgcgaacga catttataat gaacgtgaat tgctcaacag tatgaacatt 360
tcgcagccta ccgtagtgtt tgtttccaaa aaggggttgc aaaaaatttt gaacgtgcaa 420
aaaaaattac caataatcca gaaaattatt atcatggatt ctaaaacgga ttaccaggga 480
tttcagtcga tgtacacgtt cgtcacatct catctacctc ccggttttaa tgaatacgat 540
tttgtaccag agtcctttga tcgtgacaaa acaattgcac tgataatgaa ttcctctgga 600
tctactgggt tacctaaggg tgtggccctt ccgcatagaa ctgcctgcgt cagattctcg 660
catgccagag atcctatttt tggcaatcaa atcattccgg atactgcgat tttaagtgtt 720
gttccattcc atcacggttt tggaatgttt actacactcg gatatttgat atgtggattt 780
cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatccct tcaggattac 840

aaaattcaaa gtgcgttgct agtaccaacc ctattttcat tcttcgccaa aagcactctg 900
attgacaaat acgatttatc taatttacac gaaattgctt ctgggggcgc acctctttcg 960
aaagaagtcg gggaagcggt tgcaaaacgc ttccatcttc cagggatacg acaaggatat 1020
gggctcactg agactacatc agctattctg attacacccg agggggatga taaaccgggc 1080
gcggtcggta aagttgttcc atttttgaa gcgaaggttg tggatctgga taccgggaaa 1140
acgctgggcg ttaatcagag aggcgaatta tgtgtcagag gacctatgat tatgtccggt 1200
tatgtaaaca atccggaagc gaccaacgcc ttgattgaca aggatggatg gctacattct 1260
ggagacatag cttactggga cgaagacgaa cacttcttca tagttgaccg cttgaagtct 1320
ttaattaaat acaaaggata tcaggtggcc cccgctgaat tggaatcgat attgttacaa 1380
caccccaaca tcttcgacgc gggcgtggca ggtcttcccg acgatgacgc cggtgaactt 1440
cccgccgccg ttgttgtttt ggagcacgga aagacgatga cggaaaaaga gatcgtggat 1500
tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagttgt gtttgtggac 1560
gaagtaccga aaggtcttac cggaaaactc gacgcaagaa aaatcagaga gatcctcata 1620
aaggccaaga agggcggaaa gtccaaattg taa 1653

<210> 21
<211> 550
<212> PRT
<213> P. pyralis (firefly) luciferase protein GenBank: AAA29795.1
<400> 21

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro
1 5 10 15
Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg
20 25 30
Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
35 40 45
Val Asn Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala
50 55 60
Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val
65 70 75 80
Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
85 90 95
Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg
100 105 110
Glu Leu Leu Asn Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val
115 120 125
Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
130 135 140
Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly
145 150 155 160

Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe
165 170 175
Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile
180 185 190
Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val
195 200 205
Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp
210 215 220
Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
225 230 235 240
Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu
245 250 255
Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu
260 265 270
Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
275 280 285
Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr
290 295 300
Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser
305 310 315 320
Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
325 330 335
Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr
340 345 350
Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe
355 360 365
Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
370 375 380
Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly
385 390 395 400
Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly
405 410 415
Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
420 425 430
Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln
435 440 445
Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile
450 455 460
Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu

465 470 475 480
Pro Ala Ala Val Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys
485 490 495
Glu Ile Val Asp Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu
500 505 510
Arg Gly Gly Val Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly
515 520 525
Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys
530 535 540
Gly Gly Lys Ser Lys Leu
545 550

<210> 22
<211> 558
<212> DNA
<213> Gaussia princeps luciferase GenBank: AY015993.1
<400> 22
atgggagtga aagttctttt tgcccttatt tgtattgctg tggccgaggc caaaccaact 60
gaaaacaatg aagatttcaa cattgtagct gtagctagca actttgctac aacggatctc 120
gatgctgacc gtggtaaatt gcccggaaaa aaattaccac ttgaggtact caaagaaatg 180
gaagccaatg ctaggaaagc tggctgcact aggggatgtc tgatatgcct gtcacacatc 240
aagtgtacac ccaaaatgaa gaagtttatc ccaggaagat gccacaccta tgaaggagac 300
aaagaaagtg cacagggagg aataggagag gctattgttg acattcctga aattcctggg 360
tttaaggatt tggaacccat ggaacaattc attgcacaag ttgacctatg tgtagactgc 420
acaactggat gcctcaaagg tcttgccaat gtgcaatgtt ctgatttact caagaaatgg 480
ctgccacaaa gatgtgcaac ttttgctagc aaaattcaag gccaagtgga caaaataaag 540
ggtgccggtg gtgattaa 558

<210> 23
<211> 185
<212> PRT
<213> Gaussia princeps luciferase GenBank: AAG54095.1
<400> 23
Met Gly Val Lys Val Leu Phe Ala Leu Ile Cys Ile Ala Val Ala Glu
1 5 10 15
Ala Lys Pro Thr Glu Asn Asn Glu Asp Phe Asn Ile Val Ala Val Ala
20 25 30
Ser Asn Phe Ala Thr Thr Asp Leu Asp Ala Asp Arg Gly Lys Leu Pro
35 40 45
Gly Lys Lys Leu Pro Leu Glu Val Leu Lys Glu Met Glu Ala Asn Ala
50 55 60

Arg Lys Ala Gly Cys Thr Arg Gly Cys Leu Ile Cys Leu Ser His Ile
65 70 75 80
Lys Cys Thr Pro Lys Met Lys Lys Phe Ile Pro Gly Arg Cys His Thr
85 90 95
Tyr Glu Gly Asp Lys Glu Ser Ala Gln Gly Gly Ile Gly Glu Ala Ile
100 105 110
Val Asp Ile Pro Glu Ile Pro Gly Phe Lys Asp Leu Glu Pro Met Glu
115 120 125
Gln Phe Ile Ala Gln Val Asp Leu Cys Val Asp Cys Thr Thr Gly Cys
130 135 140
Leu Lys Gly Leu Ala Asn Val Gln Cys Ser Asp Leu Leu Lys Lys Trp
145 150 155 160
Leu Pro Gln Arg Cys Ala Thr Phe Ala Ser Lys Ile Gln Gly Gln Val
165 170 175
Asp Lys Ile Lys Gly Ala Gly Gly Asp
180 185

<210> 24
<211> 4107
<212> DNA
<213> Streptococcus pyogenes M1 GAS Cas9 protein id: AAK33936.1
<400> 24

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720
ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900
ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560
tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860
ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160
catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220
gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280
attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340
atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400
gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520
attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580
gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760
ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820
actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880
aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240
gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480
aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540
tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660
caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840
attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900
ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960
cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactga 4107

<210> 25
<211> 1368
<212> PRT
<213> Streptococcus pyogenes M1 GAS Cas9 protein id: AAK33936.1
<400> 25

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 26

<211> 20

<212> DNA

<213> Artificial Sequence

<220>
<223> gRNA7 for Z chromosome locus chrZ_42174515_-1
<400> 26
gtaatacaga gctaaaccag 20
<210> 27
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<223> gRNA8 for Z chromosome locus chrZ_9157091_1
<400> 27
acagacctat gatatgtgag 20
<210> 28
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<223> gRNA9 for Z chromosome locus chrZ_27767602_-1
<400> 28
gagcttgtga gtgataatcg 20
<210> 29
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<223> gRNA10 for Z chromosome locus chrZ_78779927_1
<400> 29
gtaaagagtc agatacacag 20
<210> 30
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<223> gRNA11 for Z chromosome locus chrZ_63364946_-1
<400> 30
cagtgggtac tgaagctgtg 20
<210> 31
<211> 1012
<212> DNA

<213> Artificial Sequence
<220>
<223> Left arm for gRNA7 integration to Chromosome Z
ZNF367_HABP4_ENSGALG00000012629_ENSGALG00000012628_chrZ_42174515_
-1
<400> 31
gtgaacggat aggcagcaag cattcagaac aatgaaaaag atctgcgttc tattgtaaag 60
caaacaaagg tgatgacgga acgtttattt gagatacagg caccatagta caagatatca 120
aagaggaatt aattcaaccc tcccaacaac agttgggaaa aaaaaaggat aagctgagga 180
atgcagcctt ctccgatttt tactctggtg ccggaaggag gaggaggagc gggggaaaac 240
ccagcggact cagactgcat acgcactctc ggaagtacct ctgaggcggg gtgagaggaa 300
gggtggatct aggctccccc cgctccacac actcacgtgc tcgtgtactg gttcaggctg 360
tgacttggca actcctctac atctatccat ctagttagac tcaacaagac catcgtgatt 420
caggagtctg aagatgctta actctctcag aacgtctagt tctcctggcc tggtaaatgc 480
tatgtttctc atactgcctc tctgaagaat gcttcgaatg cttctagtga tgctctaaag 540
ttctaaacag aaaaatctgg agacagttct ggtctttaga tagaaaaaat gccaacatgc 600
caaaggatgg ttacatcctt caagcaacct tgttgcatgc tgtacaatag actcatgtaa 660
taacttagcc gtagtcatcg tatctcttat ttacctgttc gttattacat tttcctggta 720
ctgctttata tttagtcagt tgtcctttta agacaaattt tttggtgtgt gctaataggc 780
agttaccaaa tgttctagag ggagggaata taatcagtga gtatgtagat gtatatagat 840
gtataaccag taagtataca gagaagttag ggtggtcttt tgagagcata tagctggaaa 900
gattatcaca tgggaaaagg taatcagaaa taaatggaaa aatgctcacc agtgtcatag 960
gctcacaaga aaactcccca aggagcaatc tccagtaaca gacccttcct ca 1012
<210> 32
<211> 1023
<212> DNA
<213> Artificial Sequence
<220>
<223> Right arm for gRNA7 integration to Chromosome Z
ZNF367_HABP4_ENSGALG00000012629_ENSGALG00000012628_chrZ_42174515_
-1
<400> 32
gtcatgtcgg tggaggagaa agtggctgta gaaaaggtaa cggacaagaa atccctcact 60
ggcttagctg taaaggtcag aggaaaatca tcttctgttc tccatacaca tctcctgtaa 120
ttaggcacct gtgcccttta aaaaagtaaa gtaatcaaag agtcatttgg tgaaaataga 180
agccaaacag ttacaaaatt tagttaatat tcaaagacaa aacaactctg gttccaaata 240
taaaaccctt tgtgtaaaca ctgcgggaag gtgccaaaga gccacctaca caaagagatg 300
ccacagagta atttgcgtac cccaacaata ggatcatata tgggcaaacc aaaagaccaa 360
caaagacccc ttgcaataat cccttgattg gcagaagcaa gtgcagcgtt agatttagta 420

tgtgtgacaa ttctgagaat aggaaggaat ttcactgaaa ctgtgtgaag atttgtatgt 480
taaatacata taatgagggt tatttagctc tggggtgtgc atgctatgtg gagagatccc 540
catgtgccca gcactgcaat agactaatgt cagctttta aaccatcatt tggtttgaag 600
agtttattat gattttcaga aacagaatta tgtattcagt gtctgtacat tcttacaagc 660
ctgctgttct gtacacatga aaaagctact tatggtgcaa atcagtatag agaagcagtt 720
ttgacataca gtggtcatga ctgggtgatc tgctgtcaga aggtaagaca ctggtataca 780
gaattgcaaa cgtgagatga agacctcaag tgcaactctg tcatcatacc acttctcttg 840
ctattatcta ttcagtgttt gttctaaata ttcaggcaga aggcttggtt tcacatcagg 900
tgttacattt aagtattctt tgcccatttt ttgctgtttg tatgtgtaac ttcagattct 960
cacattgact gtgtagtgta aatttgcaaa tagatgtaac agcttctctt taacatccca 1020
tgc 1023

<210> 33
<211> 1014
<212> DNA
<213> Artificial Sequence
<220>
<223> Left arm for gRNA8 integration to Chromosome Z
ENSGALG00000025945_AVD_ENSGALG00000025945_ENSGALG00000023622_chrZ
_9157091_1
<400> 33
ctctgcatgg tcacaacggc cccaacagcc tcccatcccg atgtcaccaa tcagttcact 60
attactggtg agcaaccagt ccagcactgc atccccctgg tggggctgtc tgttacgtgg 120
ctcaggaagt tatcctcagt gcattccagg tgctcctgga tcgcctgtag ctcaccgtgc 180
tacttttcca gcaggtgtca gggtggtaga agtcccccag caggacgaga gcctgcgatt 240
gtgacagctt ctgtagctgg agggagaatg cttcatcaac aggctccgct tgatcagtgt 300
ccctgtactg ataccacaag gcttcttttg ctgctcccat cttaactctc accccatcgt 360
ggtgttatac gtgcataaca tcatgcagtg cactaagagt taaagagtta atgctccagt 420
tccgtcgatt gtcaataatt ccaggtgtac ctttctcaga agagaagaac tacaaatccc 480
ttaagatctt actgttccat tttcattctc cattcagaag gaaaggtaaa actgcagtgt 540
gtgctcccta gcattagagt aatttctccc ttagctcgtt gccaccattt tgtttttagg 600
cccacctcct tacctttcc ctacacccct cctgtgtctc tgagacgtgg gtccatgggt 660
ccctcaacca tagtagacac aaactcagat caaaccaagc cagagctttg acaaccccta 720
agctttcagc ttactcaggg ttgctctgca gggactgctc ttcacactct attccttcct 780
tgatatagag ggcaatgtct cctctcctcc ttccttacct gtccctcctc atctgcctgt 840
tgctgtcagt ggccacactt cagccatggg aatcatccca ccctgctttt gtgaaggaaa 900
ctacattatg gttctcaggt agcacagtgg cttccatcac ctcctgtttg cttcccaagc 960
tgtgtgtgtt ggtgtagagg cacttcagct gggcagctga agaactggac gcac 1014

<210> 34
<211> 1032

<212> DNA
<213> Artificial Sequence
<220>
<223> Right arm for gRNA8 integration to Chromosome Z
ENSGALG00000025945_AVD_ENSGALG00000025945_ENSGALG00000023622_chrZ
_9157091_1
<400> 34
caccttcca cctcctgcta aagcacattc cctacagtat gtttcacatg acaccgtgct 60
ggcagatttt taatatcatc agagaatgac actccatatc ctctctgggc agccttttc 120
agtgctctgt cacccacaaa gtaaagtatt tgctcatgtt ctaatggaac ttcctatgtt 180
ctagtttgtg accattgaac cttgttctgt cactggacag tgctgaagga gcctggccct 240
atccacttga ctcccacact ttacatattt acaagcatcg ataagacccc cctcagtctt 300
ctccagacta gatagcccca ggtctcccaa ccattccttt tatgggagat gcttcaggcc 360
cctcatcacc tctgtggccc tctgctggac tgcctgcagt agttccaggt ttttttttt 420
gaactgggga gcccagaact tgacacggta ttccagatgt ggcctcacca gcgcagagga 480
gaggggaagg ctcacctccc tcgacctggt ggccatgttt tttttaacag acctcaggat 540
accactggcc ttcttggccc aagggcacac tgctggctca tggcccatgg gttggcttcc 600
aggactctca ggtccttctc cgtagagctc ctctcaagca ggacaaccct cagcctgtac 660
tgatgtgtgc ggtgcggtta ttcctcccca cgtgcaagac tctacatctg ccattgttaa 720
atctcataag cttcctctct gcccaactct ccactgtgta cgggccttag tgaatggcag 780
cacagacttc tggtgtgagg atagcattca gaacaagctg ctgtaccttc ccagtcacag 840
aggtgagggt gactggcttg tactttccca gctttgcctt cttgcccttt tggaagactg 900
gagcgacatt ggccttctgg aggaaacggt cctcaggcac ccctcctgtt ctccatgacc 960
tttgaaagat gatcgagagc accttggcag tcacttctgc cagctccttc agcacgcatg 1020
agtgcatccc gc 1032
<210> 35
<211> 1025
<212> DNA
<213> Artificial Sequence
<220>
<223> Left arm for gRNA9 integration to Chromosome Z
RIC1_ERMP1_ENSGALG00000015036_ENSGALG00000000438_chrZ_27767602_-1
<400> 35
gcagtacgta cgtttgtgaa cagctgtttt caagggtgat gcacaggaag agtaaaattt 60
cattaaaaaa tctccaagga gcacatggag agctcactca gaagcctcac cgctgccact 120
gaacaagact gatgcattag tttcatgaaa tcaaggctat gtgtcccact ggtttcatta 180
ttttgttact ctttctttaa aacaaaaaaa ttgttttatt gcttatatac atcaactatg 240
ttgttatata ttttccaaag gccgctccaa aagtaatgcc tcctgtttca ttctgttggc 300
ccaccgtgtc agaggcagat ggtggtacgg cagtagagat ggaaccttcc caccaatatt 360

ccattacatg ttgccatgtg atagacagca gcagatgggc actctgacag aatggtatct 420
gacatggaag cgtggatgga gcaaagttgt gtcactgaat tcctccataa ggaaatgagc 480
acccactaac attcactgat gcttgctgaa catttctgga gaccaaacag tggatgtgag 540
cacagtgagg tggtgggtgg tgcatttcag cagtggtgac aatgacgttg ggtcacctct 600
gctggtgcag gttttgatga gtgcagcctg tgaggtcttg ttcatcaatg gagaaaatgc 660
acaactaatg gtggtgactg ctgaaaaata gtgttttgta gctgagaatt ttctctgtca 720
aacagtgcta attgtgctct ttgtatctgt tgtagtttcc atgggaataa ataggaggca 780
ttacttgcag actgacctac atatatgtgg ccccaggcaa gtcaacagct gtacgaaaag 840
tccatgcttc ctcattataa acggaggaaa aaaagttgtt tacagctgta atgggattat 900
aaaaggcaaa tggggatagt acagtggtga gaaccagatc tatgaaggaa aagccctaag 960
aaaaagagag tgcacagata ctcttaacca tctaataact gtttcctcca tcctacagct 1020
cagag 1025

<210> 36
<211> 1040
<212> DNA
<213> Artificial Sequence
<220>
<223> Right arm for gRNA9 integration to Chromosome Z
RIC1_ERMP1_ENSGALG00000015036_ENSGALG00000000438_chrZ_27767602_-1
<400> 36
tggcttccaa ttcctcctgt ttatcaccca cgttgcatcc attgatgtag aggcacttca 60
gtcgggctat cagcaatgtt accacctgtg aggagccctc ttgagatctt ttaggacaat 120
ttagaggttt tccccactgt ttcctaaaat tacagcattc cctgtccctt cttagtgata 180
tagaactcca tccccttcca ccatcaaacc tagcttaagc tctggtaaat gatccagcca 240
gtctgcttcc caaaactctt gctcagttgc attctatctg atgcctacat gcccaacacc 300
ttcaaggcct gtccaagatc atagaacaca aaggcttgat catgacaaca tccatgcagc 360
caatcattca catgacccat tcctctcctc ctttccaggt cccaacggcc aactgagagt 420
tgtgctctga gctcttcagt gtccttccga gggatgtaaa gtcccttta atgttctgca 480
ttttctttgt cacagccttc tgagacccaa cttgaatgtg gaggagtgga gagtaatcat 540
cccgctctga ttatcagata cattagcctc ttcttggcac ctctggcaga cagaagactt 600
ccctggagag attacctgga tgataactgg gggcctcagt gcctcacagc agcaagtccc 660
cagttactaa gactctacac ctctttggtg gcgctagttc tgatgtagtt cttcagttgc 720
ttaacatgct tgttgtttcc caagatttga ctgctatctg caccctccct ttcttccagc 780
cccagagcct tatatcatga gctgagaaca ttctgtgttc aaaatacaag acacatttat 840
gcaatttcgt attctctacc catatttttg tgatagaaat tagggctact tatttcccat 900
agtgatttcc atgctgccta ccaggttcta agtcaggtca ccctgccagc ttcttccaaa 960
ccatgctcag gagtgcatca aatcacacgc agacccaagg aaaaaatgta ccaaacaata 1020
ttaccacgca tgcaaactcc 1040

<210> 37

<211> 1024
<212> DNA
<213> Artificial Sequence
<220>
<223> Left arm for gRNA10 integration to Chromosome Z
TRIM36_PGGT1B_ENSGALG00000008188_ENSGALG00000008197_chrZ_78779927
_1Chromosome Z
<400> 37
ctaccttgcc ccactccagc agctccacga ggtactcgaa gtgcccatgt ccggccatcc 60
ccagcatggc acagctgtca cctgggagca gcgggggtcc ggcaggcatc gccggcaccc 120
acagcctggg ctgggctgag gatgggcttg gtgggcagct gggtgctgcc gtggcagctg 180
agtgccgtcg tgcaggtggg ctcggtgctc tctgcagccc acggctcggg acccgtggtg 240
ggcatactgc tgggcaaggg gacggctcgg atgtgcatcc cagagcaggg tgcatgatgt 300
gtgcgcagcg gcacacagca gcgcacttgt ccccgaggca ccacccggtc tgatgaaatt 360
ccaggttttg tttaagatga aaatgcatgt gggcatacat ttgcacatga gaacacctgc 420
taatgaaaca cacagagttg tatgtgcagt ctccctggct ggtggcatgt ctgtgggggc 480
agttaggcag gaagaaaagc tccaccatct gtttctgtgt cccgggcttg gtgctgccac 540
tagagatggg atgcacacag catgcctgcg tgctgccgag gcagctcagt acctgctgct 600
gggggagagc agccctcagg gagctgctgc aaagaccccg tgggatggct ggtgaacagt 660
gctggagaca tggcagatgt cgcacccctg ctccacacca ccctccgctg cagctgcctg 720
cacgatgcca tttcagcgtg ccacggcttg tttcttcctc ttgcagtctc aggttgccgc 780
agcagcagct gcctgaggtc tggtggcaaa gggcagagca cccagcccac tgctctccat 840
gagtgagcgg gtagggggca cgccgtgctt cgcttctggt atcgggtggc ctgtggagca 900
catcacccct ggacagtgaa aacatgtcaa agggtgtttt ataacagtat agcatatccc 960
tttgaggctg aatttcctga gcttatggat atggtagtaa tgctatacca ggctctctct 1020
gcag 1024
<210> 38
<211> 1022
<212> DNA
<213> Artificial Sequence
<220>
<223> Right arm for gRNA10 integration to Chromosome Z
TRIM36_PGGT1B_ENSGALG00000008188_ENSGALG00000008197_chrZ_78779927
_1Chromosome Z
<400> 38
cagctgctgt acaatggctg tatggcagta tcgatgtgaa cttcctgata ataaacagaa 60
ctcagcagca aataaaccca gattgttctg atcagtaacg agaggctgtg caaaaagctg 120
agaaatgtac agcccttcct gctcaccact acagcccttc atgcagcggc ctgctgcact 180
agggcctgct tggccagagg cagggctgca gcttgagata cgcagcaccc agtaccccag 240

caggctgcca tgtgcttctg tcaggccatg aattgcacaa gttggtatag ttttatacag 300
tgtcatgcag aagcacaact cttaccttgg cacactgata tggaaagcaa agtgtaaagg 360
aactgtatgt ttcttggctt gtgagttcca ttggttctcc caggacagca gtgacagctc 420
tagattttct ggttgatctc ttccatctaa gtttttactt ctgaatacat ctgtgcagca 480
aagcctgatg ctttttttt tccccctccc ctgggatcgt gaaatattat tttcccatct 540
ccataataca gatgatccac tcaaacatgc atcctcagtc gcagtcctca gaagggcctt 600
ttcccaccat acacctacac gctacaccag actttattgg tcaccatgcc ttgagggatt 660
ctgcctaaca caggcacaca ggaagggctg ataccagtaa ttcattgctg aggagagaag 720
gggcacttct tttatagtag agccacagaa tcaccaggat tggaaaacac ctccaagatc 780
atccaatcca accacccacc taccaccagt atttcccact aaaccatgcc ccttggtacc 840
acatctaaaa gcatcttctg caccttcaca gcttcattgc ctttctcagg acgtgctcca 900
gggccttggt gtctttcttg tattgagtgg cccaaaattg aacacaatac tcgaggtgca 960
gcttcaccaa agctgagtgc agaggggtga tcacctctac taggccttgt tttacccctt 1020
cc 1022

<210> 39
<211> 1003
<212> DNA
<213> Artificial Sequence
<220>
<223> Left arm for gRNA11 integration to Chromosome Z
RPS23_ATG10_ENSGALG00000015617_ENSGALG00000027714_chrZ_63364946_-
1
<400> 39
gggttgaaac tgcgtgatca ttgtagtctt tttcaaccca ggccattcta tgattccatg 60
aagactctgt aaggacccca atgccatcag ctgtccccac agcttcagta cccactgcca 120
ccacattttt actgagatag gtagtttggt agccttcctc ccatcttgga taggggaggg 180
ttaactgctg gcatttttgt tctttcctgc aatatgtgtt ttctgcatgc attctctttt 240
cacccaaaat tttaatgtgg acggactttg aggatattct gcatctgccc aatatttctt 300
acagcctcac atctactgat ttatgcacag tatcttatat ataatgtata ttatattatg 360
tatattatgc actgaatcac atctactgtg tatgcacagt aacttctcct atggtactgt 420
aagcccagaa atcccagata tgtcactaca acgtgtctgc tatgtattgc ttctgtgaga 480
cacagatgtg ataatcagag gctgtacaca gtagaaatgc atacatatcc tactgtgaaa 540
tctctgaaat ctagccttaa tcttggaaca gaaatgaata tggtgacatc ttgatctgat 600
agaattggtt gccagtagca cagctgtaat catccatgat atgaatcaaa ccaagcacag 660
gtaaacaggt gagagaaatc atgaacaatt acatgcaaac ggaccagcta aaatgtgttt 720
gtttgttgtt ttttttttt caggtgattc ttgattacag taagatcaga agctgctaca 780
ttagcagacc agccactgca ctcaaggctg tgattcacag cttgcagacc tgacagcagt 840
tctgtggaag aggcaggtcc ctgtcaacca gtttaatcaa taaatcagtc tcgtgtacac 900
aaataatgtt atcctgcacc actgctggtg ttacactatt tcacccaagt ttatcaccag 960

caaactgagt cttatcgttc ctactgtgct ttgcttttct tgc 1003

<210> 40

<211> 1003

<212> DNA

<213> Artificial Sequence

<220>

<223> Right arm for gRNA11 integration to Chromosome Z RPS23_ATG10_ENSGALG00000015617_ENSGALG00000027714_chrZ_63364946_-1

<400> 40

cagaaacacc agagcttacg gtgtgaattc attgcacatc ttattagatc aacaaaatgg 60
aataaaagaa tacagaagat tactactcca ttgggcatgt ggacgtttta caggcctgga 120
taaattagat cttaaaaaca aacaaacaaa caaacaaaaa aactccttgc aaagataatg 180
ttatgtaata ttagttgcaa gaagtaagca aacacacaga actgggagca gaagcaaagc 240
actaagttat taaagcaagt tgcacatttt gagttgcatt ttgccactgg ttttataaac 300
atgtttagca tgtctggtca gaatttgggc accaggatgc ttttaagatg tctgtctatg 360
gaacctgtca gtgctcaaga ataacttctg ttatttggat gctgcaccaa agaattcaga 420
ggaagacgag ccaagccaga cgttatcata gtcactagta aagtggttct aagcctaatt 480
aagacatgtc agaactatgt gttgtgcaac caaatcctcc aaaagagaaa tcagaggtga 540
acttgtgcaa taaatataga agacacgtaa atcctgaggc agttagctaa ccatatgaag 600
ccaatcatac ctgactgctg gacgcaggag actgaacctt acagaccctg gagaatcact 660
gtttggccta gttaggcctg aatggaattt accaagattc atgactttaa catggatcag 720
gtgcaaagaa aagaaggctc agttagttcc tacaggctac cagatctttt tccacctctg 780
ctcagcctgg agctgtgggc tctacctgcc cctgaagtga agtggcatca actgcaacac 840
tttttgcaga ggcaaaaatc atcaagtcgt gcctctgcat tttaaggtga tggattccaa 900
gagatagatt ccaaacccaa cactgagatt cctctggtga tgcagatgac tgcttactgg 960
gatccttctc tctctatagc ctaatcccat gcagcaccaa aag 1003

<210> 41

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> For cloning using the gRNA3, Z chromosome, "Left arm"

<400> 41

cctgcaggac atgtggacac tctgccaggt actgggggag gatcccctcc actgctccct 60
gggagatggc aaacatctag agagggacac tgcagggatg ctgcatgaac acctccagcc 120
ttctgcaggc tgtgatgggg tcagggctga aggtcaacac aagcaatgac tgcttgctga 180
caatgtgggc attgccgacc 200

<210> 42

<211> 258
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA3, Z chromosome, "Right arm"
<400> 42
tggtgagtgg ggccaggaga gcaggacagg agtgtggact cggaggcgcg ggcagaggag 60
gtagctcaga gcatgagata tattcgccag gtgtcagtgg acttctggca gtgcatactg 120
cagagagctt tggttgctgg ccatctggat gatgacaaaa tctcagccac agaccatgtg 180
ggttggaaga gtcctcagga gatcaccttt ccacctcctg ctaaagcaca ttccctacag 240
tatgtttcac atgacacc 258
<210> 43
<211> 436
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA4, Z chromosome, "Left arm"
<400> 43
tcaacagctg tacgaaaagt ccatgcttcc tcattataaa cggaggaaaa aaagttgttt 60
acagctgtaa tgggattata aaaggcaaat ggggatagta cagtggtgag aaccagatct 120
atgaaggaaa agccctaaga aaaagagagt gcacagatac tcttaaccat ctaataactg 180
tttcctccat cctacagctc agagttaaga cttacagagg actctagtac ttagtaagat 240
gaatacgagc taatagtggc aaaaataatc ccagtgcctc aacactgacc tgggaaaaag 300
gggcatgtat agaccttctg atattgtgat gctgtgtttg tacacttatt atctacattt 360
tcagaaatta ggttaaactt cagaaaactg aagatctcca gggcttgtag cagaccctga 420
ccaccagact ggtccc 436
<210> 44
<211> 521
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA4, Z chromosome, "Right arm"
<400> 44
tcaaaccgat tgtggctgtt tctcagaggc aactctttgc gctctagccc ctctataaca 60
tggggcaact tcccctgccc caccttccct tcctgtatct tctgaaaagc ttgtagccct 120
caattgtcac gctccagcca tatgaatcat cccaccacgt ccctgtgttc acaattacct 180
tctagtcttc taattgcacc atggcttcca attcctcctg tttatcaccc acgttgcatc 240
cattgatgta gaggcacttc agtcgggcta tcagcaatgt taccacctgt gaggagccct 300
cttgagatct tttaggacaa tttagaggtt ttccccactg tttcctaaaa ttacagcatt 360

ccctgtccct tcttagtgat atagaactcc atccccttcc accatcaaac ctagcttaag 420
ctctggtaaa tgatccagcc agtctgcttc ccaaaactct tgctcagttg cattctatct 480
gatgcctaca tgcccaacac cttcaaggcc tgtccaagat c 521

<210> 45
<211> 505
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA5, Z chromosome, "Left arm"
<400> 45
gcctgtccaa actgaccttg gggccccggg cacagctgct cccgagcaag gagaggcaga 60
gaattgtgga ggaaaccctc ttccacctag ctccaagggt gcacgctcaa tgcttggatg 120
aagattcaag ttctcaatga agaactgcat tcaaagacat ctttggaggt gacccagact 180
gcctctggga tgaatccaga acaaacagaa tcttcccagc aacacctttg tttatcatac 240
acgttcagaa atgcagtgcc tgcctggtat ttttttaaac tcatcacaga ggaatctctg 300
agtggagcag aggagtgttt cctcttgctt ttttcttccc cctttctgta agaaatgcat 360
atgccagttt ccccctaaat gttttcaaac tccaaattgt ttcctgctgg tgacattctc 420
ctctccagca ctgcctacac cccagctgcg tctgatagga aaagcaccca gtaccatgct 480
ggcatggcat ggtcagtgac accca 505

<210> 46
<211> 599
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA5, Z chromosome, "Right arm"
<400> 46
ctgccagagc agagggctgc tgaaacgccg agtgccaact gacactgttc agctgacagc 60
ctcacgagag tgctgcgagg gttaagtgag gcaacaaaat acaagtactc aaaaatagaa 120
tggaatggaa tggaataaca gagggaatgg ggaacaggct gcagtgagaa ggaaaccccc 180
gtgcaccgac caacccgctc ccagtagcct ggtccctacc ttgccccact ccagcagctc 240
cacgaggtac tcgaagtgcc catgtccggc catccccagc atggcacagc tgtcacctgg 300
gagcagcggg ggtccggcag gcatcgccgg cacccacagc ctgggctggg ctgaggatgg 360
gcttggtggg cagctgggtg ctgccgtggc agctgagtgc cgtcgtgcag gtgggctcgg 420
tgctctctgc agcccacggc tcgggacccg tggtgggcat actgctgggc aaggggacgg 480
ctcggatgtg catcccagag cagggtgcat gatgtgtgcg cagcggcaca cagcagcgca 540
cttgtccccg aggcaccacc cggtctgatg aaattccagg ttttgtttaa gatgaaaat 599

<210> 47
<211> 453
<212> DNA

<213> Artificial Sequence
<220>
<223> For cloning using the gRNA6, Z chromosome, "Left arm"
<400> 47
ttgcagtctc aggttgccgc agcagcagct gcctgaggtc tggtggcaaa gggcagagca 60
cccagcccac tgctctccat gagtgagcgg gtaggggggca cgccgtgctt cgcttctggt 120
atcgggtggc ctgtggagca catcacccct ggacagtgaa aacatgtcaa agggtgtttt 180
ataacagtat agcatatccc tttgaggctg aatttcctga gcttatggat atggtagtaa 240
tgctatacca ggctctctct gcaggttgct cactgaaaca atataccctt tctttctcaa 300
gaaatgggac tcatgaatag ctaccaggcc tttctgctac tagatactgt agaattacat 360
atatcagatg gctcattaat cacatagctt gttattggga cagaggcagg ttttcagcat 420
ttcccacact gcttttctgc aaatggatta gga 453
<210> 48
<211> 439
<212> DNA
<213> Artificial Sequence
<220>
<223> For cloning using the gRNA6, Z chromosome, "Right arm"
<400> 48
gggaccatca ctcacttcct aggccatgct tcgtagttta actccatgca aggtattttc 60
ttgctctgat ctaactctag atcactaaat ggcacttgtg caggacactt ttccactgtc 120
attttggggg gtagaagtgt ggtcagagcc agggaggctt gcagggccca caccagctgt 180
ggcccaagca gctgctgtac aatggctgta tggcagtatc gatgtgaact tcctgataat 240
aaacagaact cagcagcaaa taaacccaga ttgttctgat cagtaacgag aggctgtgca 300
aaaagctgag aaatgtacag cccttcctgc tcaccactac agcccttcat gcagcggcct 360
gctgcactag ggcctgcttg gccagaggca gggctgcagc ttgagatacg cagcacccag 420
taccccagca ggctgccat 439

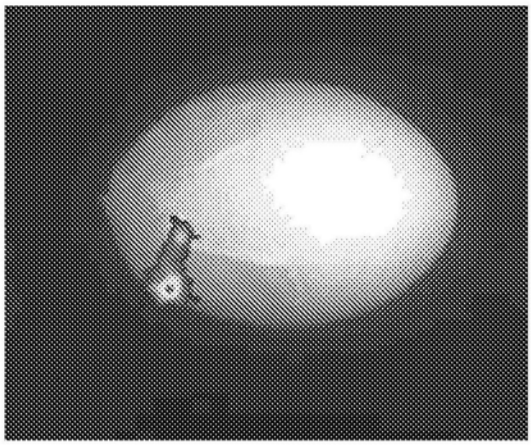


图1A

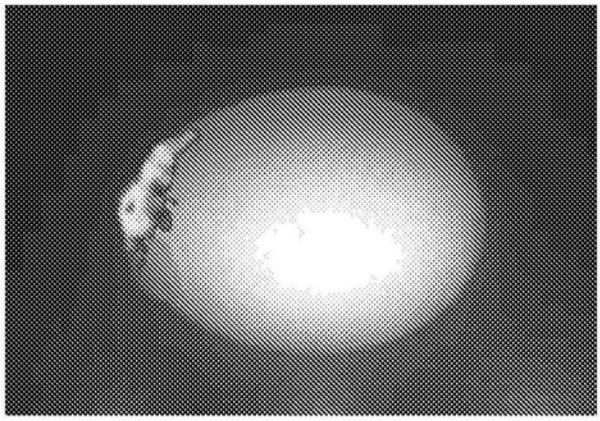


图1B

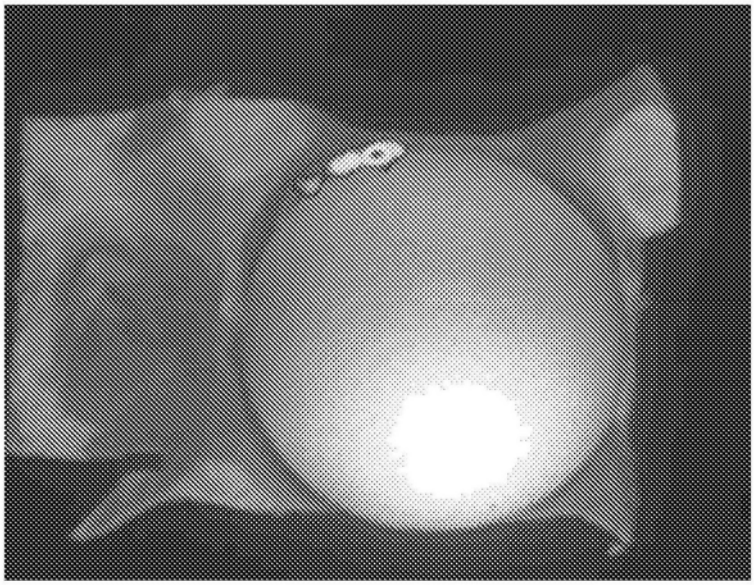


图2A

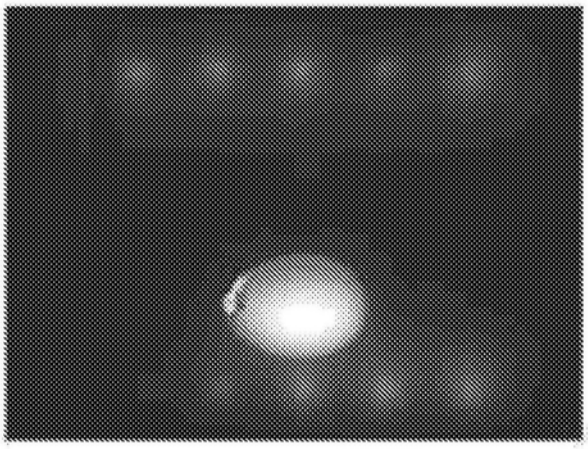


图2B



图3A

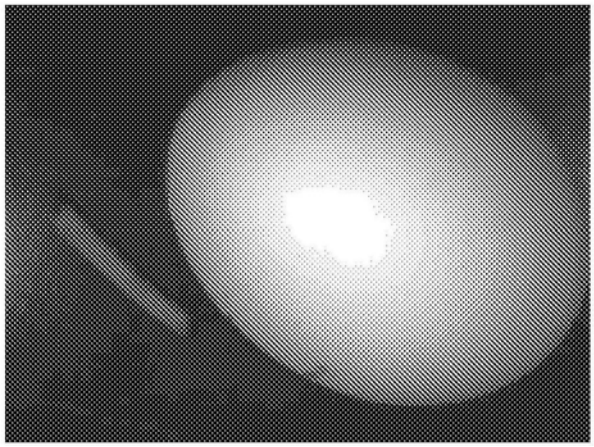


图3B

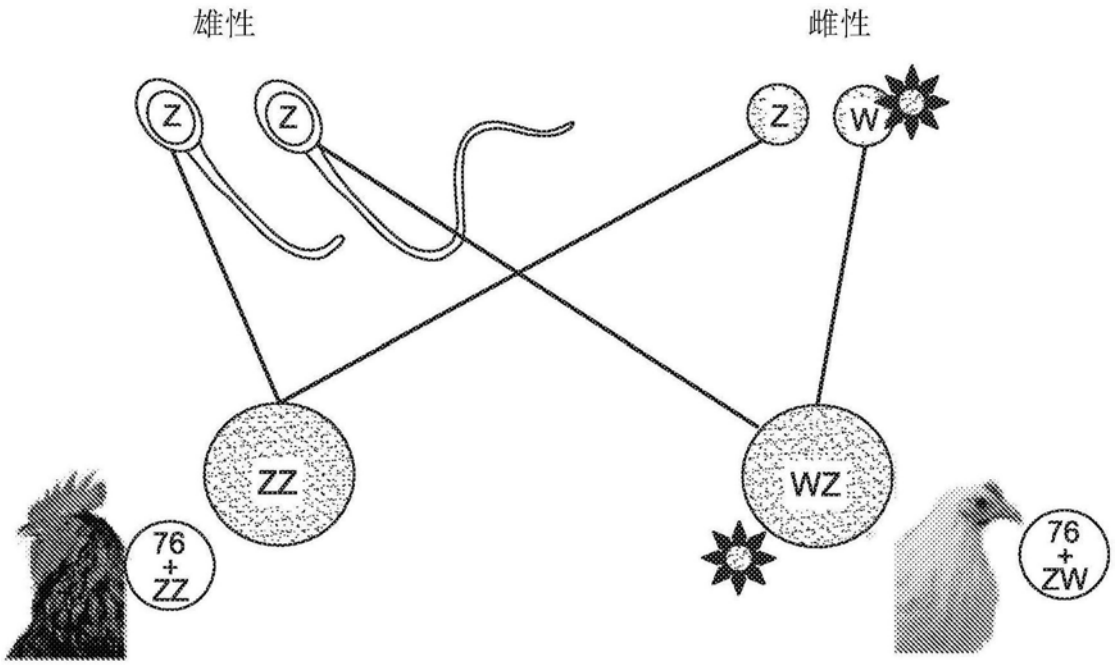


图4

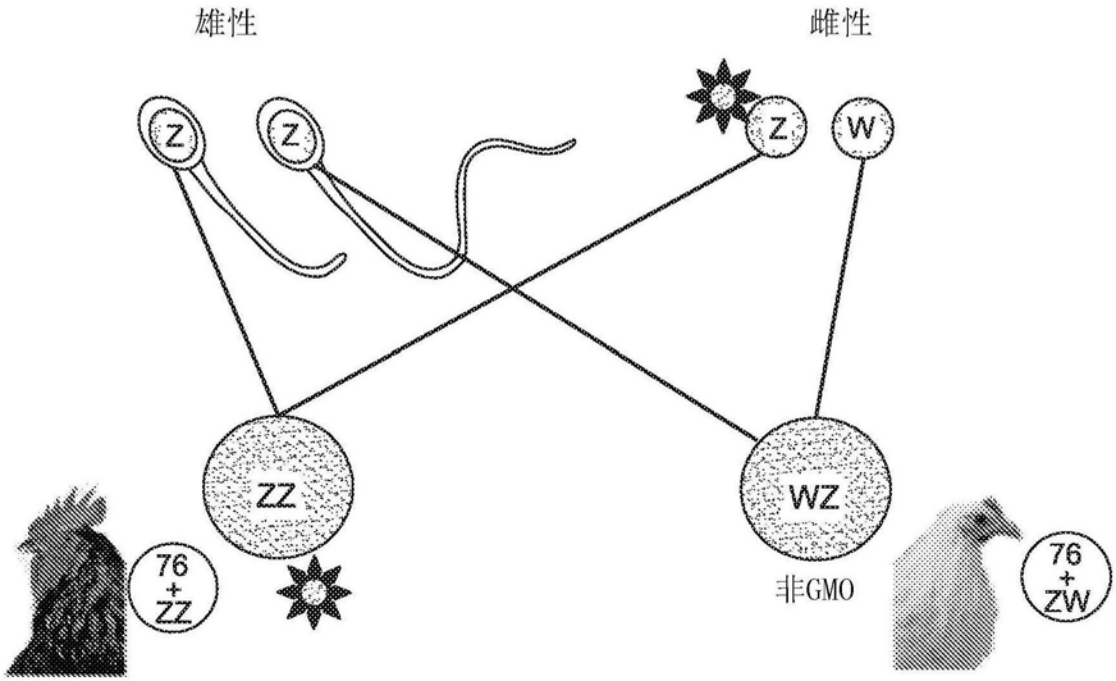


图5