ORIGINAL

# 4-Fragment Gateway cloning format for MosSCI-compatible vectors integrating Promoterome and 3'UTRome libraries of *Caenorhabditis elegans*

Toshiaki Kogame[1,2]

[1]*Systems Biology of Gene Regulatory Elements, Max-Delbrück-Center for Molecular Medicine, Berlin, Germany,* [2]*Division of Enzyme Pathophysiology, The Institute for Enzyme Research (KOSOKEN), the University of Tokushima, Tokushima, Japan*

**Abstract : The technique of *Mos1*-mediated Single Copy Insertion (MosSCI) now has become the essential technique which facilitates transgenic experiments for *Caenohabditis elegans* (*C. elegans*). Gateway system which is adopted to MosSCI-compatible vectors offers an advantage of simultaneous cloning with entry vectors cloned in the Gateway system format. On the other hand, the format for MosSCI-compatible vectors restricts flexibility in designing the vectors to only 3-fragment integration. Thus, construct of complex transgene such as the expression vector for translational gene fusion is tedious work even with Gateway system. We have developed the new recombination format called LeGaSCI (Library-enhanced Gateway for MosSCI) to expand the conventional 3-fragment to 4-fragment format which still retains the capacity to accept Promoterome and 3'UTRome libraries of *C. elegans*. In the new recombination format, 2 different Gateway format were combined. Cloning reaction for the tissue-specific expression vector of GFP-tagged protein with 3'UTR successfully occurred without any expected insertion, deletion or frame-shift mutation. Moreover, The MosSCI transgenic line was successfully generated with the construct. Collectively, we established the new Gateway system format which allows us to assemble 4-fragment insertion with the widest variety of entry clone vectors from *C. elegans* libraries. J. Med. Invest. 62 : 161-166, August, 2015**

***Keywords*** **:** *Gateway cloning, MosSCI, C. elegans, transgenic, plasmid construct*

## 1. INTRODUCTION

Transgenesis of *Caenohabditis elegans* (*C. elegans,* Supplementary Table 1 for the abbreviated terms) had until recently been a notorious problem due to the unstable gene expression affected by transgene silencing effect (1). However, *Mos1*-mediated Single Copy Insertion (MosSCI) has enabled the generation of stable transgenic *C. elegans* lines with one copy of the desired plasmid inserted into the decided locus of the genome (2). Subsequently many related techniques and variants have since been reported (3, 4). All of the transgenic techniques rely on the generation of transgenes and the cloning procedures are often laborious and time-consuming. The conventional cloning protocol requires DNA work including : PCR amplification with the DNA polymerase ; restriction enzyme digest ; and insertion into the vector by ligation. Due to the imperfect fidelity of the DNA polymerase, mutations in PCR products are often encountered, especially when the amplicons are longer. Sometimes the use of restriction enzymes which produce unspecific digested ends cannot be avoided, and therefore the orientation of insertion cannot be specified. Finally the confirmatory Sanger sequencing is almost always required to confirm the integrity of cloned DNA. The advent of Gateway cloning has revolutionalized standard cloning procedures. Gateway cloning was developed by the research for the site-specific recombination by the bacteriophage lambda recombination protein Integrase and its recognition of the "*att* site" (5). When the phage is integrated into *E. coli* genome, attB which stands for the bacterial attachment site on *E. coli* chromosome interacts with attP which corresponds to the specific sequence on the phage genome. Hence the recombination is referred to as BP reaction, which results in attL and attR sites which stands for Left and Right attachment site of the integrated phage. When the phage is excised from *E. coli* genome, attL site binds to attR site mediated by the Integrase, and the LR reaction inversely produces attB and attP sites again (6). Harnessing the site-specific recombination, the Gateway cloning consists of two processes. The BP reaction allows cloning of the PCR product flanked with attB sites into the entry vector with attP sites in the first step. The second step is the subcloning of the inserted sequence, from the entry clone into the destination vector by LR reaction. The recombination reaction using the integrase poses no risk of causing any mutations in the cloned DNA fragments, unlike conventional DNA work with the DNA polymerase. Therefore the sequence of the insert only needs to be validated once by Sanger sequencing, after it has been inserted by BP reaction, instead of sequencing after every step. Therefore once entry clones are generated they can be easily shuttled into multiple destination vectors, using the LR reaction, without any sequence validation. Moreover, there are several subtypes of *att* site, and each particular *att* site can specify the order and partners of recombination among the entry clones and destination vectors. Thus it allows for multiple clonings to be performed simultaneously whilst also enabling control over order, orientation and number of DNA inserts. This property has streamlined the large-scale cloning procedure. Especially in the *C. elegans* field, previous research has made good use of Gateway cloning and cloned functional DNA elements in a genome-wide scale in order to clarify their functions. The results were published as the milestones of "omics" research known as the studies of Promoterome, ORFeome and 3'UTRome (7-9). The *C.*

Received for publication January 20, 2015 ; accepted February 17, 2015.

Address correspondence and reprint requests to Toshiaki Kogame, MD, Department of Dermatology, Kansai Medical University, 2-5-1 Shinmachi, Hirakata, Osaka 573-1010, Japan and Fax : +81-72-804-2764.

*elegans* researchers adopted the Gateway cloning of the 3-fragment format, when they pioneered the omics research project (10). Thus, the entry clones from such libraries can efficiently assemble in the order of "Promoter : : ORF : : 3'UTR" into destination vectors in a single reaction (1). MosSCI-compatible vectors were also designed to serve as the destination vectors of the "user-friendly" Gateway cloning in the 3-fragment format, taking the genome-wide resources of the libraries into their consideration (2). However, in other words, this Gateway format constrains the flexibility and refuses any vectors inconsistent with the 3-fragment format. To date, some variations of Gateway cloning format are commercially available. They employed different alignment of Gateway *att* sites from the format adopted by *C. elegans* libararies (1). Thus, old legacies of the entry clone libraries in the 3-fragment format can be hardly used in the other formats. Due to the necessity of a more flexible construct which still accepts the *C. elegans* libraries, we searched for a new Gateway format and finally established the new recombination format named LeGaSCI (Library-enhanced Gateway for MosSCI).

Supplementary table 1 : The list of abbreviated terms.
The technical or abbreviated terms used in this work are shown in the list.

| abbreviation | full name |
|---|---|
| MosSCI | *Mos1*-mediated Single Copy Insertion |
| LeGaSCI | Library-enhanced Gateway for MosSCI |
| *att* site | attachment site |
| attP | the phage attachment site |
| attB | the bacterial attachment site |
| attL | Left attachment site |
| attR | Right attachment site |
| *C. elegans* | *Caenohabditis elegans* |
| *E. coli* | *Escherichia coli* |
| UTR | untranslated region |
| GFP | green fluorescent protein |
| UPRT | *uracil phosphoribosyltransferase* |
| iPAR-CLIP | *in vivo* Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation |

## 2. MATERIALS AND METHODS

### *2.1. Strains*

Wild type *C. elegans* strain N2 Bristol and strain EG6699 which carries *Mos1* insertion in Chr II were obtained from the Caenorhabditis Genetics Center (CGC ; Minneapolis, MN, USA). #195 worms which carry the construct *Pmyo-2* : : GFP : : UPRT : : *unc-54 3'UTR* were generated by the direct insertion method of MosSCI technique with the strain EG6699 (3).

### *2.2. Plasmids and Gateway cloning*

pCFJ90 and pCFJ68 were obtained from Addgene plasmid #19327 and Addgene plasmid #19325, respectively (Addgene, USA). HA-UPRT-pBS and pBCN27-R4R3 were kindly provided by Dr. C. Doe (11) and Dr. J. Semple (12), respectively. The transgene was assembled in LeGaSCI format with the combination of MultiSite Gateway® Three-Fragment Vector Construction Kit (Invitrogen, Gaithersburg, MD, USA, Cat.# 12537-023) and MultiSite Gateway® Pro plus (Invitrogen, Gaithersburg, MD, USA, Cat.# 12537-100). Briefly, *Pmyo-2* promoter, GFP-tag, UPRT and *unc-54 3'UTR* were cloned from pCFJ90, pCFJ68, HA-UPRT-pBS and pCFJ90 using the primer cassettes shown below. *Pmyo-2*, F : ggggactgcttttttgtacaaacttgcatttctgtgtctgacgatc, R : ggggacaactttgtatagaaaagttg catttatatctgagtagta. GFP-tag, F : ggggacaagtttgtacaaaaaagcaggctttatgagtaaaggagaagaact, R : ggggacaacttttgtatacaaagttgttttgtatagttcgtccatgc. ORF of UPRT, F : ggggacaactttgtatacaaaagttgttatggcgcaggtcccagcgag, R : ggggaccactttgtacaagaaagctgggtactacatggttccaaagtacc. *unc-54 3'UTR,* F : ggggacagctttcttgtacaaagtggtt catctcgcgcccgtgcctct, R : ggggacaactttgtataataaagttgt aagttggaaacagttatgtt. The PCR products were cloned into pDONR P4-P1R, pDONR P1-P5r, pDONR P5-P2 and pDONR P2R-P3 by BP reaction, respectively. 10 ng of each entry clone was subsequently mixed with 150 µg of MosSCI-compatible vector, pBCN27-R4R3. The solution was filled up to 8 µl with water and 2 µl of the LR clonase was added. The LR reaction was performed at 25°C for 24 hours followed by transformation of *E. coli.* The resultant construct named #195 was purified by the standard mini prep protocol.

### *2.3. Microinjection for MosSCI technique*

Microinjection was performed following the protocol of the direct insertion method described in the web page of Wormbuilder (http : //www.wormbuilder.org/).

## 3. RESULTS

### *3.1. The entry vectors of pDONR P1-P5r and pDONR P5-P2 can expand the potential of the MosSCI-compatible vectors to four-fragment integration*

Generally speaking, there are 2 major formats of Gateway cloning, namely the 3-fragment format and the 4-fragment format. The 3-fragment format which MosSCI-compatible vectors belong to carries the attR4-attR3 cassette in destination vectors whilst the 4-fragment format carries attR1-attR2 cassette in destination vectors and outperforms the flexibility and insertion number for plasmid construction conferred by the revised alignment of *att* sites (1, 13). Therefore the entry clones designed for one format are not compatible to another. Although both Gateway formats are used widely in the world according to any purpose, the 3-fragment format has more advantage in *C. elegans* field due to the accessibility to the genome-wide libraries than the 4-fragment format. Nevertheless, we investigated loopholes for these systems in order to find the new Gateway format which possesses advantages of both formats. After focusing attention on the combination of each component of two formats, it was noticed that entry clones in the middle position of MosSCI-compatible vectors can be substituted by two entry clones derived from pDONR P1-P5r and pDONR P5-P2 in the 4-fragment format, because the entry clones generate the recombination intermediate such as attL1 : : fragment 1 : : attB5 : : fragment 2 : : attL2 by the LR reaction and the middle position in the 3-fragment format is designed to accept the DNA fragment flanked with the attL1 and attL2 (Figure 1). It was tempting to investigate if the new combination of entry clones could perform the recombination into the attR4-attR3 cassette in the proper order, that is : Promoter, ORF1, ORF2 and 3'UTR from upstream to downstream. pDONR P4-P1R and pDONR P2R-P3 were used to clone a promoter and a 3'UTR sequence respectively, since the Promoterome and 3'UTRome libraries were cloned in the 3-fragment format (1). pDONR P1-P5r and pDONR P5-P2 were designed to clone ORF1 and ORF2, assuming a tag sequence in ORF1 and a gene of interest in ORF2 for N-terminally tagged proteins or *vice versa* for C-terminal tagged proteins. We generated entry clones that code the *myo-2* promoter, a GFP tag, the ORF of UPRT (*uracil phosphoribosyltransferase*) and the *unc-54 3'UTR*, respectively (2, 11). The four entry clones and a MosSCI-compatible vector, pBCN27-R4R3, were mixed for the LR reaction and the recombinant plasmid was purified. Because the usage in Figure 1 is not originally

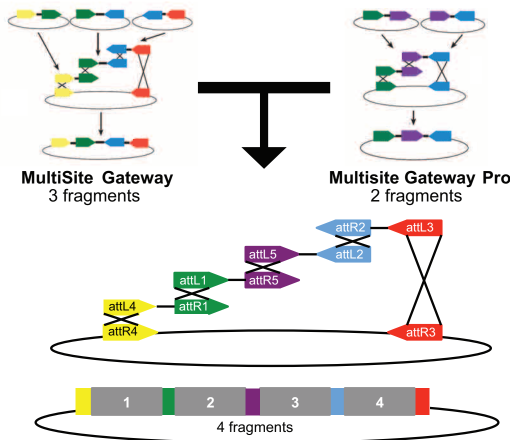


Figure 1 : A new recombination platform integrates existing genome-wide *C. elegans* Promoterome and 3'UTRome libraries.
A : Four-fragment recombination allows integration of Promoterome library (position 1), ORF1 and ORF2 (positions 2 and 3) and 3'UTRome library (position 4) to allow easy construction of *C. elegans* transgenes for tissue-specific expression of tagged proteins (ORF1 fused to ORF2).

intended by the manufacturer, it was necessary to confirm the precise sequence of the resulting construct by Sanger sequence. The sequence result showed the orderly alignment of the four fragments (*Pmyo-2* : : GFP : : UPRT : : *unc-54 3'UTR* in pBCN27). The recombinant linker sequences derived from the *att* sites of the entry clones were also yielded without unexpected mutations (Figure 2).

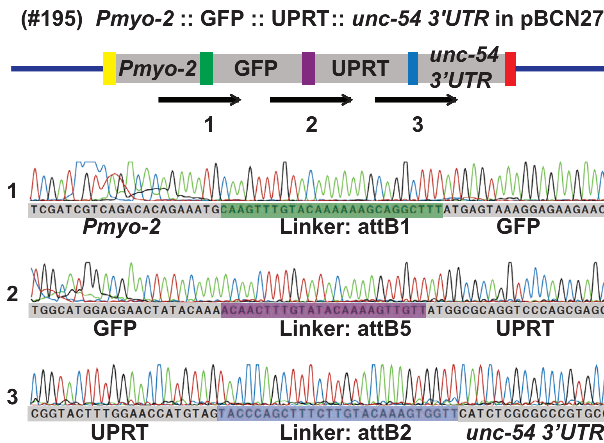


Figure 2 : Sanger sequencing revealed that each fragment was assembled with linker sequences as intended.
A confirmatory sequence showed that the 4 fragments from entry vectors were assembled in the expected position without inversion or truncation and that the recombinant linker sequences carry no mutations such as deletion, insertion or flame-shift mutation.

We confirmed that the new Gateway format recombines four entry clones into a single MosSCI-compatible vector without any inversion, deletion or mutation of the inserted fragments. Additionally, this indicates that the primer design for the BP reaction complies with the original protocol and does not require any modification. The new Gateway format was named LeGaSCI (Library-enhanced Gateway for MosSCI).

*3.2. The recombinant construct served as a translational gene fusion vector in MosSCI technique*

MosSCI is one of the most reliable techniques to generate stable transgenic lines of *C. elegans* (1, 2). Thus LeGaSCI format can improve the flexibility of vector-design for MosSCI and enables the streamlining of the generation of stable transgenic lines carrying more complex transgenes. We next investigated if the recombinant construct can generate stable transgenic lines using MosSCI. MosSCI was performed with the recombinant construct shown in Figure 2 (*Pmyo-2* : : GFP : : UPRT : : *unc-54 3'UTR* in pBCN27). It was designed to express GFP-tagged UPRT protein in the pharynx. Thus the GFP signal should only be seen in the pharynx of MosSCI lines. After microinjection was performed with the recombinant construct, the stable transgenic MosSCI lines were selected by recovery of *unc-119* mutant phenotype and loss of negative selection markers according to the MosSCI technique protocol (3). The GFP signal from the generated MosSCI line was searched under the fluorescent microscope. The pharynx in the MosSCI line exhibited high intensity GFP signal whereas no signal was observed in wild type (Figure 3). This indicates GFP-tagged UPRT protein was tissue-specifically expressed in the pharynx as intended.

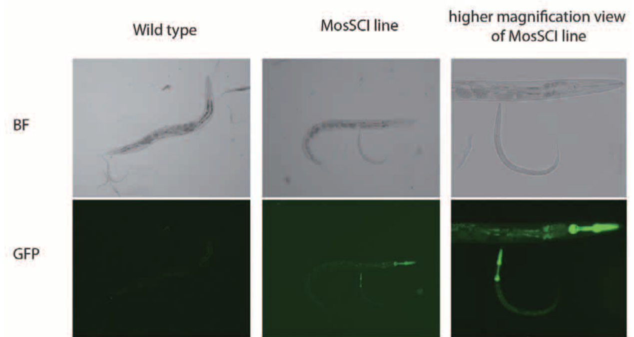


Figure 3 : The MosSCI line generated by the construct exhibited the high intensity signal of GFP in the pharynx.
Bright field (BF) and GFP fluorescence microscopy of wild-type N2 and MosSCI line #195 (carrying the construct of *Pmyo-2* : : GFP : : UPRT : : *unc-54 3'UTR* in pBCN27). The wild-type N2 shows the slight background signal from the gut, whilst #195 indicated the high intensity signal of GFP from the pharynx. The high magnification image delineates the shape of the pharynx which the *myo-2* promoter is known to drive the gene expression in.

Collectively, the recombinant construct generated in LeGaSCI format was confirmed to serve as a translational gene fusion vector in MosSCI. Thus we have established the more flexible 4-fragment new Gateway format, which can accept the Promoterome and 3'UTRome libraries.

## DISCUSSION

Here the 4-fragment Gateway format optimized for *C. elegans* MosSCI named LeGaSCI is established. It augments the potentials of MosSCI by coupling flexibility of the insert number, with the widest accessibility to entry clones, allowing insertion of the Promoterome as well as the 3'UTRome libraries, which are both publicly available (1). Further 13 MosSCI vectors were successfully generated in LeGaSCI format without any mutations detected upon Sanger sequencing (Supplementary table 2 and 3). This indicates that LeGaSCI format can accept any entry vectors from the genome-wide libraries. LeGaSCI format is particularly intended for the

Supplementary table 2 : The list of the entry clones.
Entry clones were generated as indicated or obtained from Addgene (USA).

| Entry clones | Cloned sequence | PCR template | Backbone vector | Fwd primer | Rvs primer |
|---|---|---|---|---|---|
| A1 | *Pmyo-2* | pCFJ90 | pDONR P4-P1R | GGGGACAACTTTGTATAGAAAAGTTG atttctgtgtctgacgatc | GGGGACTGCTTTTTTGTACAAACTTG cattttatatctgagtagta |
| A2 | *Phsp-16.48* | pJL44 | pDONR P4-P1R | GGGGACAACTTTGTATAGAAAAGTTG tggacggaaatagtggtaaa | GGGGACTGCTTTTTTGTACAAACTTG tcttgaagtttagagaatga |
| A3 | *Pglh-2* | pJL43.1 | pDONR P4-P1R | GGGGACAACTTTGTATAGAAAAGTTG gaattcatattaccattatt | GGGGACTGCTTTTTTGTACAAACTTG tacaagtttggcgtttgcaa |
| A7 | *Phis-72* | *C. elegans* Genome | pDONR P4-P1R | GGGGACAACTTTGTATAGAAAAGTTG ttaaattgagaattgagaatgg | GGGGACTGCTTTTTTGTACAAACTTG ttggctgagtatcaaacgtta |
| A8 | *Ppie-1* | pCM1.127* | pDONR P4-P1R | | |
| | | | | | |
| B3 | A(5)-tag | pMK38 | pDONR P1-P5r | GGGGACAAGTTTGTACAAAAAAGCAGGCTTT atgatgggcagtgtcgagct | GGGGACAACTTTTGTATACAAAGTTGT cactttatgcttccggctcg |
| B4 | luciferase | SCP2_CMVenh-luc-1 (22) | pDONR P1-P5r | GGGGACAAGTTTGTACAAAAAAGCAGGCTTT atggaagacgccaaaaacat | GGGGACAACTTTTGTATACAAAGTTGT cacggcgatctttccgccct |
| B6 | mCherry | pCFJ104 | pDONR P1-P5r | GGGGACAAGTTTGTACAAAAAAGCAGGCTTT atggtctcaaagggtgaaga | GGGGACAACTTTTGTATACAAAGTTGT cttatacaattcatccatgc |
| B7 | GFP | pCFJ68 | pDONR P1-P5r | GGGGACAAGTTTGTACAAAAAAGCAGGCTTT atgagtaaaggagaagaact | GGGGACAACTTTTGTATACAAAGTTGT tttgtatagttcgtccatgc |
| | | | | | |
| C1 | A(3)-tag | pMK43 (23) | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atgatgggcagtgtcgagc | GGGGACCACTTTGTACAAGAAAGCTGGGTA tcaagctctgctcttgcact |
| C2 | GFP | pCFJ68 | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atgagtaaaggagaagaact | GGGGACCACTTTGTACAAGAAAGCTGGGTA ttatttgtatagttcgtc |
| C3 | mCherry | pCFJ104 | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atggtctcaaagggtgaaga | GGGGACCACTTTGTACAAGAAAGCTGGGTA ttacttatacaattcatcca |
| C5 | OsTIR | pNHK36 (23) | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atgacgtacttcccggagga | GGGGACCACTTTGTACAAGAAAGCTGGGTA ctataggattttaacaaaat |
| C6 | TEV | pDS5 (24) | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atggcgatgcaagcggccaa | GGGGACCACTTTGTACAAGAAAGCTGGGTA ttacaattgagtcgcttcct |
| C7 | UPRT | pbluescript-HA-UPRT | pDONR P5-P2 | GGGGACAACTTTGTATACAAAAGTTGTT atggcgcaggtcccagcgag | GGGGACCACTTTGTACAAGAAAGCTGGGTA ctacatggttccaaagtacc |
| | | | | | |
| D1 | *glh-2 3'UTR* | pJL43.1 | pDONR P2R-P3 | GGGGACAGCTTTCTTGTACAAAGTGGTT cttgtccatattgtatcagta | GGGGACAACTTTGTATAATAAAGTTGT gcaaaaaaatgaatcataac |
| D2 | *unc-54 3'UTR* | pCFJ90 | pDONR P2R-P3 | GGGGACAGCTTTCTTGTACAAAGTGGTT catctcgcgcccgtgcctct | GGGGACAACTTTGTATAATAAAGTTGT aagttggaaacagttatgtt |
| D3 | *tbb-2 3'UTR* | pCM1.36* | pDONR P2R-P3 | | |

* : plasmids obtained from Addgene (USA)
Numbers in parentheses indicate numbers of references

Supplementary table 3 : The list of the construct for MosSCI.
With the entry clones shown in the supplementary table 2, 13 MosSCI constructs were generated besides the construct indicated in the main text.

| Construct | Cloned sequence | Slot1 | Slot2 | Slot3 | Slot4 | Backbone vector |
|---|---|---|---|---|---|---|
| #124 | *Ppie-1* : : luc : : A(3)-tag : : *glh-2 3'UTR* | A8 : *Ppie-1* | B4 : luciferase | C1 : A(3)-tag | D1 : *glh-2 3'UTR* | pCFJ150 |
| #145 | *Ppie-1* : : mCherry : : OsTIR : *glh-2 3'UTR* | A8 : *Ppie-1* | B6 : mCherry | C5 : OsTIR | D1 : *glh-2 3'UTR* | pCFJ150 |
| #148 | *Phsp-16.48* : : GFP : : TEV : : *glh-2 3'UTR* | A2 : *Phsp-16.48* | B7 : GFP | C6 : TEV | D1 : *glh-2 3'UTR* | pCFJ150 |
| #150 | *Ppie-1* : : GFP : : UPRT : : *glh-2 3'UTR* | A8 : *Ppie-1* | B7 : GFP | C7 : UPRT | D1 : *glh-2 3'UTR* | pBCN27-R4R3 |
| #151 | *Ppie-1* : : luc : : GFP : : *glh-2 3'UTR* | A8 : *Ppie-1* | B4 : luciferase | C2 : GFP | D1 : *glh-2 3'UTR* | pCFJ150 |
| #152 | *Ppie-1* : : luc : : mCherry : : *glh-2 3'UTR* | A8 : *Ppie-1* | B4 : luciferase | C3 : mCherry | D1 : *glh-2 3'UTR* | pCFJ150 |
| #158 | *Phsp-16.48* : : GFP : : TEV : : *tbb-2 3'UTR* | A2 : *Phsp-16.48* | B7 : GFP | C6 : TEV | D3 : *tbb-2 3'UTR* | pBCN27-R4R3 |
| #176 | *Ppei-1* : : A(5)-tag : : GFP : : *tbb-2 3'UTR* | A8 : *Ppie-1* | B3 : A(5)-tag | C2 : GFP | D3 : *tbb-2 3'UTR* | pBCN27-R4R3 |
| #192 | *Pmyo-2* : : A(5)-tag : : GFP : : *unc-54 3'UTR* | A1 : *Pmyo-2* | B3 : A(5)-tag | C2 : GFP | D2 : *unc-54 3'UTR* | pCFJ210 |
| #194 | *Phsp-16.48* : : GFP : : OsTIR : : *unc-54 3'UTR* | A2 : *Phsp-16.48* | B7 : GFP | C5 : OsTIR | D2 : *unc-54 3'UTR* | pBCN27-R4R3 |
| #195 | *Pmyo-2* : : GFP : : UPRT : : *unc-54 3'UTR* | A1 : *Pmyo-2* | B7 : GFP | C7 : UPRT | D2 : *unc-54 3'UTR* | pBCN27-R4R3 |
| #196 | *Phsp-16.48* : : GFP : : UPRT : : *unc-54 3'UTR* | A2 : *Phsp-16.48* | B7 : GFP | C7 : UPRT | D2 : *unc-54 3'UTR* | pBCN27-R4R3 |
| #197 | *Pglh-2* : : GFP : : HA-UPRT : : *unc-54 3'UTR* | A3 : *Pglh-2* | B7 : GFP | C7 : UPRT | D2 : *unc-54 3'UTR* | pBCN27-R4R3 |
| #198 | *Phis-72* : : GFP : : HA-UPRT : : *tbb-2 3'UTR* | A7 : *Phis-72* | B7 : GFP | C7 : UPRT | D3 : *tbb-2 3'UTR* | pBCN27-R4R3 |

construction of tagged protein expression vectors for. In order to construct MosSCI vectors expressing tagged proteins, the translational PCR fusion (also known as overlap extension PCR or PCR stitching) is recommended as a standard procedure (1). Sequences of the tag and the gene of interest are stitched by PCR and subcloned into one of the three entry vectors. Hence the 3-fragment format has two more slots in which to accept the entry clones including the libraries. However, the confirmatory Sanger sequence should be repeated for the subcloned tag sequence owing to the theoretically imperfect fidelity of the DNA polymerase. Thus, this procedure requires duplication of effort and becomes the bottleneck of high-throughput cloning using the Gateway system. On the other

hand, without re-sequencing, tagged protein expression vectors are assembled from entry clones, in a single LR reaction, once a repertoire of entry clones for the validated tag-coding sequences has been prepared in LeGaSCI format.

*C. elegans* researchers suffer from insufficient resources of antibodies against many sorts of native proteins. The tag sequence can also be used as the affinity tag for protein purification. Thus the solution for the issue is to use transgenic lines that express fusion proteins with affinity tags against which antibodies are readily available. Therefore LeGaSCI format aids genome-wide scale experiments in *C. elegans,* and allows streamlining and simplification of the process. For instance, recently the novel method, "iPAR-CLIP" was established by one group (15). As the name suggests, the experiments are based on the technique of Cross-Linking and Immunoprecipitation which purifies the binding partners of the target molecule. Thus they cross-linked the RNA molecules bound to the protein of interest and purified them. They then annotated all the RNA bound to the protein by a deep sequencer. The technique takes advantage of the *C. elegans* transgenic line which expresses the Gld-1 : : GFP : : FLAG fusion protein. They performed CLIP with the anti-FLAG antibody and clarified genome-wide mRNAs bound to the Gld-1 protein. Thus LeGaSCI format can provide easy production of transgenic lines for any tagged protein. Because of the apparent necessity for transgenic lines expressing tagged proteins in the *C. elegans* field, one group has already launched the genome-scale project, TransgenOme for the tag-based analysis of genome-wide transgene behavior. They used the fosmid library which is designed for expression vectors of C-terminally GFP-tagged proteins because the fosmids usually contain approximately 35-40 kb stretch of genomic sequences which encompass the intact regulatory elements. They generated the transgenic lines with the fosmid constructs and clarified multiple gene behaviors with transgenic lines of 230 different genes (16). The project is still ongoing and constantly increasing the number of strains available for the public, however it has yet to be completed for as many as approximately 20000 genes (17). Therefore you may not be able to obtain the transgenic lines of interest from the TransgenOme repository. Moreover, modification of the tag can be required in CLIP experiments since it is known that the tag sequence can interfere with protein folding and alter the conformation. Although a lot of the protocols to generate fosmid constructs are described in literature, it is not as easy as Gateway cloning (18). Thus LeGaSCI format can provide easier access to transgenic resources than TransgenOme library in some contexts.

*C. elegans* is a multicellular organism, with a transparent body, hence fluorescent proteins such as GFP can be easily used to monitor gene expression, giving information on tissue-specificity. Thus *C. elegans* was used to monitor the spatial and temporal regulation of *in vivo* gene function using fluorescent proteins. Such transgenic approaches, especially in the high-throughput setting, always suffer from experimental noise due to the artificial manipulation of a living system. Hence it is essential to recapitulate the gene functions in a more reliable experimental design. Though the *C. elegans* Promoterome studies revealed promoter activity in a genome-wide scale, part of the results from the promoter : : reporter constructs conflict the data from the reporter lines of the TransgenOme, whose constructs carry intact gene structure and the native regulatory elements (16). This clearly indicates the importance of regulatory elements outside of promoters. Furthermore, many reports clarified that the 3'UTR plays important roles in regulating gene expression, especially in the gonadal and embryonic stages in *C. elegans* (19, 20). Thus integration of the regulatory elements is fundamentally important to monitor gene behaviors in the developmental stage. Previous researchers used *C. elegans* as the platform to analyze splicing isoform of genome-wide transcripts in order to recapitulate the spatial and temporal regulation of the gene expression *in vivo.* They established the method to generate fluorescent alternative splicing reporters (21). The reporter construct in the method was designed with Gateway destination vector carrying attR1-attR2 cassette. For the control of tissue specificity, they constructed 20 destination vectors that possess different kinds of promoters and 3'UTRs for major tissues in order to conduct genome- and organism-wide applications. Hence 20 destination vectors provide with the resolution of tissue-specificity, which enables the analysis of the expression profile of the splicing isoforms. However, we can not exclude the experimental noise owing to imperfect regulatory elements for gene regulation in the approach, and strategy can be improved by constructing reporter transgenes with regulatory elements more specific to tissue or cell types. Promoterome library consists of ~ 6000 promoters for predicted genes and 3'UTRome library is composed of 3'UTRs for 7105 CDSs (6741 genes) (7, 9). Hence, LeGaSCI format can provide literally countless combination of regulatory element components, and therefore augments the resolution of tissue- or cell type-specific analysis by the number of the combination. There is another point to consider in regards to experimental noise : The previous reporter constructs conducted for genome-wide analysis such as TransgenOme are not designed for MosSCI. Therefore they do not ensure one copy insertion in the decided locus of the genome whilst the new format can avoid this issue. In conclusion, combining MosSCI and regulatory elements from the libraries can thus provide one of the most reliable platforms to monitor the gene expression using fluorescent tagged proteins (Figure 4).

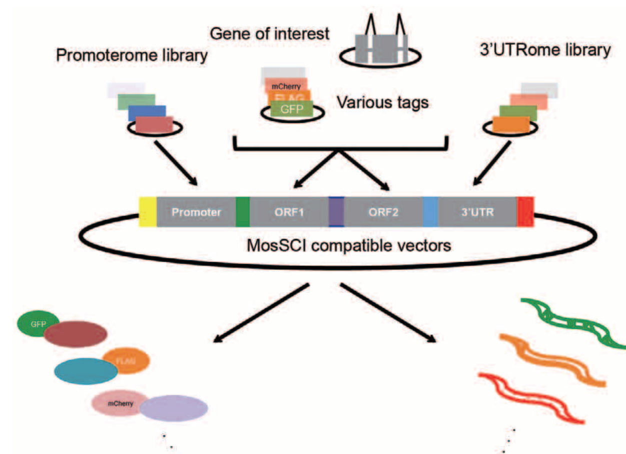


Figure 4 : The Overview of the "LeGaSCI" format.
The 2 slots in the middle of MosSCI-compatible vectors enable to flexibly design the various tagged proteins once tags are cloned in LeGaSCI format. The entry clones from Promoterome and 3'UTRome libraries are integrated into the format in suitable order as protein expression vectors, which allows to design the fine-tuned gene expression by the regulatory elements.

Collectively, we established the new Gateway format named LeGaSCI intended for tagged proteins, integrating the regulatory elements from the Promoterome and 3'UTRome libraries. LeGaSCI format adds more flexibility to the designing of *C. elegans* transgenic lines while it streamlines the time-consuming cloning process. We believe LeGaSCI format allows the realization of the potential of the *C. elegans* libraries in genome-wide researches and expands the field of *C. elegans* in life science to more complex applications with more high-throughput experimental designs.

## STATE OF ALL FUNDING SOURCES THAT SUPPORTED THE WORK

This work was supported as a subproject through funding by the German Bundesministerium fuer Bildung und Forschung (BMBF) for the Berlin Institute of Medical Systems Biology (BIMSB), 0314362.

## CONFLICTS OF INTEREST DISCLOSURES :

None declared.

## ACKNOWLEDGEMENT

This project was carried out in the laboratory of Nikolaus Rajewsky at the Max-Delbrueck Center and was supported as a subproject through funding by the German Bundesministerium fuer Bildung und Forschung (BMBF) for the Berlin Institute of Medical Systems Biology (BIMSB), 0314362. Dr. E. Jorgensen and Dr. C. Frøkær-Jensen kindly provided me with materials as well as technical advices for MosSCI technique. pBCN27-R4R3, HA-UPRT-pBS, SCP2_CMVenh-luc-1, pDS5 were the kind gifts from Dr. J. Semple, Dr. C. Doe, Dr. JT Kadonaga and Dr. M. Knop, respectively. Dr. M. Kanemaki generously allowed us to use pNHK36, pMK38 and pMK43. pPCM1.127 (Addgene Plasmid #21384) and pCM1.36 (Addgene Plasmid #17249) were obtained from the plasmid repository at Addgene (USA). R. Eccles kindly revised the manuscripts.

## REFERENCE


1. Merritt C, Seydoux G : Transgenic solutions for the germline. WormBook : 1-21, 2010
2. Frøkjær-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, Jorgensen EM : Single-copy insertion of transgenes in *Caenorhabditis elegans.* Nat Genet 40(11) : 1375-83, 2008
3. Frøkjær-Jensen C, Davis MW, Ailion M, Jorgensen EM : Improved *Mos1*-mediated transgenesis in *C. elegans.* Nat Methods 9(2) : 117-8, 2012
4. Frøkjær-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella M, Pozniakovsky A, Moerman DG, Jorgensen EM : Random and targeted transgene insertion in *Caenorhabditis elegans* using a modified *Mos1* transposon. Nat Methods 11 (5) : 529-34, 2014
5. Hartley JL, Temple GF, Brasch MA : DNA Cloning Using *In Vitro* Site-Specific Recombination. Genome Res 1 ; 10(11) : 1788-95, 2000
6. Enquist LW, Nash H, Weisberg RA : Strand exchange in site-specific recombination. Proc Natl Acad Sci U S A 76(3) : 1363-7, 1979
7. Dupuy D, Li QR, Deplancke B, Boxem M, Hao T, Lamesch P, Sequerra R, Bosak S, Doucette-Stamm L, Hope IA, Hill DE, Walhout AJ, Vidal M : A First Version of the *Caenorhabditis elegans* Promoterome. Genome Res 14(10B) : 2169-75, 2004
8. Brasch MA, Hartley JL, Vidal M : ORFeome cloning and systems biology : standardized mass production of the parts from the parts-list. Genome Res 14(10B) : 2001-9, 2004
9. Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, Mis E, Zegar C, Gutwein MR, Khivansara V, Attie O, Chen K, Salehi-Ashtiani K, Vidal M, Harkins TT, Bouffard P, Suzuki Y, Sugano S, Kohara Y, Rajewsky N, Piano F, Gunsalus KC, Kim JK : The landscape of *C. elegans* 3'UTRs. Science 23 ; 329(5990) : 432-5, 2010
10. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, Moore T, Hudson JR Jr, Hartley JL, Brasch MA, Vandenhaute J, Boulton S, Endress GA, Jenna S, Chevet E, Papasotiropoulos V, Tolias PP, Ptacek J, Snyder M, Huang R, Chance MR, Lee H, Doucette-Stamm L, Hill DE, Vidal M : Nat Genet 34(1) : 35-41, 2003
11. Miller MR, Robinson KJ, Cleary MD, Doe CQ : TU-tagging : cell type-specific RNA isolation from intact complex tissues. Nat Methods 6(6) : 439-41, 2009
12. Semple JI, Garcia-Verdugo R, Lehner B : Rapid selection of transgenic *C. elegans* using antibiotic resistance. Nat Methods 7(9) : 725-7, 2010
13. Petersen LK, Stowers RS : A Gateway MultiSite recombination cloning toolkit. PLoS One 6(9) : e24531, 2011
14. Boulin T, Etchberger JF & Hobert O : Reporter gene fusions. WormBook 1-23, 2006
15. Jungkamp AC, Stoeckius M, Mecenas D, Grün D, Mastrobuoni G, Kempa S, Rajewsky N : In vivo and transcriptome-wide identification of RNA binding protein target sites. Mol Cell 9 ; 44(5) : 828-40, 2011
16. Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, Hasse S, Rupprecht M, Vinis E, Tinney M, Preston E, Zinke A, Enst S, Teichgraber T, Janette J, Reis K, Janosch S, Schloissnig S, Ejsmont RK, Slightam C, Xu X, Kim SK, Reinke V, Stewart AF, Snyder M, Waterston RH, Hyman AA : A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. Cell 17 ; 150(4) : 855-66, 2012
17. Hodgkin J : What does a worm want with 20,000 genes? Genome Biol 2(11), 2001
18. Tursun B, Cochella L, Carrera I, Hobert O : A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans.* PLoS One 4(3) : e4625, 2009
19. Merritt C, Rasoloson D, Ko D, Seydoux G : 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. Curr Biol 14 ; 18(19) : 1476-82, 2008
20. Oldenbroek M, Robertson SM, Guven-Ozkan T, Spike C, Greenstein D, Lin R : Regulation of maternal Wnt mRNA translation in *C. elegans* embryos. Development 140(22) : 4614-23, 2013
21. Kuroyanagi H, Ohno G, Sakane H, Maruoka H, Hagiwara M : Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic *Caenorhabditis elegans.* Nat Protoc 5(9) : 1495-517, 2010
22. Juven-Gershon T, Cheng S, Kadonaga JT : Rational design of a super core promoter that enhances gene expression. Nat Methods 3(11) : 917-22, 2006
23. Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M : An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods 6(12) : 917-22, 2009
24. Taxis C, Stier G, Spadaccini R, Knop M : Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. Mol Syst Biol 5(267) : 267, 2009