






## Article

# Identification of Genomic Safe Harbors in the Anhydrobiotic Cell Line, Pv11

Yugo Miyata <sup>1,2,†</sup>, Shoko Tokumoto <sup>1,†</sup>, Tomohiko Arai <sup>3,†</sup>, Nurislam Shaikhutdinov <sup>4,5</sup>, Ruslan Deviatiiarov <sup>5,6</sup> , Hiroto Fuse <sup>3</sup>, Natalia Gogoleva <sup>5</sup>, Sofya Garushyants <sup>4</sup>, Alexander Cherkasov <sup>4</sup>, Alina Ryabova <sup>5</sup> , Guzel Gazizova <sup>5</sup>, Richard Cornette <sup>1</sup> , Elena Shagimardanova <sup>5</sup>, Oleg Gusev <sup>5,6,7,8</sup> , and Takahiro Kikawada <sup>1,3,\*</sup> 

- <sup>1</sup> Division of Biomaterial Sciences, Institute of Agrobiological Sciences, National Agriculture and Food Research Organization (NARO), Tsukuba 305-0851, Japan; miyata.mche@tmd.ac.jp (Y.M.); tokumotos023@affrc.go.jp (S.T.); cornette@affrc.go.jp (R.C.)
  - <sup>2</sup> Department of Medical Chemistry, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan
  - <sup>3</sup> Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-8562, Japan; 9047097639@edu.k.u-tokyo.ac.jp (T.A.); 4182125849@edu.k.u-tokyo.ac.jp (H.F.)
  - <sup>4</sup> Center of Life Sciences, Skolkovo Institute of Science and Technology, 121205 Moscow, Russia; nurislam.shaikhutdinov@skoltech.ru (N.S.); garushyants@gmail.com (S.G.); cherkasovav@yandex.ru (A.C.)
  - <sup>5</sup> Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, 420012 Kazan, Russia; ruselusalbus@gmail.com (R.D.); negogoleva@gmail.com (N.G.); urban-nomad@yandex.ru (A.R.); grgazizova@gmail.com (G.G.); ryukula@gmail.com (E.S.); gaijin.ru@gmail.com (O.G.)
  - <sup>6</sup> Endocrinology Research Center, 115478 Moscow, Russia
  - <sup>7</sup> Graduate School of Medicine, Juntendo University, Tokyo 113-8421, Japan
  - <sup>8</sup> Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan
- \* Correspondence: kikawada@affrc.go.jp
- † These authors contributed equally to this work.



**Citation:** Miyata, Y.; Tokumoto, S.; Arai, T.; Shaikhutdinov, N.; Deviatiiarov, R.; Fuse, H.; Gogoleva, N.; Garushyants, S.; Cherkasov, A.; Ryabova, A.; et al. Identification of Genomic Safe Harbors in the Anhydrobiotic Cell Line, Pv11. *Genes* **2022**, *13*, 406. <https://doi.org/10.3390/genes13030406>

Academic Editor: Liangxue Lai

Received: 6 February 2022

Accepted: 21 February 2022

Published: 24 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Genomic safe harbors (GSHs) provide ideal integration sites for generating transgenic organisms and cells and can be of great benefit in advancing the basic and applied biology of a particular species. Here we report the identification of GSHs in a dry-preservable insect cell line, Pv11, which derives from the sleeping chironomid, *Polypedilum vanderplanki*, and similar to the larvae of its progenitor species exhibits extreme desiccation tolerance. To identify GSHs, we carried out genome analysis of transgenic cell lines established by random integration of exogenous genes and found four candidate loci. Targeted knock-in was performed into these sites and the phenotypes of the resulting transgenic cell lines were examined. Precise integration was achieved for three candidate GSHs, and in all three cases integration did not alter the anhydrobiotic ability or the proliferation rate of the cell lines. We therefore suggest these genomic loci represent GSHs in Pv11 cells. Indeed, we successfully constructed a knock-in system and introduced an expression unit into one of these GSHs. We therefore identified several GSHs in Pv11 cells and developed a new technique for producing transgenic Pv11 cells without affecting the phenotype.

**Keywords:** Pv11 cells; genomic safe harbor sites; anhydrobiosis; transgenesis; cell engineering

## 1. Introduction

Transgene integration, one of the most commonly used and effective techniques in biological research, can be achieved by either of two approaches: random integration of the gene of interest or site-specific knock-in. Random integration is the simpler method, but it can have unwanted and unpredictable effects on the host cell phenotype, depending on the integration site. In contrast, site-specific knock-in results in more controlled outcomes and has been facilitated by recent advances in genome-editing technologies. Furthermore,

in some species and cell lines, several research groups have developed genomic sites called genomic safe harbors (GSHs) that enable transgenes to function as designed without adverse effects on the host [1–6]. Transgene integration into such GSHs has promoted fundamental and applied biological research, but identifying GSHs is still challenging, especially in non-model species, owing to insufficient development of the genome database and genetic engineering tools.

GSHs are sites in the genome where new genetic material can be integrated without affecting the host phenotype and where the newly integrated material functions as designed [1]. For example, in basic research, a variety of genetic materials have been integrated into a mouse GSH, the Rosa26 locus, allowing exogenous gene expression or highly targeted gene knockdown experiments [7]. In applied research, GSHs have been used to establish cells that can stably produce proteins of medical interest, for example, therapeutic antibodies [4,8] and species-specific glycosylated proteins [9]. In addition, some researchers have proposed a therapy for genetic diseases using cells that are genetically modified via human GSHs [10]. Therefore, the identification and exploitation of GSHs can significantly advance our understanding of the biology of a particular species as well as leading to the development of new biotechnological applications.

GSHs have usually been identified by a four-step approach [4,11,12]: (i) creation of a transgenic cell pool by random integration; (ii) selection of cells showing stable transgene expression and no/few phenotype changes due to the exogenous gene material; (iii) genome analysis of the selected cells; (iv) confirmation of whether the integration sites are GSHs by a site-specific knock-in method. However, there are a number of difficulties in executing this strategy. Thus, for step (i), suitable genetic engineering tools are needed to create transfectants, and these might not be available for the species of interest. Step (ii) is time-consuming and requires a major effort to monitor the functional and phenotypic stability of the integrated genetic material and the transformant cells, respectively. For step (iii), a genome sequence and database are normally required to identify integration sites, while for step (iv), an appropriate site-specific knock-in method, such as the CRISPR/Cas9 system, must be available for the subject species. Therefore, GSH identification is not a straightforward process, especially in non-model species.

Pv11 is a culturable cell line derived from an insect, the sleeping chironomid *P. vanderplanki*, which inhabits semi-arid regions in Africa [13]. *P. vanderplanki* larvae display extreme desiccation tolerance [14], and Pv11 has inherited this ability, such that the cells can be preserved in the dry state at room temperature, while retaining their ability to proliferate once rehydrated [13]. When Pv11 cells are dried, any exogenous protein they contain can be preserved at room temperature for up to 372 days; thus, because of their desiccation tolerance, Pv11 cells potentially have a number of industrial applications, for instance, as water-free storage containers for biomaterials at room temperature [15]. The identification of GSHs in Pv11 cells will provide a major step towards such applications.

To explore the molecular mechanisms underlying the desiccation tolerance of Pv11 cells, we have exploited several gene-manipulation techniques, including a random integration method [16] and a site-specific knock-in method using the CRISPR/Cas9 system [17,18], and have generated several key resources. We have produced a transgenic cell pool, so-called KH cells, by random integration of an AcGFP1 (*Aequorea coerulea* GFP)-expressing plasmid [16]. Furthermore, we have a well-annotated genomic database and have developed the CRISPR/Cas9 system in Pv11 cells [19]. Thus, the materials needed to identify GSHs of Pv11 cells are already in place.

Here we describe the successful identification of GSHs in Pv11 cells. Following the approach used by others, we derived subpopulations from a KH cell pool and then analyzed the transgene integration sites in these cells at the whole-genome level. We found four candidate GSHs in Pv11 cells. We confirmed their potential using an AcGFP1 expression unit consisting of a functional promoter, the AcGFP1-coding sequence, and the polyadenylation signal, which were inserted into each site using the CRISPR-mediated targeted knock-in method. The proliferation rate and desiccation tolerance of the knock-in

cells were analyzed, leading to the identification of three GSHs on chromosome 1 (Chr1) of the Pv11 genome, Chr1:21143572, Chr1:21155382, and Chr1:21164645. Knock-in cells containing exogenous material at the Chr1:21164645 site displayed stable expression of the transgene after more than one year in culture. In addition, we constructed a basic donor vector system to knock-in exogenous genetic material into the Chr1:21164645 site. These results demonstrate the construction of a new gene-manipulation system for Pv11 cells, which should facilitate future advances in the basic biology and applied biotechnology of these cells.

## 2. Materials and Methods

### 2.1. Cell Culture

Pv11 and Pv11-KH cells were grown as described previously [18]. Briefly, for their culture, IPL-41 medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 2.6 g/L tryptose phosphate broth (Becton, Dickinson and Company, Franklin Lakes, NJ, USA), 10% (*v/v*) fetal bovine serum, and 0.05% (*v/v*) of an antibiotic and antimycotic mixture (penicillin, amphotericin B, and streptomycin; MilliporeSigma, Burlington, MA, USA) was used, and the medium is designated hereafter as complete IPL-41 medium. A density of  $3 \times 10^5$  cells/mL were seeded into a cell culture flask with plug seal cap and grown at 25 °C for 6–7 days.

### 2.2. Cloning a Subpopulation of KH Cells

Single cell sorting was performed using a MoFlo Astrios cell-sorter (Beckman Coulter Life Sciences, Indianapolis, IN, USA) as described previously [18]. Briefly, the cells were stained with DAPI (Dojindo, Kumamoto, Japan), and DAPI and AcGFP1 were excited with 355 and 488 nm lasers, respectively. In total, 1000 wild-type Pv11 cells were seeded as a feeder layer in each well of a 96-well plate prior to sorting. The sorted and the feeder cells were grown for two weeks, and then a second sorting (bulk cell sorting) was performed to eliminate the feeder cells.

### 2.3. Desiccation and Rehydration

Pv11 cells were subjected to desiccation—rehydration as described previously [13]. Briefly, cells were incubated in preconditioning medium (600 mM trehalose containing 10% (*v/v*) complete IPL-41 medium) for 48 h at 25 °C. Forty-microliter aliquots of the cell suspension were dropped into 35-mm petri dishes, and the dishes were desiccated and maintained at <10% relative humidity and 25 °C for more than seven days. An hour after rehydration by complete IPL-41 medium, cells were stained with propidium iodide (PI; Dojindo) and Hoechst 33342 (Dojindo), and images were acquired using a conventional fluorescence microscope (BZ-X700; Keyence, Osaka, Japan). The survival rate was calculated as the ratio of the number of live cells (Hoechst-positive and PI-negative) to that of total cells (Hoechst-positive).

### 2.4. Proliferation Analysis

Pv11 cells and the other cell lines were seeded at a density of  $1 \times 10^5$  cells/mL, and the live cell numbers were counted after staining with PI and Hoechst 33342 (Dojindo), as described in Section 2.3.

### 2.5. High-Molecular-Weight DNA Extraction and Purification

High-molecular-weight DNA was extracted with NucleoBond HMW DNA Kit (TaKaRa Bio, Shiga, Japan). The DNA solution was further treated with a Short Read Eliminator Kit (Circulomics, Baltimore, MD, USA) to deplete short DNA fragments. DNA concentrations were measured using a Qubit2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

### 2.6. Library Preparation for MinION Sequencing

Library preparation was performed using a Ligation Sequencing Kit (version SQK-LSK109; Oxford Nanopore, Oxford, UK), following the manufacturer's instructions. These libraries were sequenced on the MinION platform, using a Flow Cell R9.4.1 and MinKNOW software (20.06.5).

### 2.7. Base-Calling and Data Analysis

The raw data were acquired as fast5 files, and base-called with Guppy basecaller software (v4.2.2). Low quality and short reads (Phred score < 7 and length < 1000 bp, respectively) were removed using NanoFilt (version 2.7.1). To detect the integration sites of a transfected plasmid (Data S1) [16], we performed the following three steps: (i) to extract the reads containing a sequence/sequences of the exogenous DNA material, a BLASTn (2.10.1+) search was carried out, using the exogenous sequence as the query against the database of sequencing reads that passed the quality filters (E-value <  $1 \times 10^{-30}$ , -outfmt 6); (ii) 500 bases of sequence upstream and downstream of the integrated exogenous sequences found in step (i) reads were extracted using SeqKit (version 0.13.2) with the following parameters: `grep -nrp, subseq -r`; (iii) the integration sites in the genome were found by BLASTn search of these upstream and downstream regions (see step (ii)) against the Pv11 genome (Genbank Assembly Accession ID: GCA\_018290105.1, E-value <  $1 \times 10^{-30}$ , -outfmt 6, max\_target\_seq 1, max\_hsps 1). The results were visualized using the DensityMap tool [20].

### 2.8. Expression Vectors

For Cas9 expression, the previously constructed vector, pPv121-hSpCas9, was used; gRNA expression vectors were also constructed as reported previously [17]. Briefly, pPvU6b-DmtRNA-BbsI was digested with BbsI, and annealed oligonucleotides were ligated into the cut vector. Vector names and oligonucleotides used for the gRNA expression vector construct are listed in Table S1.

### 2.9. Genomic PCR and Sanger Sequencing Analysis

Accurate genome sequences around GSH candidates were analyzed by genomic PCR and Sanger sequencing. Pv11 cell genomic DNA was extracted with a NucleoSpin Tissue kit (Takara Bio) and subjected to PCR using specific primer sets (Table S2). After gel purification of the PCR products, sub-cloning was carried out using a TOPO cloning kit (Thermo Fisher Scientific, Waltham, MA, USA), and the plasmids were sequenced.

### 2.10. Donor Vector Construction

Donor vectors containing expression units for AcGFP1 or zeocin resistance (ZeoR) were constructed using PCR, a HiFi Assembly kit (New England BioLabs, Ipswich, MA, USA) and a Zero Blunt TOPO PCR cloning kit (Thermo Fisher Scientific, Waltham, MA, USA). As a PCR template, the AcGFP1 or ZeoR expression unit of a previously constructed vector, pPv121-AcGFP1-Pv121-ZeoR, was used [21]. Primers in the PCR included the gRNA-target and 40 bp homology sequences of each knock-in site. PCR products were cloned into pCR4 Blunt-TOPO (Thermo Fisher Scientific, Waltham, MA, USA); all primer sequences and vector names are listed in Table S3.

Targeted integration at the Chr1:21164645#9 site was performed using donor vectors containing the bidirectional AcGFP1 and HaloTag expression unit and different homology arms (all PCR primers are listed in Table S4). Donor vectors were constructed using PCR, a HiFi Assembly kit (New England BioLabs) and a DNA Ligation Kit Mighty Mix (Takara Bio), based on pCR4-Pv.00443#1-P2A-GCaMP3 and pCRII-Pv.00443#1-P2A-AcGFP1-P2A-ZeoR [18]. The former vector was digested with SpeI and NotI. The 3910 bp backbone vector and the following two PCR fragments were assembled: (i) the 1000-base left homology arm with the gRNA-target sequence and SpeI site at the 5'- and 3'-end, respectively; (ii) the 1000-base right homology arm with the NotI site and the gRNA-

target sequence at the 5'- and 3'-end, respectively. In the HiFi assembly, the original SpeI and NotI sites of pCR4-Pv.00443#1-P2A-GCaMP3 were eliminated. The assembled vector was named pCR4-21164645#9\_1kbpHA-SpeI\_NotI, and further used as a PCR template to make PCR fragments with the 40–750 base homology arms. The PCR products were inserted into the digested 3910 bp backbone vector (pCR4-21164645#9\_40bpHA-SpeI\_NotI, pCR4-21164645#9\_125bpHA-SpeI\_NotI, pCR4-21164645#9\_250bpHA-SpeI\_NotI, pCR4-21164645#9\_500bpHA-SpeI\_NotI, pCR4-21164645#9\_750bpHA-SpeI\_NotI). In the case of the vector without a homology arm, AcGFP1 was used as a PCR template, which provided a sufficient length of fragment for HiFi assembly in the next step. A SpeI site and the gRNA-target sequence were added at the 5'-end of the fragment, while a NotI site and the gRNA-target sequence were added at the 3'-end. The PCR product was inserted into the digested 3910 bp backbone vector (pCR4-21164645#9\_0bpHA-SpeI\_AcGFP1\_NotI).

Next, pCRII-Pv.00443#1-P2A-AcGFP1-P2A-ZeoR was digested with HindIII and XbaI. The 3407 bp backbone vector and the following two PCR fragments were assembled: (i) the 121 promoter-controlled HaloTag expression unit with a SpeI site at the 5'-end, (ii) the 121 promoter-controlled AcGFP1 expression unit with a NotI site at the 3'-end. The assembled vector was named pCRII-SpeI-HaloTag-121-121-AcGFP1-NotI.

Lastly, pCR4-21164645#9\_0bpHA-SpeI\_AcGFP1\_NotI, pCR4-21164645#9\_40bpHA-SpeI\_NotI, pCR4-21164645#9\_125bpHA-SpeI\_NotI, pCR4-21164645#9\_250bpHA-SpeI\_NotI, pCR4-21164645#9\_500bpHA-SpeI\_NotI, pCR4-21164645#9\_750bpHA-SpeI\_NotI, and pCRII-SpeI-HaloTag-121-121-AcGFP1-NotI were digested with SpeI and NotI and ligated. The vectors were named pCR4-21164645#9\_0bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_40bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_125bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_250bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_500bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_750bpHA-HaloTag-121-121-AcGFP1, pCR4-21164645#9\_1kbpHA-HaloTag-121-121-AcGFP1 (the complete sequences are shown in Data S2–S8).

### 2.11. Transfection and Site-Specific Transgene Knock-In and Establishment of Clonal Cell Lines

Transfection for site-specific knock-in was carried out using a NEPA21 Super Electroporator (Nepa Gene, Chiba, Japan) as described previously [21]. In total, 5 µg each of the gRNA- and SpCas9-expression vectors plus 0.03 pmol each of donor vectors were transfected into Pv11 cells. Five days after transfection, the cells at a density of  $1 \times 10^5$  cells per mL were treated with 400 µg/mL zeocin. After the zeocin selection, the medium was changed to complete IPL-41 medium, and the cells were incubated for an additional two weeks. To establish clonal cell lines, single cell sorting was performed, using a MoFlo Astrios cell-sorter (Beckman Coulter Life Sciences), as described in Section 2.2. The sorted and the feeder cells were grown in the complete IPL-41 medium without zeocin for two weeks and then treated with zeocin for two weeks to eliminate the feeder cells. Once the establishment of clonal cell lines was confirmed, they were cultured in the complete IPL-41 medium without zeocin. A portion of each cell line was cryopreserved with CELLBANKER 1plus (Takara Bio). The remaining cells were passaged over one year in the complete IPL-41 medium without zeocin. Hence, cell lines cultured continuously for more than one year, the above-mentioned cryopreserved cells thawed and cultured for only three weeks, and intact Pv11 cells were used to analyze the expression level of AcGFP1 protein by the cell-sorter and cell phenotypes of survival rate and proliferation ability.

### 2.12. Optimization of the Homology Arm Length for Maximum Knock-In Efficiency

For knock-in of both HaloTag and AcGFP1 expression units at Chr1:21164645, 0.03 pmol of a donor vector containing the expression units were transfected with 5 µg each of the gRNA- and SpCas9-expression vectors plus 0.03 pmol of a donor vector containing the ZeoR expression unit with 40 bp homology arms. Five days after transfection, the cells at a density of  $1 \times 10^5$  cells per mL were treated with 400 µg/mL zeocin. After 10 days in culture with zeocin selection, the cells were subjected to flow cytometry analysis by a CytoFLEX S (Beckman Coulter Life Sciences). For HaloTag labeling, HaloTag TMRDirect



Ligand (Promega, Fitchburg, WI, USA) was added to the medium 16 h before the analysis. The fluorescence of DAPI, AcGFP1, and HaloTag TMRDirect Ligand was detected by excitation with a 405, 488, and 561 nm laser, respectively.

### 2.13. Statistical Analysis

All data were expressed as mean  $\pm$  standard deviation (SD). Differences between two groups were examined for statistical significance using Student's *t*-test. Statistical significance among more than three groups was examined by ANOVA followed by a Tukey post-hoc test. A *p*-value  $< 0.05$  denoted a statistically significant difference. GraphPad Prism 8 software (GraphPad, San Diego, CA, USA) was used for the statistical analyses.

## 3. Results

### 3.1. Cloning Subpopulations with Improved Anhydrobiotic Ability from a KH Cell Pool

To isolate clonal KH cell subpopulations, single-cell sorting was performed (Figure 1A). We acquired two cell lines, B2 and 4C, whose survival rates after rehydration were higher than the original KH cells (Figure 1B). The proliferation rates of B2 and 4C were the same and faster than that of KH cells, respectively, although all KH-derived cells grew more slowly than wild-type Pv11 cells (Figure 1C). Thus, the two cell lines displayed a less-impaired phenotype than the original KH cells.

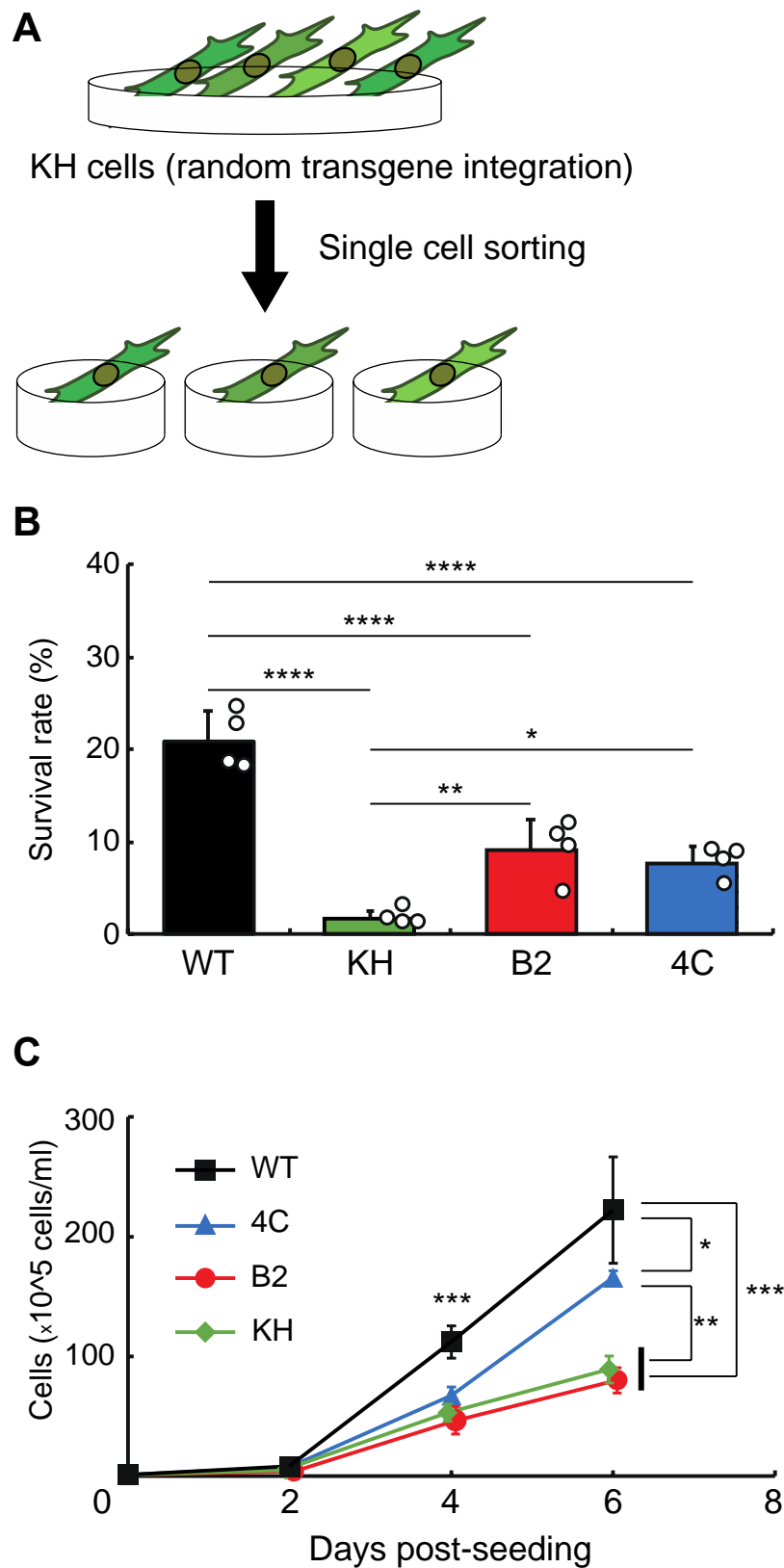
### 3.2. Genome-Wide Analysis and Identification of the Integration Sites in B2 and 4C Lines

Next, to identify the integration sites in the two cell lines, high-molecular-weight genomic DNA was extracted (Figure S1). DNA libraries were prepared and sequenced with a MinION sequencer (Table S5). As illustrated in Figure 2A, fragmented plasmid sequences were detected throughout the whole genome (Figure 2B, and Table S6). In contrast, the AcGFP1 expression unit was detected only on chromosome 1 (Chr1; Figure 2B,C, and Data S9 and S10) in both clones. Three of these integration sites, Chr1:280397, Chr1:21155382 and Chr1:21164645, were located in intergenic regions, while a fourth, Chr1:21143572, was located in the intron of transcription unit, *g12121*, whose expression is relatively low in Pv11 cells (Figure 2C, and Table S7, accession number GSE171333 [17]).

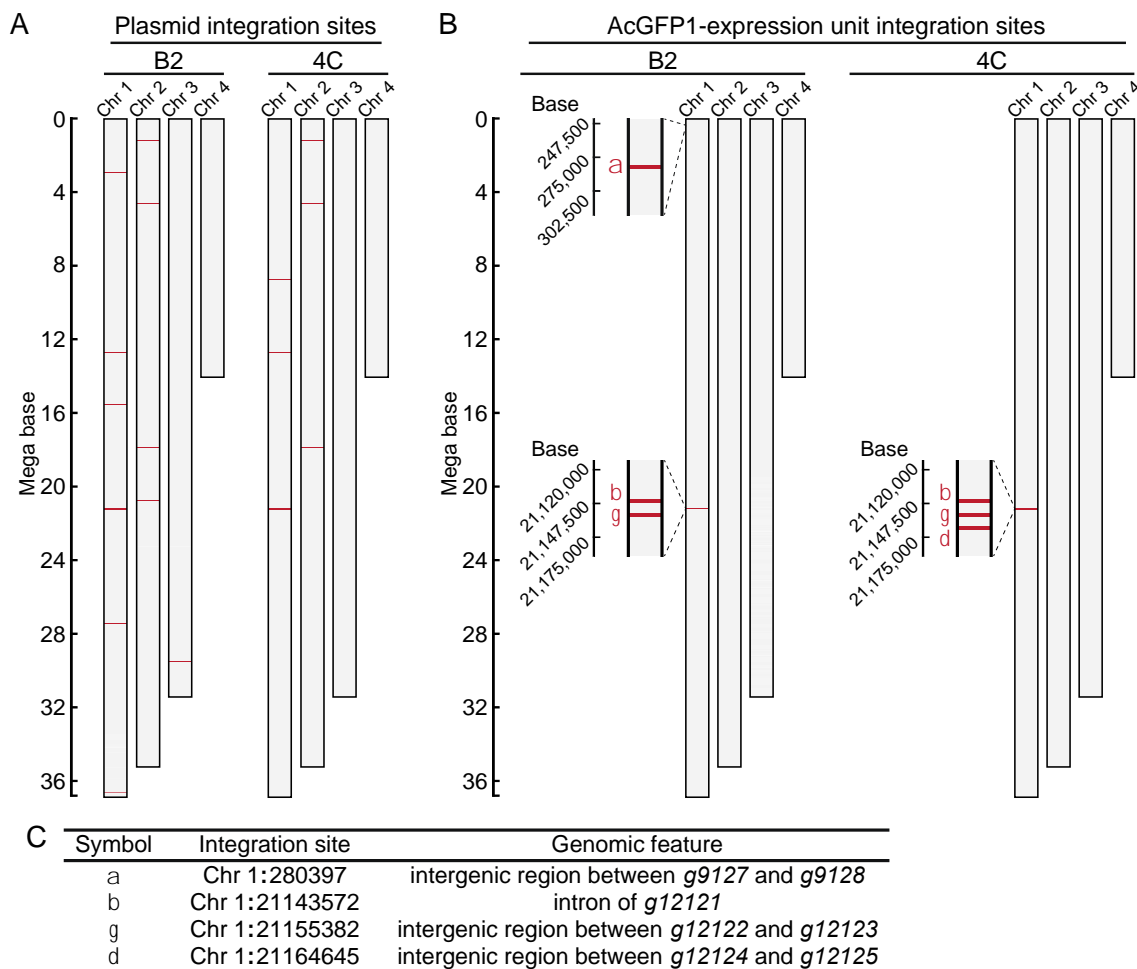
### 3.3. Identification of Genomic Safe Harbors in Pv11

Next, we examined whether genomic integration at the above sites affects the anhydrobiotic ability or proliferation rate of the corresponding cells. As shown in Figure 3A, AcGFP1 and ZeoR expression units were inserted individually into each genomic site in wild-type Pv11 cells by the CRIS-PITCh method [17,18,22]. Although precise integration was not achieved at Chr1:280397 (Figure S2), the other three sites allowed exogenous DNA integration as designed, and this was confirmed by Sanger sequencing (Data S11–S16). The knock-in cell lines displayed similar desiccation survival rates and proliferation rates to wild-type Pv11 cells (Figure 3B,C). Therefore, the three sites, Chr1:21143572, Chr1:21155382, and Chr1:21164645, were identified as potential GSHs in Pv11.

We then checked the stability of the protein expression level and the cellular phenotypes in a knock-in cell line with a copy of AcGFP1 at Chr1:21164645. The cells were grown for more than one year and, as shown in Figure 4, long-term culture had no effect on AcGFP1 fluorescence intensity (Figure 4A), nor the anhydrobiotic ability (Figure 4B) and proliferation rate (Figure 4C) of the cells.



**Figure 1.** Selection of clonal cell lines from the KH cell pool. (A) The experimental scheme is shown. To establish clonal cell lines from KH cells, single-cell sorting was performed. (B) The survival rates after desiccation–rehydration treatment of the clonal cell lines, B2 and 4C, are shown. (C) The proliferation rates of the B2 and 4C cell lines are shown. Values are expressed as mean  $\pm$  standard deviation (SD);  $n = 4$  in each group. \*\*\*\*  $p < 0.0001$ ; \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ .

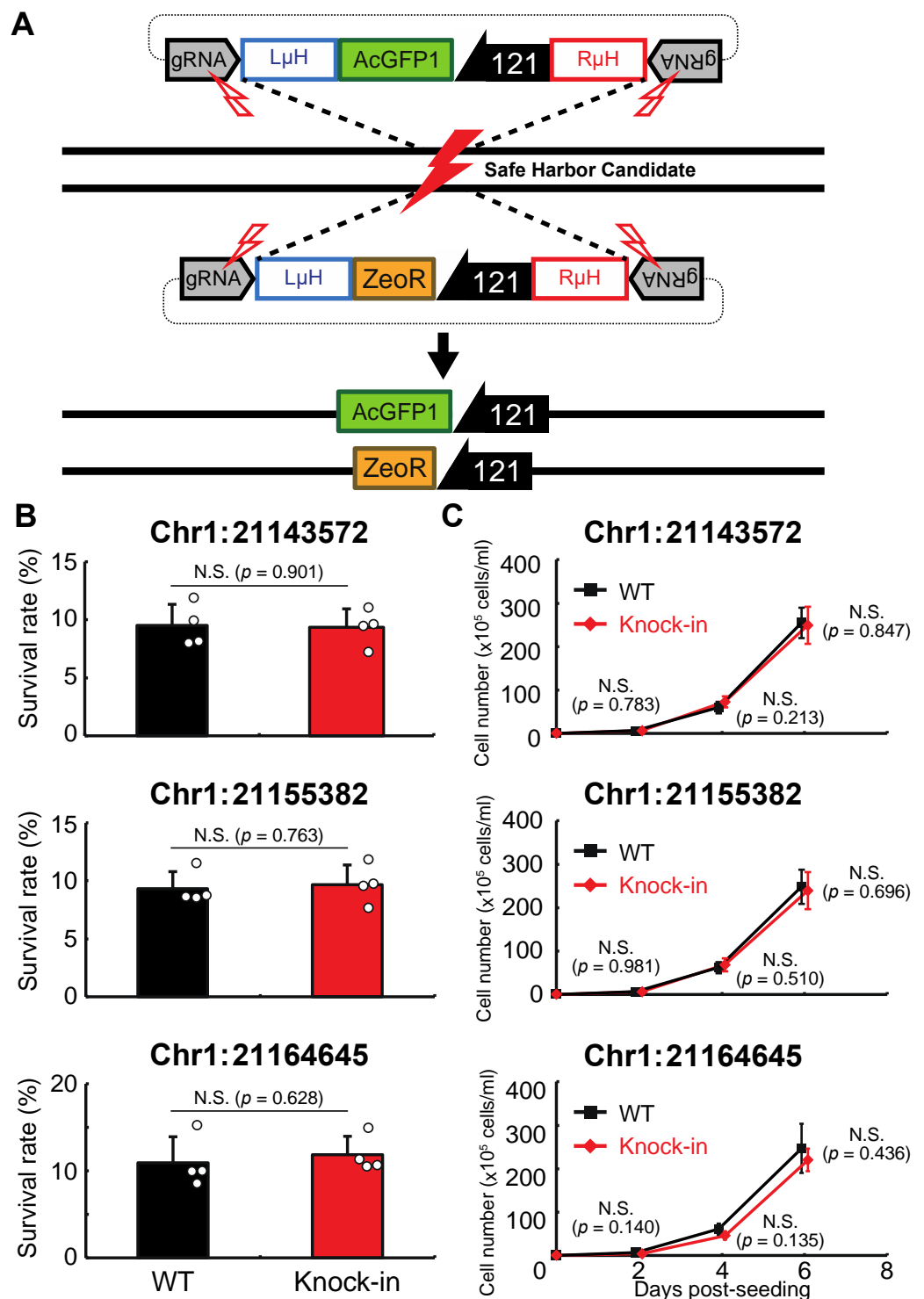


**Figure 2.** Genome-wide analysis of integration sites of the exogenous plasmid sequence in clonal cell lines, B2 and 4C. (A) Integration sites of the plasmid sequence are shown. (B) Integration sites of the AcGFP1-expression unit are shown. (C) Identified GSH candidates are listed, and their genomic features are described.

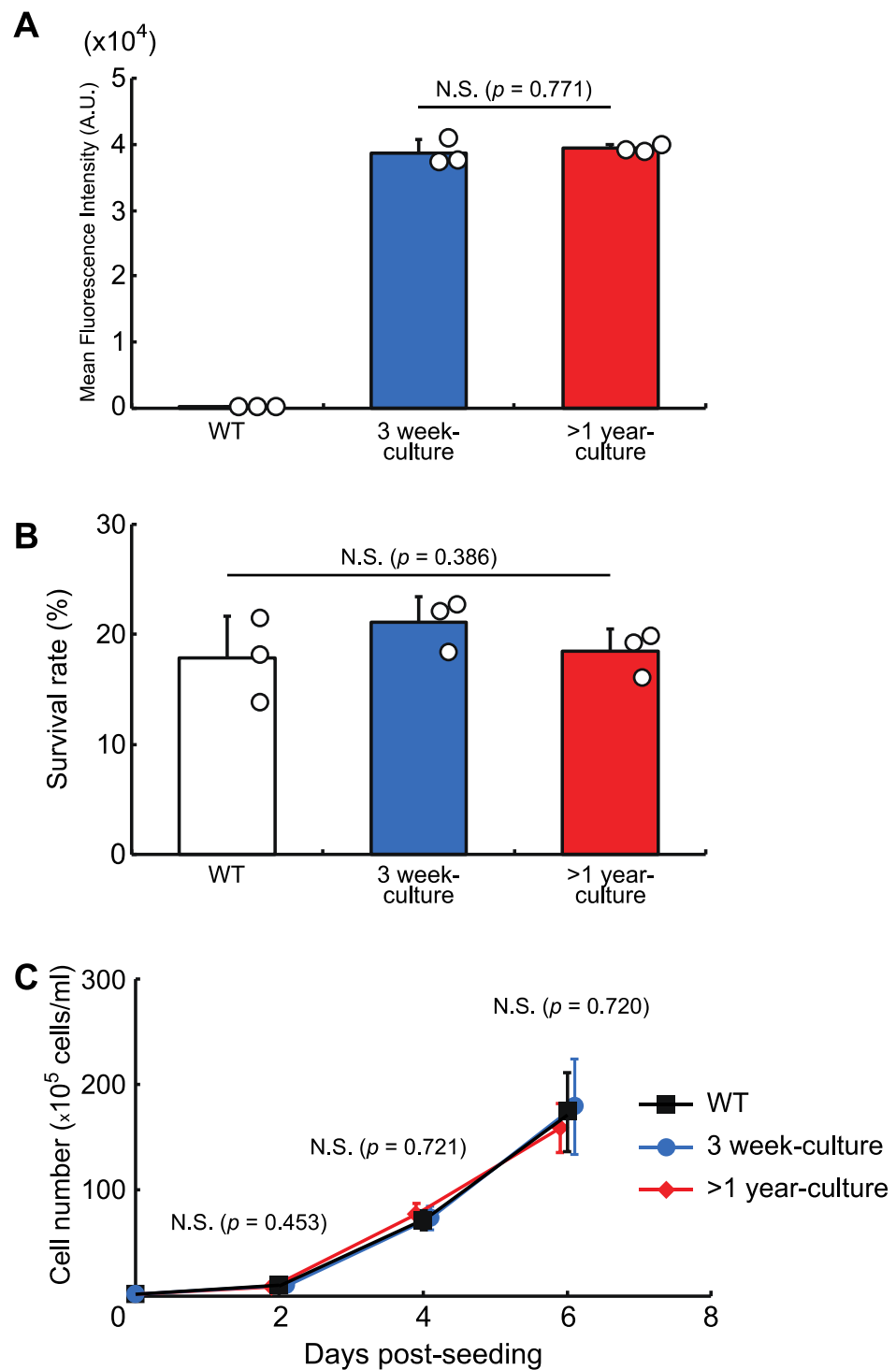
### 3.4. Construction of a GOI Knock-In System for *Pv11*

Next, we attempted to construct a gene-of-interest (GOI) expression system at the Chr1:21164645 site. A donor vector containing bidirectional HaloTag and AcGFP1 expression unit was designed, the former as an example of a GOI, and the latter as a marker of successful integration (Figure 5A). In our first attempt, a 40 bp homology arm length was used as shown in Figure 3 and described in previous studies [17,18,22]. However, the integration efficiency was low ( $7.1 \pm 1.3\%$  of the target cells were HaloTag<sup>+</sup>/AcGFP1<sup>+</sup> cells; Figure 5B,C), possibly because the insert size was much longer than the homology arm length. Therefore, to determine the optimal homology arm length for this knock-in system, we constructed a series of donor vectors with bidirectional HaloTag and AcGFP1 expression unit flanked by homology arms of various lengths (from 0 to 1000 bp, Data S2–S8; Figure 5A). Each of these donor vectors was transfected with the previously used construct containing the ZeoR expression unit (Figure 3), and after zeocin selection the protein expression levels of HaloTag and AcGFP1 were analyzed (Figure 5B). There was no significant difference in the knock-in efficiencies of the 250, 500, 750, and 1000 bp HA groups ( $23.7 \pm 2.7\%$ ,  $26.0 \pm 2.6\%$ ,  $26.0 \pm 4.3\%$ , and  $25.6 \pm 2.9\%$  of cells were HaloTag<sup>+</sup>/AcGFP1<sup>+</sup>, respectively), but the 0, 40, and 125 bp HA groups all gave a lower knock-in efficiency ( $8.8 \pm 2.4\%$ ,  $7.1 \pm 1.3\%$ , and  $20.7 \pm 1.7\%$  of cells were HaloTag<sup>+</sup>/AcGFP1<sup>+</sup>, respectively; Figure 5C). Thus, a homology arm length greater than 250 bp is needed for efficient knock-in of the GOI expression construct at the Chr1:21164645 site.

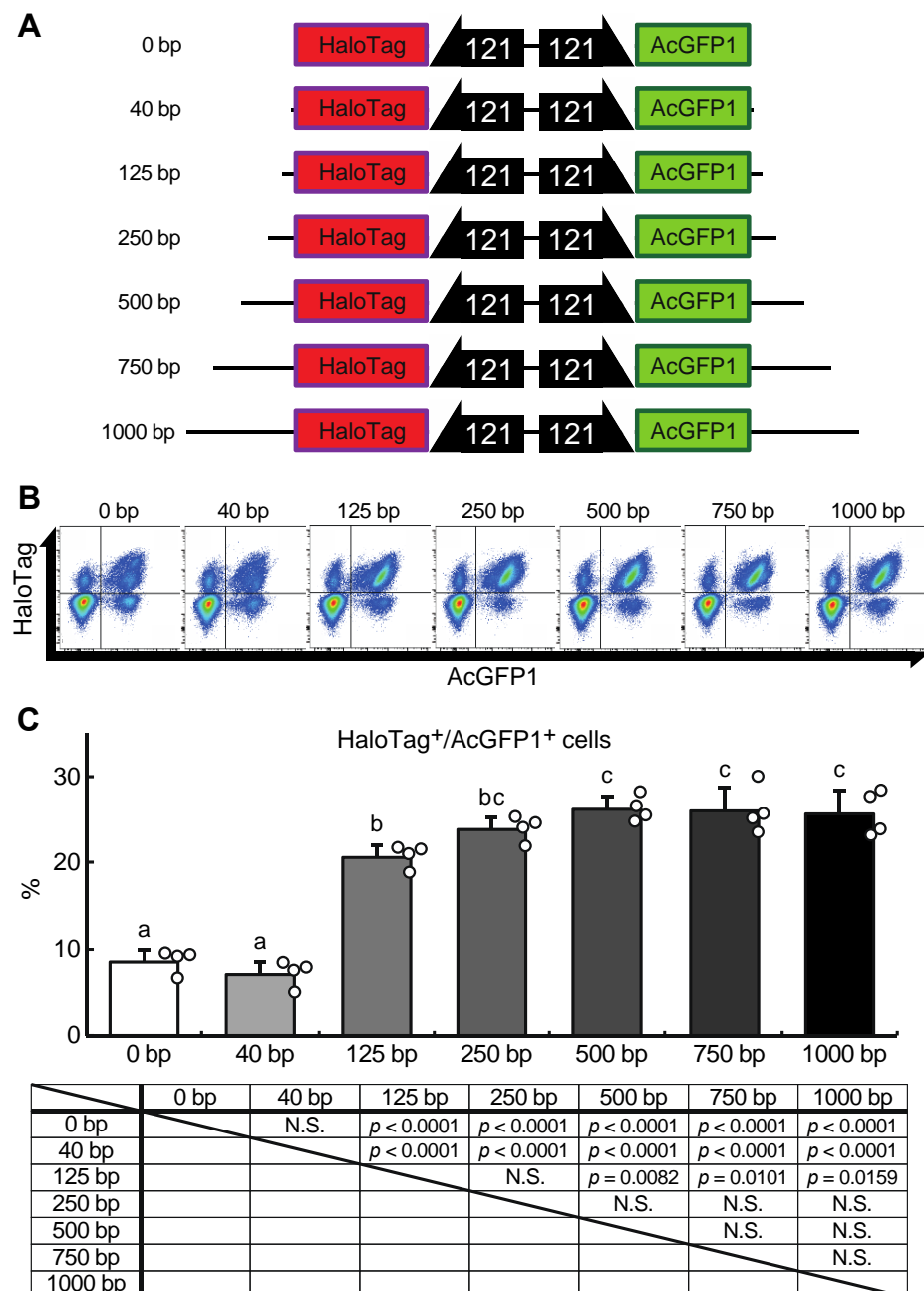




**Figure 3.** Integration of the expression units of AcGFP1 and zeocin-resistance (ZeoR) genes into GSH candidates. (A) The knock-in scheme for AcGFP1 and ZeoR expression units is shown; the donor vectors harboring AcGFP1 and ZeoR genes under control of the 121 promoter were transfected into Pv11 cells. (B) The survival rate after desiccation–rehydration treatment is shown for each knock-in cell line. (C) The proliferation rate of the knock-in cell lines is shown. Values are expressed as mean  $\pm$  SD;  $n = 4$  in each group. N.S., not significant.



**Figure 4.** The effects of long-term culture after knock-in at the GSH, Chr1:21164645. (A) AcGFP1-expression stability after more than one year in culture without zeocin. (B) Cell survival rate following desiccation–rehydration treatment after more than one year in culture without zeocin. (C) Cell proliferation rate after more than one year in culture without zeocin. Values are expressed as mean  $\pm$  SD;  $n = 3$  in each group in (A,B);  $n = 4$  in each group in (C). N.S., not significant.



**Figure 5.** A donor vector construct for GOI expression and optimization of the homology arm length for maximum knock-in efficiency at the Chr1:21164645 site. **(A)** Schematic outline of donor vectors with different homology arm lengths in the range 0–1000 bp. **(B)** Representative dot plot data of transfected cells after zeocin selection showing AcGFP1 and HaloTag fluorescence. **(C)** The proportions of AcGFP1<sup>+</sup> and HaloTag<sup>+</sup> cells in the live-cell population analyzed using a flow cytometer after 10 days in culture with zeocin selection, and the result of the statistical analysis. Values are expressed as mean  $\pm$  SD;  $n = 4$  in each group. N.S., not significant. Different letters above each bar indicate significant differences among groups at  $p = 0.05$  as shown in the statistical analysis. The darker shade indicates the longer homology arm length.

To check whether the GOI knock-in construct affects anhydrobiosis and cell proliferation, a clonal HaloTag- and AcGFP1-positive cell line was established, and the genotype and phenotype were analyzed. Sanger sequencing of the genome showed precise integration of the donor vector as designed in Figure 5A (Data S17–S19), while the cells exhibited the same anhydrobiotic ability and proliferation rate as wild-type Pv11 cells (Figure S3).

#### 4. Discussion

The major achievements of the current study are: (i) identification of GSHs in Pv11 cells and (ii) the construction of a basic knock-in system at one of the GSHs, Chr1:21164645. To identify GSHs in Pv11 cells, a transgenic cell pool obtained by random integration was used as source material, and two clonal cell lines with less-impaired phenotypic differences than the overall cell population were established. Whole-genome sequencing was performed on the cell lines, identifying four potential GSHs that were then used for knock-in experiments. At three of these sites, precise Cas9-mediated integration was achieved, and all three sites fulfilled the GSH criteria, i.e., there were no deleterious effects on anhydrobiotic ability or proliferation rate following knock-in. Furthermore, one of the GSHs, Chr1:21164645, maintained the function of the integrated gene after more than one year in culture and this site was subsequently used to construct a GOI-expression system. The resources generated in this study should be useful for advancing our understanding of anhydrobiosis, but also for biotechnological applications of Pv11 cells.

We previously reported a method to stably express GOIs in Pv11 cells [17,18]. The method exploits the high-expression system of the endogenous *Pv.00443* gene (as known as *g7775*): GOI coding sequences plus P2A are knocked into the 5'-flanking site of the stop codon of *Pv.00443* resulting in polycistronic expression. Although this approach is appropriate for constitutive expression of GOIs, it cannot be applied to integration of artificial expression units, for example, inducible gene expression systems [23], such as the Tet-On inducible expression system we previously developed in Pv11 cells [24]. Knocking in the system will be invaluable for tight control of the expression of proteins that inhibit the growth rate in Pv11 cells, and thus will facilitate the use of Pv11 cells as water-free storage containers for proteins. Another example is small RNA-expression systems including RNA Polymerase III promoters [25]. Polymerase III promoters have often been used to express shRNAs and gRNAs and can also be used for genome-wide screening [26,27]. Thus, our finding of GSHs in Pv11 cells will be of great benefit in further revealing the molecular mechanisms underlying anhydrobiosis.

We constructed a basic vector to knock-in a GOI at the Chr1:21164645 site (Figure 4) and found that homology arm lengths of 250–500 bp were sufficient for maximum knock-in efficiency. However, this efficiency was still only around 25%. To increase the efficiency, further modification of the knock-in system might be attempted, for example, by inhibition of some protein activities associated with the specific repair of DNA double-strand breaks [28]. Indeed, a recent study reported a drastic shift from non-homologous end joining-/microhomology-mediated end joining-mediated repair processes to a homology-directed repair-mediated process by inhibition of DNA polymerase  $\theta$  and DNA-dependent protein kinase, which led to a huge improvement in knock-in efficiency [29]. Such methods might also enable simultaneous insertion of exogenous DNAs into multiple GSH sites, or multi-copy integration at the same location, which can lead to increased recombinant protein production in a cell. Thus, improving the efficiency of our knock-in system may be a key factor in biotechnological applications of Pv11 cells.

Exogenous gene integration into genomic sites can occasionally cause changes in genomic structure or gene expression pattern, which in turn can cause alteration in cellular phenotypes (e.g., oncogenesis) [1]. Such phenotypic changes are a major concern, especially in human gene therapy applications [1]. In general, global gene expression patterns are assessed to confirm that the likelihood of any potential phenotypic changes in the future is low [3,10]. In other words, the most important point is whether the intrinsic cellular phenotypes can be maintained after transgenesis. Although we did not examine the gene expression patterns of the transgenic cell lines, intrinsic characteristics of Pv11 cells, namely the ability to anhydrobiosis and proliferation, were maintained even after one year in the cell lines (Figures 3 and 4). Hence, we believe that the current results are sufficient to derive our conclusion.

To begin to understand why the three sites, Chr1:21143572, Chr1:21155382, and Chr1:21164645, behave as GSHs in Pv11 cells, we acquired ATAC-seq data (GEO ID:

GSE190481) to assess chromatin accessibility. Figure S4 shows that the integration sites of the fragmented plasmid are in or near ATAC-seq peaks, which suggests that random integration of exogenous DNAs is most likely to occur in open chromatin regions. However, we cannot otherwise find any specific features of the GSH regions, such as larger or more consecutive ATAC-seq peaks, that might provide further clues. Further examination of chromatin conformation analysis may be needed, for example, using 5C and Hi-C methods [30,31]. These techniques detect active or inactive chromatin compartments on a genome-wide scale, and the information would help settle the question of why GSHs in Pv11 cells locate in a specific region of Chr1.

## 5. Conclusions

We identified GSHs in Pv11 cells that, when used for exogenous gene integration, had no effect on two phenotypes tested: anhydrobiotic ability and proliferation rate. Testing of cells obtained using one of the GSHs showed that the integrated gene functioned as designed. Furthermore, we constructed a basic tool kit to knock-in a GOI expression unit at this GSH. These resources will facilitate further genome engineering strategies, such as an inducible-expression system, and contribute to advancements in basic biology and applied biotechnology in Pv11 cells.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13030406/s1>, Figure S1: Agarose gel electrophoresis image of extracted genomic DNAs, Figure S2: Agarose gel electrophoresis image of genomic PCR in knock-in cell lines at the Chr1:280397 site, Figure S3: Anhydrobiotic ability and proliferation rate in cells produced by knock-in of a GOI-expression unit, Figure S4: Visualization of ATAC-seq peaks, Table S1: Oligo DNA sequences for construction of gRNA-expression vectors, Table S2: Primers for genomic DNA sequencing, Table S3: Primers for construction of the donor vectors in Figure 3, Table S4: Primers for construction of the donor vectors used in Figure 5, Table S5: Number of reads in the course of the analysis, Table S6: Integration sites of the fragmented plasmid sequences, Table S7: TPM values of genes around GSHs with annotations, Data S1: Complete sequence of the transfected plasmid in KH cells, Data S2–S8: Complete sequences of the plasmids used in Figure 4, Data S9 and S10: Raw reads of MinION sequencing relating to Figure 2B, Data S11–S19: Genome sequences of knock-in cell lines shown in Figures 3 and 4.

**Author Contributions:** Y.M. designed the project, performed the experiments, analyzed the data, contributed to discussions, and wrote the manuscript. S.T. and T.A. performed the experiments, analyzed the data, contributed to discussions, and wrote the manuscript. N.S., H.F., N.G., S.G., A.C., A.R. and G.G. performed the experiments and analyzed the data. R.D., R.C., E.S. and O.G. contributed to discussions and participated in editing the manuscript. T.K. organized the project, wrote the manuscript, and contributed to discussions. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a Grants-in-Aid for Scientific Research (KAKENHI) Grant (JP19J12030 to S.T.), and funded by a Strategic International Collaborative Research project promoted by the Ministry of Agriculture, Forestry and Fisheries, Japan (JPJ008837 to T.K.), and Adaptable and Seamless Technology transfer Program through Target-driven R&D (A-STEP) from Japan Science and Technology Agency (JST) (JPMJTR20UH to R.C.), and by a grant from the Russian Science Foundation No. 20-44-07002 (to E.S.). This work has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreements No. 734434 (to T.K.). The sequencing was supported by Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-15-2020-899) (O.G. and A.R.).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Raw sequencing data of the whole-genome libraries and the ATAC-seq generated in this study are available at NCBI GEO under accession number GSE190481 (token: qvslwuusvrihjyt).



**Acknowledgments:** We are grateful to Ken Naito (NARO) and Tomoe Shiratori (NARO) for the technical support of MinION sequencing and routine maintenance of Pv11 cells and transgenic cell lines, respectively.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Papapetrou, E.P.; Schambach, A. Gene insertion into genomic safe harbors for human gene therapy. *Mol. Ther.* **2016**, *24*, 678–684. [\[CrossRef\]](#)
- Cantos, C.; Francisco, P.; Trijatmiko, K.R.; Slamet-Loedin, I.; Chadha-Mohanty, P.K. Identification of “safe harbor” loci in indica rice genome by harnessing the property of zinc-finger nucleases to induce DNA damage and repair. *Front. Plant Sci.* **2014**, *5*, 302. [\[CrossRef\]](#)
- Brady, J.R.; Tan, M.C.; Whittaker, C.A.; Colant, N.A.; Dalvie, N.C.; Love, K.R.; Love, J.C. Identifying improved sites for heterologous gene integration using ATAC-seq. *ACS Synth. Biol.* **2020**, *9*, 2515–2524. [\[CrossRef\]](#)
- Shin, S.; Kim, S.H.; Shin, S.W.; Grav, L.M.; Pedersen, L.E.; Lee, J.S.; Lee, G.M. Comprehensive analysis of genomic safe harbors as target sites for stable expression of the heterologous gene in HEK293 cells. *ACS Synth. Biol.* **2020**, *9*, 1263–1269. [\[CrossRef\]](#)
- Hilliard, W.; Lee, K.H. Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis. *Biotechnol. Bioeng.* **2021**, *118*, 659–675. [\[CrossRef\]](#)
- Pellenz, S.; Phelps, M.; Tang, W.; Hovde, B.T.; Sinit, R.B.; Fu, W.; Li, H.; Chen, E.; Monnat, R.J., Jr. New human chromosomal sites with “safe harbor” potential for targeted transgene insertion. *Hum. Gene Ther.* **2019**, *30*, 814–828. [\[CrossRef\]](#)
- Gurumurthy, C.B.; Lloyd, K.C.K. Generating mouse models for biomedical research: Technological advances. *Dis. Models Mech.* **2019**, *12*, dmm029462. [\[CrossRef\]](#)
- Kawabe, Y.; Komatsu, S.; Komatsu, S.; Murakami, M.; Ito, A.; Sakuma, T.; Nakamura, T.; Yamamoto, T.; Kamihira, M. Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems. *J. Biosci. Bioeng.* **2018**, *125*, 599–605. [\[CrossRef\]](#)
- Larsen, J.S.; Karlsson, R.T.G.; Tian, W.; Schulz, M.A.; Matthes, A.; Clausen, H.; Petersen, B.L.; Yang, Z. Engineering mammalian cells to produce plant-specific N-glycosylation on proteins. *Glycobiology* **2020**, *30*, 528–538. [\[CrossRef\]](#)
- Papapetrou, E.P.; Lee, G.; Malani, N.; Setty, M.; Riviere, I.; Tirunagari, L.M.; Kadota, K.; Roth, S.L.; Giardina, P.; Viale, A.; et al. Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat. Biotechnol.* **2011**, *29*, 73–78. [\[CrossRef\]](#)
- Hamaker, N.K.; Lee, K.H. Site-specific integration ushers in a new era of precise CHO cell line engineering. *Curr. Opin. Chem. Eng.* **2018**, *22*, 152–160. [\[CrossRef\]](#)
- Lee, J.S.; Park, J.H.; Ha, T.K.; Samoudi, M.; Lewis, N.E.; Palsson, B.O.; Kildegaard, H.F.; Lee, G.M. Revealing key determinants of clonal variation in transgene expression in recombinant CHO cells using targeted genome editing. *ACS Synth. Biol.* **2018**, *7*, 2867–2878. [\[CrossRef\]](#)
- Watanabe, K.; Imanishi, S.; Akiduki, G.; Cornette, R.; Okuda, T. Air-dried cells from the anhydrobiotic insect, *Polypedilum vanderplanki*, can survive long term preservation at room temperature and retain proliferation potential after rehydration. *Cryobiology* **2016**, *73*, 93–98. [\[CrossRef\]](#)
- Cornette, R.; Kikawada, T. The induction of anhydrobiosis in the sleeping chironomid: Current status of our knowledge. *IUBMB Life* **2011**, *63*, 419–429. [\[CrossRef\]](#)
- Kikuta, S.; Watanabe, S.J.; Sato, R.; Gusev, O.; Nesselov, A.; Sogame, Y.; Cornette, R.; Kikawada, T. Towards water-free biobanks: Long-term dry-preservation at room temperature of desiccation-sensitive enzyme luciferase in air-dried insect cells. *Sci. Rep.* **2017**, *7*, 6540. [\[CrossRef\]](#)
- Sogame, Y.; Okada, J.; Kikuta, S.; Miyata, Y.; Cornette, R.; Gusev, O.; Kikawada, T. Establishment of gene transfer and gene silencing methods in a desiccation-tolerant cell line, Pv11. *Extremophiles* **2017**, *21*, 65–72. [\[CrossRef\]](#)
- Tokumoto, S.; Miyata, Y.; Deviatiiarov, R.; Yamada, T.G.; Hiki, Y.; Kozlova, O.; Yoshida, Y.; Cornette, R.; Funahashi, A.; Shagimardanova, E.; et al. Genome-wide role of HSF1 in transcriptional regulation of desiccation tolerance in the anhydrobiotic cell line, Pv11. *Int. J. Mol. Sci.* **2021**, *22*, 5798. [\[CrossRef\]](#)
- Miyata, Y.; Fuse, H.; Tokumoto, S.; Hiki, Y.; Deviatiiarov, R.; Yoshida, Y.; Yamada, T.G.; Cornette, R.; Gusev, O.; Shagimardanova, E.; et al. Cas9-mediated genome editing reveals a significant contribution of calcium signaling pathways to anhydrobiosis in Pv11 cells. *Sci. Rep.* **2021**, *11*, 19698. [\[CrossRef\]](#)
- Gusev, O.; Suetsugu, Y.; Cornette, R.; Kawashima, T.; Logacheva, M.D.; Kondrashov, A.S.; Penin, A.A.; Hatanaka, R.; Kikuta, S.; Shimura, S.; et al. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat. Commun.* **2014**, *5*, 4784. [\[CrossRef\]](#)
- Guizard, S.; Piegu, B.; Bigot, Y. DensityMap: A genome viewer for illustrating the densities of features. *BMC Bioinform.* **2016**, *17*, 204. [\[CrossRef\]](#)
- Miyata, Y.; Tokumoto, S.; Sogame, Y.; Deviatiiarov, R.; Okada, J.; Cornette, R.; Gusev, O.; Shagimardanova, E.; Sakurai, M.; Kikawada, T. Identification of a novel strong promoter from the anhydrobiotic midge, *Polypedilum vanderplanki*, with conserved function in various insect cell lines. *Sci. Rep.* **2019**, *9*, 7004. [\[CrossRef\]](#)

22. Sakuma, T.; Nakade, S.; Sakane, Y.; Suzuki, K.T.; Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat. Protoc.* **2016**, *11*, 118–133. [[CrossRef](#)]
23. Kallunki, T.; Barisic, M.; Jaattela, M.; Liu, B. How to choose the right inducible gene expression system for mammalian studies? *Cells* **2019**, *8*, 796. [[CrossRef](#)]
24. Tokumoto, S.; Miyata, Y.; Usui, K.; Deviatiiarov, R.; Ohkawa, T.; Kondratieva, S.; Shagimardanova, E.; Gusev, O.; Cornette, R.; Itoh, M.; et al. Development of a Tet-On inducible expression system for the anhydrobiotic cell line, Pv11. *Insects* **2020**, *11*, 781. [[CrossRef](#)]
25. Ma, H.; Wu, Y.; Dang, Y.; Choi, J.G.; Zhang, J.; Wu, H. Pol III promoters to express small RNAs: Delineation of transcription initiation. *Mol. Ther. Nucleic Acids* **2014**, *3*, e161. [[CrossRef](#)]
26. Crotty, S.; Pipkin, M.E. In vivo RNAi screens: Concepts and applications. *Trends Immunol.* **2015**, *36*, 315–322. [[CrossRef](#)]
27. Yu, J.S.L.; Yusa, K. Genome-wide CRISPR-Cas9 screening in mammalian cells. *Methods* **2019**, *164–165*, 29–35. [[CrossRef](#)]
28. Yang, H.; Ren, S.; Yu, S.; Pan, H.; Li, T.; Ge, S.; Zhang, J.; Xia, N. Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. *Int. J. Mol. Sci.* **2020**, *21*, 6461. [[CrossRef](#)]
29. Arai, D.; Nakao, Y. Efficient biallelic knock-in in mouse embryonic stem cells by in vivo-linearization of donor and transient inhibition of DNA polymerase theta/DNA-PK. *Sci. Rep.* **2021**, *11*, 18132. [[CrossRef](#)]
30. Dekker, J.; Marti-Renom, M.A.; Mirny, L.A. Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. *Nat. Rev. Genet.* **2013**, *14*, 390–403. [[CrossRef](#)]
31. Belmont, A.S. Large-scale chromatin organization: The good, the surprising, and the still perplexing. *Curr. Opin. Cell Biol.* **2014**, *26*, 69–78. [[CrossRef](#)]