Journal of Cell and Molecular Research (2021) 13 (1), 1-9 DOI: 10.22067/JCMR.2021.71604.1015

Research Article

A Multi-faceted Approach for Prediction of Genome Safe Harbor Loci in the Chicken Genome

Nima Dehdilani [1,2], Mohsen Fathi Najafi [3], Hesam Dehghani [1,2,4,*]

[1] *Division of Biotechnology, Faculty of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran*
[2] *Stem Cell Biology and Regenerative Medicine Research Group, Research Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran*
[3] *Education and Extension Organization (AREEO), Agricultural Research, Razi Vaccine and Serum Research Institute, Mashhad, Branch, Mashhad, Iran*
[4] *Department of Basic Sciences, Faculty of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran*

Received 23 July 20201 Accepted 23 August 2021

Abstract

To achieve a reliable and persistent expression, the transgene should be precisely integrated into the genome safe harbor (GSH) loci. Little attention has been paid to find the safe harbor loci of the chicken (*Gallus gallus domesticus*) genome. Identification and characterization of GSH loci that allow the persistent and reliable expression of knock-in genes could be a major area of interest within the field of transgenic technology and is central to the development of transgenic livestock. Randomly integrated transgenes might encounter position effects and epigenetic silencing, so unstable phenotypes, as well as unreliable and unpredictable expression of the knock-in transgene could occur. In contrast to random gene insertion, site-specific gene targeting provides a superior strategy that exploits homologous recombination to insert a transgene of interest into a pre-determined locus. In this study, based on bioinformatics, gene expression atlas, and Hi-C analyses, the GSH region was predicted in the chicken genome between *DRG1* and *EIF4ENIF1* genes. To do so, we introduce a fast and easy-to-use pipeline that allows the prediction of orthologue GSH loci in all organisms, especially chickens. In addition, the procedure to design Cas9/gRNA expression and targeting vectors for targeting these predicted GSH regions is described in detail.

***Keywords*:** Genome safe harbor loci, Genetically engineered birds, Transgenic chicken, CRISPR/Cas9, Gene expression atlas, Hi-C map.

Introduction

It has become increasingly important to determine regions that support the integration and long-term expression of a transgene in the genome. Considerable efforts have been underway to elucidate genomic safe harbor (GSH) loci that could potentially support long-term expressions within the field of transgenesis and recombinant protein production (Sadelain et al., 2011). The discovery of a GSH locus that allows reliable and consistent expression of a knock-in gene without triggering the functional disruption of internal genes is of utmost importance to develop bioreactors (Papapetrou et al., 2011; Ruan et al., 2015).

There is a growing number of strategies to screen and identify GSH loci. Traditional strategies are expensive, cumbersome, and time-consuming. For example, to identify and explore potential GSH loci, the "gene trapping" method has been used that relies on random integration of a promoterless reporter construct across the genome to indicate the expression of an endogenous gene (Stanford et al., 2001). Then, the integration sites are evaluated to find the regions with the highest expressions (Papapetrou et al., 2011). Such a reverse screening strategy is very laborious, as numerous sites which are subjected to reporter insertion should be analyzed. Another method to find the potential GSH loci is based on whole transcriptome sequencing, which could be expensive and needs specialized analyses (Ma et al., 2018). The *in vivo* imaging (Rizzi et al., 2017) of reporter animals which could be performed to find suitable loci for transgene integration, is also time-consuming and uneconomical. More recently, a systematic approach that combines the RNA-seq data with the High-throughput Chromosome Conformation Capture (Hi-C) data was proposed to predict the GSH regions (Hilliard and Lee, 2021). This approach is more informative and applicable, but the high cost is the main constraint against its universal use. However, it is helpful and cost-effective to do data-mining and similarity-finding

*Corresponding author's e-mail address: *dehghani@um.ac.ir*

experiments on the data adopted from validated GSH loci databases to identify potential orthologous GSH regions. Hence, it would be easy to predict actively transcribed regions, transcriptionally permissive topologically associating domains (TADs), and nucleosome-poor regions (Fishman et al., 2019; Hilliard and Lee, 2021; Zhao et al., 2019). A comparative genomics approach could also be applied to screen and detect similar sites/homologous sequences among different species (Irion et al., 2007; Li et al., 2014; Wu et al., 2016; Yang et al., 2016b).

Here, we introduce a novel, cost-effective, and easy-to-use pipeline to predict the potential GSH loci in the eukaryotic genome. We applied this pipeline to analyze the gene expression atlas, and Hi-C data to identify a potential GSH locus in the chicken genome. Then, we describe the procedure for designing CRISPR-based targeting and Cas9/gRNA expression vectors that are applicable for targeting these predicted GSH regions.

Materials and Methods

Pipeline for prediction of GSH loci

We used a five-step, fast, and easy-to-use pipeline to predict an orthologous GSH locus in the *Gallus gallus domesticus* genome (Figure 1). In the first step, genes around the validated GSH locus were found based on their sequence similarities in the chicken genome. The validated locus, the intergenic sequence between *DRG1* and *EIF4ENIF1* genes, was previously reported as a potential GSH locus in *Sus scrofa*, *Mus musculus*, and *Homo sapiens*. To this end, gene similarities around the predicted GSH locus were compared to the genes around the validated GSH locus using the NCBI Genome Data Viewer (GDV) browser (https://www.ncbi.nlm.nih.gov/genome/gdv/). In the second step, the intergenic sequence between *DRG1* and *EIF4ENIF1* genes from *Sus scrofa*, *Mus musculus*, and *Homo sapiens* was used as a template to perform a pairwise alignment (EMBOSS Water algorithm) against the similar intergenic sequence of the *Gallus gallus domesticus* (taxid: 9031) genome. In the third step, the predicted GSH locus was evaluated for the presence of possible annotated coding or non-coding genes using the NCBI GDV (*Gallus gallus* genome assembly GRCg6a) and the UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgGateway; chicken assembly GRCg6a/galGal6; Mar. 2018). In the fourth step, benefiting from chicken RNAseq data, the expression levels (transcript per millions; TPM) of the genes flanking the intergenic locus of interest were determined using the Gene Expression Atlas (https://www.ebi.ac.uk/gxa/home). In the fifth step, the chicken Hi-C data were used to predict the coordination of the GSH locus with its adjacent genes using the Hi-C map (Supplementary file 1). For visualizing the Hi-C map, the Juicebox software, version 1.9.0 (https://github.com/aidenlab/Juicebox/wiki/Download) was used to find the location of interest (coordinate system of the map corresponds to the genome version GalGal5). Then, the gene coordinates were compared with the locations of TAD boundaries. The map with the gene tracks is available in the following URL (http://sites.icgbio.ru/ontogen/wp-content/uploads/MolMechDevDepart/GCF_000002315.4_Gallus_gallus-5.0_genomic.gff.genes.bed). Alternatively, defined TAD boundaries were computationally adopted from the following URL (http://sites.icgbio.ru/ontogen/wp-content/uploads/MolMechDevDepart/subTADs-ChEF-all-HindIII-40k.hm.gzipped_matrix.jucebox_domains.annotation).

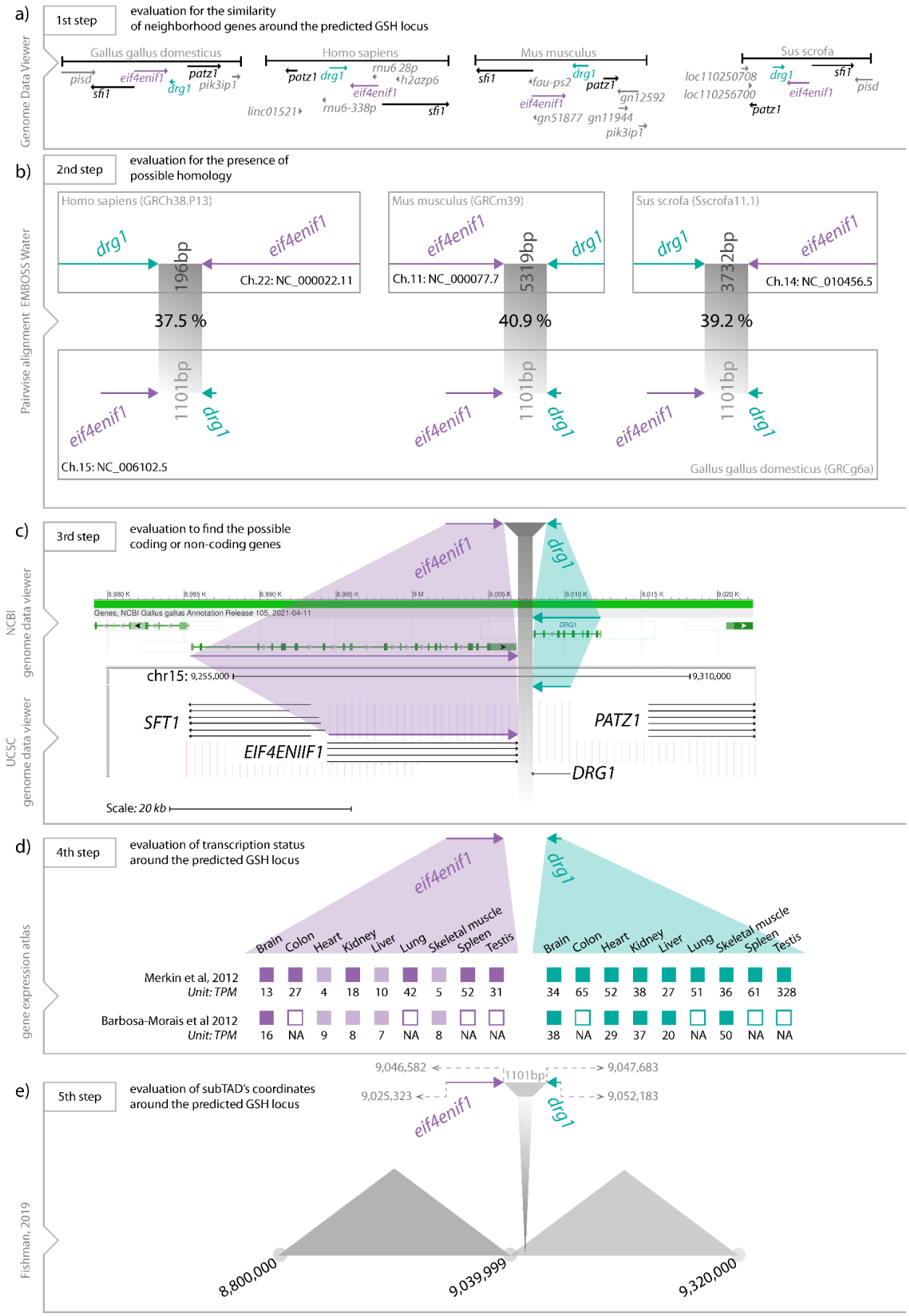
a)
1st step
evaluation for the similarity
of neighborhood genes around the predicted GSH locus
Genome Data Viewer
Gallus gallus domesticus
Homo sapiens
Mus musculus
Sus scrofa
b)
2nd step
evaluation for the presence of
possible homology
Pairwise alignment EMBOSS Water
Homo sapiens (GRCh38.P13)
Mus musculus (GRCm39)
Sus scrofa (Sscrofa11.1)
Ch.22: NC_000022.11
Ch.11: NC_000077.7
Ch.14: NC_010456.5
196bp
5319bp
3732bp
37.5 %
40.9 %
39.2 %
1101bp
Ch.15: NC_006102.5
Gallus gallus domesticus (GRCg6a)
c)
3rd step
evaluation to find the possible
coding or non-coding genes
NCBI genome data viewer
UCSC genome data viewer
Genes, NCBI Gallus gallus Annotation Release 105, 2021-04-11
chr15: 9,255,000
9,310,000
SFT1
EIF4ENIIF1
PATZ1
DRG1
Scale: 20 kb
d)
4th step
evaluation of transcription status
around the predicted GSH locus
gene expression atlas
Brain Colon Heart Kidney Liver Lung Skeletal muscle Spleen Testis
Merkin et al, 2012
Unit: TPM
13 27 4 18 10 42 5 52 31
34 65 52 38 27 51 36 61 328
Barbosa-Morais et al 2012
Unit: TPM
16 NA 9 8 7 NA 8 NA NA
38 NA 29 37 20 NA 50 NA NA
e)
5th step
evaluation of subTAD's coordinates
around the predicted GSH locus
Fishman, 2019
9,046,582
9,025,323
1101bp
9,047,683
9,052,183
8,800,000
9,039,999
9,320,000

Figure 1. A schematic depiction for deciphering a genome safe harbor locus in the chicken genome. a) Comparison of gene distribution around the validated GSH locus in human (*Homo sapiens*), mouse (*Mus musculus*), and pig (*Sus scrofa*) genome with the same region in the chicken genome using NCBI Genome Data Viewer; b) Pairwise alignment of the predicted chicken GSH locus and validated GSH loci in human, mouse, and pig genomes by EMBOSS Water algorithm for calculation of sequence identity; c) Screening for possible coding or non-coding genes in the predicted chicken GSH locus by NCBI and UCSC genome data viewers; d) Expression levels of *DRG1* and *EIF4ENIF1* in several tissues adopted from gene expression atlas; e) Coordinates of chicken subTADs around the predicted chicken GSH locus. Abbreviations: ch/chr, chromosome; bp, base pair; TPM, transcripts per million; NA, not applicable.

Designing Cas9/gRNA expression and CRISPR-based targeting vectors

To design a highly specific gRNA for the predicted chicken GSH locus, the intergenic sequence between *DRG1* and *EIF4ENIF1* genes was subjected to the CHOPCHOP search engine (https://chopchop.cbu.uib.no/). The predicted gRNA expressing sequence with high specificity to the predicted GSH locus was selected and synthesized (Macrogen, South Korea) as 20 bp forward (P1, Table 1) and reverse (P2) oligonucleotides with appropriate overhangs (cacc, for the forward oligonucleotide; and aaac, for the reverse oligonucleotide) to be cloned into the *Bbs*I site in the gRNA expression vector (pSpCas9(BB)-2A-Puro (PX459) V2.0; Plasmid #62988; Addgene, USA). The annealed gRNA oligonucleotide was phosphorylated by T4 polynucleotide kinase (PNK; Thermofischer, EK0031) for 30 min at 37 °C followed by a 30 min inactivation at 70 °C. To calculate the insert: vector molar ratio, the annealed oligonucleotides were run on a 2% agarose gel, and their integrated density index was determined using the ImageJ software (https://imagej.nih.gov/ij/download.html) . The gRNA expression vector was subjected to *Bbs*I enzymatic digestion for 1h, and the subsequent heat-inactivation for 30 min at 70 °C. Then, it was dephosphorylated with fast alkaline phosphatase (Thermofischer, EF0654) for 10 min at 37 °C, and heat-inactivated for 30 min at 70 °C. Using T4 DNA ligase (Thermofischer, EL0011), the phosphorylated gRNA was ligated to the digested and dephosphorylated gRNA expression vector for 3h at 16 °C which was followed by 12h at 4 °C.

To design a CRISPR/Cas9-based targeting vector, a plasmid containing CMV-PAC^r-IRES-EGFP cassette (Figure 2b; constructed in this laboratory) was used. This plasmid contains *Pvu*I-*Xho*I sites upstream of CMV, and *Nhe*I-*Xcm*I sites downstream of EGFP. Left and right homology arms (LHA and RHA, respectively) with approximately 500-bp length were amplified from the chicken genomic DNA by appropriate primers, including *Pvu*I cut site at the 5′-end of LHA (P3), *Xho*I cut site at the 3′-end of LHA (P4), *Nhe*I cut site at the 5′-end of RHA (P5), and *Xcm*I cut site at the 3′-end of RHA (P6). In two steps, arms were cloned into the targeting vector. At the first step, the EGFP plasmid was cut with *Nhe*I-*Xcm*I for 3h at 37 °C, followed by dephosphorylation with fast alkaline phosphatase for 10 min at 37 °C, and heat-inactivation for 30 min at 70 °C. The amplified RHA was cut with *Nhe*I and *Xcm*I for 3h at 37 °C. A 1:3 vector to insert molar ratio was used to ligate the amplified RHA into the EGFP vector by the T4 DNA ligase. The generated vector was called the pre-targeting vector. At the second step, the pre-targeting vector was cut with *Pvu*I and *Xho*I for 3h at 37 °C, followed by dephosphorylation with fast alkaline phosphatase for 10 min at 37 °C, and heat-inactivation for 30 min at 70 °C. The amplified LHA was cut with the *Pvu*I and *Xho*I for 3h at 37 °C. A 1:3 vector to insert molar ratio was used to ligate the amplified LHA into the pre-targeting vector by the T4 ligase. The generated vector was called the CRISPR/Cas9-based targeting vector.

For all cloning procedures, 5µl of the ligation mix was transformed into *E. coli* DH5α followed by overnight incubation at 37 °C. Colony PCR was performed on transformants by vector-specific, and insert-specific primers (Figure 2), and positive clones were grown for plasmid extraction using the Plasmid DNA Isolation Kit (DENAzist Asia, Iran). Cloning verification was performed using restriction enzyme digestion (Figure 2).

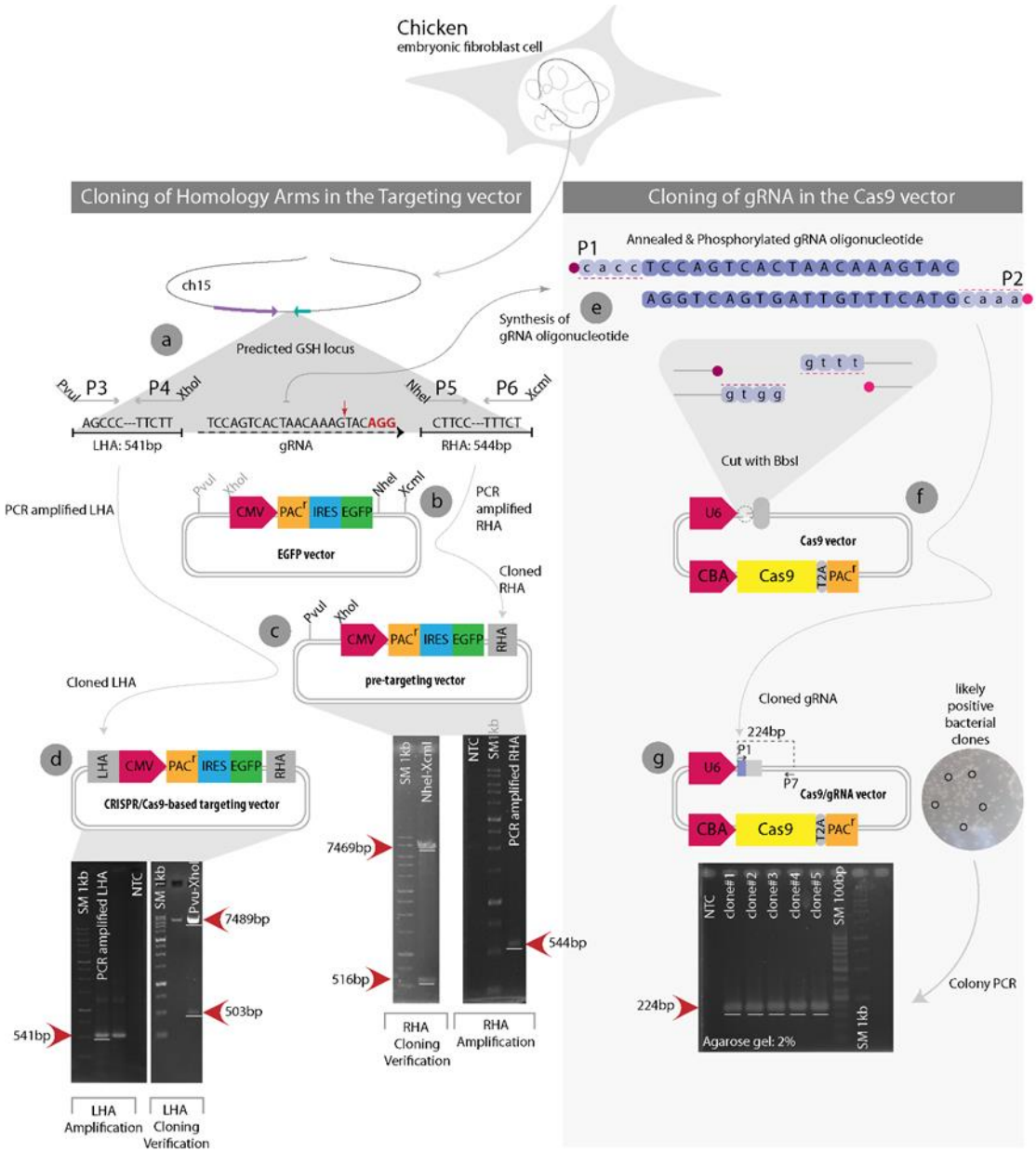

Figure 2. The proposed standard strategy for constructing the GSH targeting vector. a) A specific gRNA sequence in the predicted chicken GSH locus and the corresponding Cas9 cut site (red arrow) are depicted. P3 and P4 primers containing restriction enzyme sites for *Pvu*I and *Xho*I, as well as P5 and P6 primers containing restriction enzyme sites for *Nhe*I and *Xcm*I were designed and synthesized. b) A 544-bp PCR amplified RHA was cut with *Nhe*I/*Xcm*I. Then, the 516-bp RHA was cloned into the EGFP vector (7469-bp) to generate a pre-targeting vector (8012-bp). c) Cloning verification of RHA was performed by *Nhe*I/*Xcm*I digestion. The 516-bp RHA band and 7469-bp vector backbone were detected on the agarose gel. d) A 541-bp PCR amplified LHA was cut with *Pvu*I/*Xho*I. Then, the 503-bp LHA was cloned into the pre-targeting vector (7489-bp) to generate a CRISPR-based targeting vector (7992-bp). Verification of the LHA cloning was performed by *Pvu*I/*Xho*I. The 503-bp LHA band and 7489-bp vector backbone were detected on the agarose gel. e) gRNA oligonucleotides (P1 and P2) containing *Bbs*I overhangs were synthesized, annealed, and phosphorylated. f) gRNA oligonucleotides containing *Bbs*I overhangs were cloned into the *Bbs*I-digested Cas9 vector. g) Colony PCR was performed to verify the cloning of gRNA using P1 and P7 primers, and a 224-bp band was detected on the agarose gel. Abbreviations: LHA, left homology arm; RHA, right homology arm; SM, size marker; NTC, non-template control; bp, base pairs; gRNA, guide RNA; P, primer.

Results

Gene similarities around the validated GSH loci and predicted GSH locus

In mice (Tasic et al., 2011), humans (Zhu et al., 2014), and pigs (Ruan et al., 2015), the region between the *DRG1* and *EIF4ENIF1* genes has been

identified as a validated GSH locus to support the consistent expression of transgenes over time. The regions around the *DRG1* and *EIF4ENIF1* genes in the chicken genome are similar to the organisms mentioned above and include a gene-dense area. In the mouse, human, and pig genomes, the *sfi1* and *patz1* genes are located upstream of *EIF4ENIF1* and *DRG1* genes, respectively. Also, the direction of *EIF4ENIF1* and *DRG1* genes are towards each other in these organisms. These data suggest that the gene organization around the predicted GSH locus in the chicken genome is identical with the validated GSH loci (Figure 1a).

Finding possible similarities using the Water algorithm of pairwise alignment

To designate the orthologous GSH locus in the chicken genome, we first located the *DRG1* and *EIF4ENIF1* genes in the chicken genome. Local pairwise alignment (Water algorithm) describes the most similar region(s) within the sequences to be aligned. It was performed to find the possible similarity of known GSH intergenic locus in the human (GRCh38.P13), mouse (GRCm39), and pig (Sscrofa11.1) genomes with the same locus in chicken. To this end, the following regions, all from the intergenic locus between the *DRG1* and *EIF4ENIF1* genes, were selected and pairwise aligned to an 1100-bp region of the same locus in the chicken genome (NC_052546.1; Ch.15 from 9286985..9286084): a 195-bp region from the human genome (NC_000022.11; Ch.22 from 31434452 to 31434648), a 5318-bp region from the mouse genome (NC_000077.7; Ch.11 from 3194588 to 3199907), and a 3731-bp region from the pig genome (NC_010456.5; Ch.14 from 48153103 to 48156835). The results showed that the intergenic locus between the *DRG1* and *EIF4ENIF1* genes in the chicken genome has similarity scores of 37.5%, 40.9%, and 39.2% with the corresponding regions from the human, mouse, and pig genomes, respectively (Figure 1b).

Searching for possible coding and non-coding genes in the predicted GSH locus

The presence of any annotated coding and non-coding genes in the intergenic region between the chicken *DRG1* and *EIF4ENIF1* genes was evaluated using the UCSC Genome Browser (chicken assembly; Mar. 2018 GRCg6a/galGal6), and the NCBI Genome Data Viewer (GRCg6a). The results showed no coding/non-coding genes at a distance between these two genes (Figure 1c).

Evaluating the transcriptional status of genes adjacent to the predicted GSH locus

It has been demonstrated that the transgene can be expressed reliably in the actively transcribed regions of the genome. The transcriptional status of *DRG1* and *EIF4ENIF1* genes was evaluated using the RNAseq data adopted from the Gene Expression Atlas. RNAseq data showed that the *DRG1* gene is actively transcribed in several tissues. In contrast, the transcription of the *EIF4ENIF1* gene is variable among tissues. Hence, here the insertion site of the transgene of interest was designated near the *DRG1* gene (Figure 1d).

Designating the topological location of the predicted GSH locus using the TAD data

Several studies have confirmed that gene clusters located in a given TAD are regulated similarly (Hilliard and Lee, 2021). Transgenes inserted into a TAD containing actively transcribed genes maintain their transcriptional activity. Benefiting from the Hi-C data of chicken embryonic fibroblasts (Fishman et al., 2019), we located the topological position of the *DRG1* and *EIF4ENIF1* genes related to the adjacent TAD. It was surprising that the transgene insertion site was located in an individual active TAD near the *DRG1* gene (Figure 1e).

Constructing gRNA expression and CRISPR/Cas9-based targeting vectors

Synthetic gRNA oligonucleotides specific to the predicted GSH locus were cloned into the Cas9 expression vector and confirmed with P1 and P7 primers (Table 1). Five clones of transformants were checked with colony PCR. To construct the CRISPR/Cas9-based targeting vector, the isogenic 544-bp right homology arm was amplified using P5 and P6 primers (Table 1). Then, it was cloned into the pre-targeting vector. The pre-targeting vector was cut with *Nhe*I and *Xcm*I, and the 516-bp and 7469-bp bands were detected on the agarose gel. Then, the isogenic 541-bp left homology arm was amplified using P3 and P4 primers (Table 1), confirmed and cloned into the pre-targeting vector. For verification of LHA cloning, a CRISPR-based targeting vector was cut with *Pvu*I and *Xho*I, and 503-bp and 7489-bp bands were detected on the agarose gel.

Table 1. List of primers used in this study.

Primer	Sequence (5′ to 3′)	Length (bp)
P1	CACCTCCAGTCACTAACAAAGTAC	20-mer
P2	AAACGTACTTTGTTAGTGACTGGA	20-mer
P3	CATGCATTAGTTCGCGATCGAGCCCTAGGGGAGGTCCTG	39-mer
P4	TGGCGACCGGTACCCTCGAGAAGAATTTCCTGCTTATTTGACTTCTCC	48-mer
P5	CTTTCTAGGGTTAAGCTAGCCTTCCACTAGTATAAACAATTG	42-mer
P6	TGGTGCCACCTATGTTGTGGAGAAATAAAACTGCTCTCCC	40-mer
P7	CGGGCCATTTACCGTAAG	18-mer

Discussion

One of the most important applications for the identification of GSH loci is to use these regions for insertion of transgenes and generating transgenic animals with the ability to be used as bioproduction systems (Li et al., 2019; Ruan et al., 2015). It is possible to achieve a consistent and reliable expression of the transgene by stable chromosomal insertion of the exogenous DNA at a GSH locus (Shin et al., 2020). Since the generation of transgenic animals is expensive and time-consuming, the prediction of the potential GSH loci can be helpful, preventing a potential transgene-silencing over generations.

Traditionally, random integration of a given transgene followed by the screening of the transgene expression across the genome has been used to find the highly expressed regions (Zambrowicz et al., 1997). This method is accompanied by cumbersome and time-consuming steps, including screening of the integrated transgenes, analyzing their expression levels, and identification of reliable GSH regions. Moreover, in the random integration approach, lots of regions, including intergenic and intragenic regions, are targeted and screened, and some intragenic regions could be selected as potential GSH regions. In some previous studies, new GSH regions were identified by genome-wide comprehensive analyses (Ma et al., 2018) or using available bioinformatics data to search for potential GSH regions (Yang et al., 2016b; Irion et al., 2007; Kobayashi et al., 2012; Lee et al., 2019; Li et al., 2014; Liu et al., 2018; Rizzi et al., 2017; Ruan et al., 2015; Stanford et al., 2001; Tasic et al., 2011; Wu et al., 2016; Yang et al., 2016a; Zhu et al., 2014). It has been discussed that the utilization of intragenic regions to express a transgene may lead to transgene silencing, disruption of endogenous genes, or even inducing oncogenes (Oleg E. Tolmachov et al., 2013).

Whole transcriptome analysis has been performed to widely analyze gene expression levels in a range of organisms and tissues (Ma et al., 2018). Thus, an intergenic region between two highly expressed genes could be a potential GSH candidate (Tasic et al., 2011). Nowadays, with the advent of CRISPR/Cas technology, a transgene of interest can be precisely integrated into the candidate GSH region (Kimura et al., 2014). It has been demonstrated that the actively expressed gene-rich regions may support a reliable, consistent, and long-term expression of the transgene. For example, the intergenic region between the *DRG1* and *EIF4ENIF1* genes has been known as a GHS locus in several animals, including mice, humans, and pigs. It has been revealed that both *DRG1* and *EIF4ENIF1* genes have broad spatial and temporal EST (expression sequence tag) expression patterns (Hippenmeyer et al., 2010) and could reliably support transgene expression (Pryzhkova et al., 2020).

The DNA sequences between the *DRG1* and *EIF4ENIF1* genes in mice were compared with those from the same locus in humans to determine the level of sequence identity. Results showed that there was a 45% similarity (Zhu et al., 2014). Functional validation verified that the region was suitable as a safe location for the placement of transgenes in human cells. In another study, the prediction of the GSH locus was accomplished by the assessment of the similarity of adjacent genes and their intron/exon organization. For example, the intergenic sequence between the pig *DRG1* and *EIF4ENIF1* genes was predicted to be a GSH due to its similarity to the adjacent genes of mice (Ruan et al., 2015). Although the GSH locus between these two genes was successfully predicted in the human and pig genomes based on a similarity search, there are two caveats in these studies: ***i***) the levels of expression of these two genes were not evaluated, and ***ii***) the coordination of the *DRG1* and *EIF4ENIF1* genes related to the locations of TAD boundaries was not investigated.

It has been demonstrated that transcriptionally permissive chromatin structures can support the consistent and reliable expression of a transgene. For example, 10.9% of the Chinese hamster ovary (CHO) cell genome contains actively transcribed 3D

chromatin structures, leading to the stability of transgene expression during the cell development process (Hilliard and Lee, 2021). Therefore, the transcriptionally permissive 3D chromatin structures could be easily predicted using RNAseq and Hi-C data. Also, analyzing of NucMap data will make better predictions of GSH regions in the future (Zhao et al., 2019).

Here, we used a multi-faceted approach to predict a GSH region in the chicken genome based on similarity search, RNAseq data, and Hi-C data. The easy-to-use and fast pipeline for the prediction of GSH regions before generating the transgenic animals can facilitate industrial research and development procedures. In this study, for the first time, we introduce the GSH locus located between the *DRG1* and *EIF4ENIF1* genes in the chicken genome. It has been demonstrated that intergenic regions show less nucleosome occupancy than intragenic regions (Voong et al., 2016). So, it is less subjected to silencing. On the other hand, Tasic and colleagues showed that molecular integration tools could have better access to intergenic regions in comparison to intragenic regions in the mouse embryonic stem cells (Tasic et al., 2011). Also, consistent expression of the desirable transgene without silencing for over 30 passages has been reported (Zhu et al., 2014). Hence, the GSH locus located between the *DRG1* and *EIF4ENIF1* genes can be reliably applied to generate transgenic animals.

Acknowledgments
Hesam Dehghani would like to thank all members of his laboratory who have contributed to transgenesis and genome editing experiments.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., Strunov, A., et al. (2019) 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Research 47:648–665.

Hilliard, W. and Lee, K.H. (2021) Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis.
Biotechnology & Bioengineering 118:659–675.

Hippenmeyer, S., Youn, Y.H. Moon, H.M., Miyamichi, K. Zong, H., Wynshaw-Boris, A., et al. (2010). Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. Neuron 68: 695–709.

Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M. and Keller, G. (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nature Biotechnology 25: 1477–1482.

Kimura, Y., Hisano, Y., Kawahara, A. and Higashijima, S. (2014) Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Scientific Report 4:6545.

Kobayashi, T., Kato-Itoh, M., Yamaguchi, T., Tamura, C., Sanbo, M., Hirabayashi, M. et al. (2012). Identification of rat Rosa26 locus enables generation of knock-in rat lines ubiquitously expressing tdTomato. Stem Cells Development 21:2981–2986.

Lee, E.-S., Moon, S., Abu-Bonsrah, K.D., Kim, Y.K., Hwang, M.Y., Kim, Y.J. et al. (2019) Programmable nuclease-based integration into novel extragenic genomic safe harbor identified from Korean population-based CNV analysis. Molecular Therapy Oncolytics 14:253–265.

Li, X., Yang, Y., Bu, L., Guo, X., Tang, C., Song, J. et al. (2014) Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. Cell Research 24:501–504.

Li, Y.-S., Meng, R.-R., Chen, X., Shang, C.-L., Li, H.-B., Zhang, T.-J., et al. (2019) Generation of H11-albumin-rtTA Transgenic Mice: A tool for inducible gene expression in the liver. G3. 9:591–599.

Liu, T., Hu, Y., Guo, S., Tan, L., Zhan, Y., Yang, L., et al. (2018) Identification and characterization of MYH9 locus for high efficient gene knock-in and stable expression in mouse embryonic stem cells. PLoS One 13 (2):e0192641.

Ma, L., Wang, Y., Wang, H., Hu, Y., Chen, J., Tan, T., et al. (2018) Screen and verification for transgene integration sites in pigs. Scientific Report **8:**1–11.

Oleg E. Tolmachov, Subkhankulova, T. and Tolmachov, T. (2013) Silencing of transgene expression: a gene therapy perspective. Gene therapy tools and potential applications,

francisco martin molina, intechopen, doi: 10.5772/53379. available from: https://www.intechopen.com/chapters/43164

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M.S., et al. (2011) Genomic safe harbors permit high β -globin transgene expression in thalassemia induced pluripotent stem cells. Nature Biotechnology 29: 73–78.

Pryzhkova, M. V., Xu, M.J. and Jordan, P.W. (2020) Adaptation of the AID system for stem cell and transgenic mouse research. Stem Cell Research. 49.

Rizzi, N., Rebecchi, M., Levandis, G., Ciana, P. and Maggi, A. (2017) Identification of novel loci for the generation of reporter mice. Nucleic Acids Research 45(6): 37.

Ruan, J., Li, H., Xu, K., Wu, T., Wei, J. and Zhou, R. (2015) Highly efficient CRISPR / Cas9- mediated transgene knockin at the H11 locus in pigs. Scientific Reports 5:1–10.

Sadelain, M., Papapetrou, E.P. and Bushman, F.D. (2011) Safe harbours for the integration of new DNA in the human genome. Nature Reviews Cancer 12:51–58.

Shin, S., Kim, S.H., Shin, S.W., Grav, L.M., Pedersen, L.E., Lee, J.S., et al. (2020) Comprehensive analysis of genomic safe harbors as target sites for stable expression of the heterologous gene in HEK293 cells. ACS Synthetic Biology 9:1263–1269.

Stanford, W.L., Cohn, J.B., Cordes, S.P. and Lunenfeld, S. (2001) Gene-trap mutagenesis: past, present and beyond. Nature reviews genetics 2:756–768.

Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y., et al. (2011) Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proceedings of the National Academy of Sciences 108: 7902–7907.

Voong, L.N., Xi, L., Sebeson, A.C., Xiong, B., Wang, J.-P. and Wang, X. (2016) Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. Cell 167:1555-1570.

Wu, M., Wei, C., Lian, Z., Liu, R., Zhu, C., Wang, H., et al. (2016) Rosa26 -targeted sheep gene knock-in via CRISPR-Cas9 system. Scientific Reports. 6:1–7.

Yang, D., Song, J., Zhang, J., Xu, J., Zhu, T. and Wang, Z. (2016) Identification and characterization of rabbit ROSA26 for gene knock-in and stable reporter gene expression. Scientific Reports 6:1–8.

Yang, D., Song, J., Zhang, J., Xu, J., Zhu, T., Wang, Z., et al. (2016) Identification and characterization of rabbit ROSA26 for gene knock-in and stable reporter gene expression. Scientific Reports 6: 251-61.

Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G. and Soriano, P. (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proceedings of the National Academy of Sciences 94: 3789–3794.

Zhao, Y., Wang, J., Liang, F., Liu, Y., Wang, Q., Zhang, H., et al. (2019) NucMap: a database of genome-wide nucleosome positioning map across species. Nucleic Acids Research. 47: 163–169.

Zhu, F., Gamboa, M., Farruggio, A.P., Hippenmeyer, S., Tasic, B., Schüle, B., et al. (2014) DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. Nucleic Acids Research. 42(5):e34.