

**Development of functional cellomics for comprehensive analysis of the relationship
between neural networks and behavior in *Caenorhabditis elegans***

Yuji Yamauchi

2023

Contents

Abbreviations	...	p.2	
Introduction	...	p.3–p.15	
Chapter I.	High-throughput functional annotation of <i>Caenorhabditis elegans</i> neural network using a cellomics method	...	p.16–p.29
Chapter II.	Cre-<i>lox</i> engineering with machine learning to achieve sparse labeling at a desired sparseness level	...	p.30–p.51
Chapter III.	Neuronal cell subclass-specific proteomic analysis of the <i>Caenorhabditis elegans</i>	...	p.52–p.69
Acknowledgements	...	p.70	
Publications	...	p.71	

Abbreviations

SSR	site-specific recombinase
MORF	mononucleotide repeat frameshift
MARCM	mosaic analysis with repressive cell markers
MADM	mosaic analysis with dual markers
SPARC	sparse predictive activity through recombinase competition
STARS	stochastic gene activation with genetically regulated sparseness
AFD	amphid finger cell D
mRNA	messenger RNA
tRNA	transfer RNA
BONCAT	bioorthogonal non-normal amino acid tagging
Azf	p-azide- <i>L</i> -phenylalanine
MuPheRS	mutated phenylalanyl-tRNA synthetase
PheRS	phenylalanyl-tRNA synthetase
ATR	all-trans-retinal
NGM	nematode growth medium
Ex array	Extrachromosomal array
HSN	hermaphrodite-specific
VC	Ventral C
PCR	polymerase chain reaction
SD medium	synthetic defined medium
qPCR	quantitative PCR
RBE	recombinase binding element
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal Bovine Serum
GPR	Gaussian process regression
LC	Liquid chromatography
MS/MS	Tandem mass spectrometry
GPCR	G protein-coupled receptors
SPE	solid-phase extraction

Introduction

In multicellular organisms, complex biological phenomena emerge when many cells form a network. In the brain, for example, as many as 86 billion neurons in humans and 100 million neurons in mice form a complex network structure that generates complex brain functions such as learning and memory [1, 2]. The cell types that make up the network are also diverse, with various cells in the brain differentiated to have unique functions, including sensory neurons, interneurons, motor neurons. However, the mechanism by which neuronal networks create higher-order life phenomena remains a mystery and is one of the greatest mysteries of the 21st century.

The data-driven methodology is a powerful weapon for understanding complex biological phenomena such as brain function. In recent years, high-throughput sequencer and mass spectrometer technologies have developed remarkably [3, 4]. We are now in a situation where the dynamics of a vast number of molecules can be studied at once (**Figure 1**). As a result, data-driven analysis of the genome, transcriptome, proteome, and metabolome levels has become relatively easy, and the characterization and modeling of complex biological systems are progressing. However, it is not easy to directly adapt these omics analyses targeting biomolecules to life phenomena at the individual level. This is because it is not apparent which cells in an individual perform essential functions and should be analyzed. Therefore, to understand individual-level life phenomena that emerge when many cells form a network, a cell-level omics method that enables comprehensive annotation of the effects of each cell on life phenomena is considered necessary.

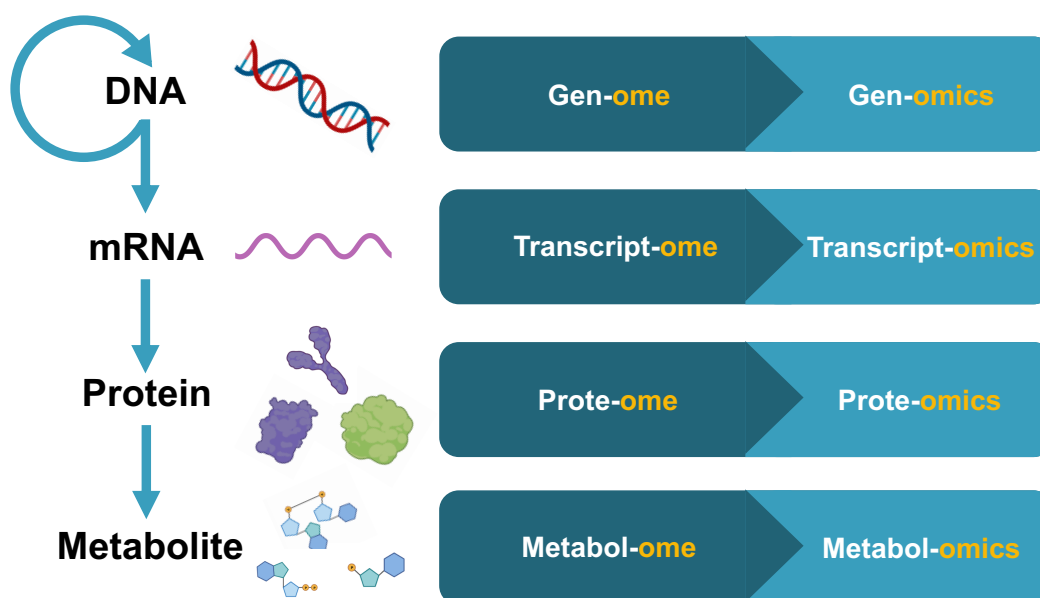


Figure 1. Multi-omics approaches. The figure is made using Biorender (<https://biorender.com/>).

***Caenorhabditis elegans* as a model organism in the field of neurology**

One of the primary goals of neuroscience is to understand how computation is performed on neural networks. *C. elegans* is a beneficial model organism for neuroscience research (**Figure 2**). *C. elegans* hermaphrodites have a simple neural network of 302 neurons, and a comprehensive map of their neural connections, the connectome, has already been revealed [5, 6]. In addition, *C. elegans* exhibit various behaviors, such as thermotaxis and chemotaxis, and various behavior assays have been established [7, 8]. However, more than 30 years after the *C. elegans* connectome was elucidated in 1986, the relationship between nematode behavior and neural networks has yet to be comprehensively clarified.

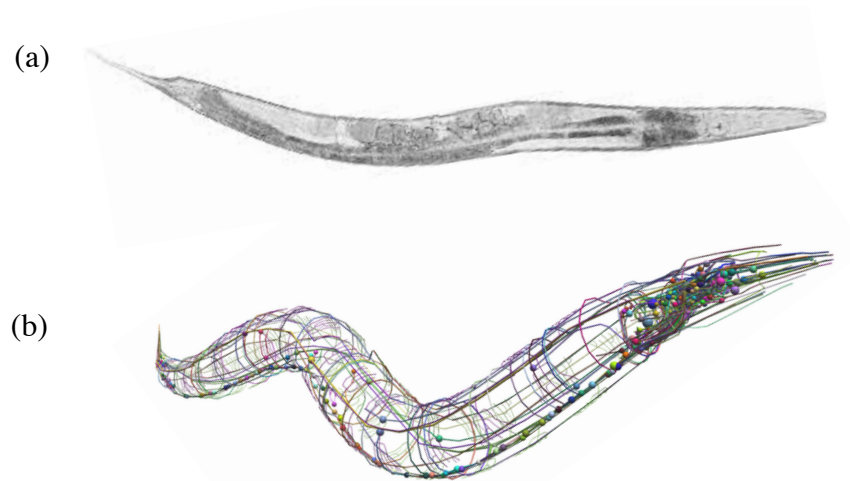


Figure 2. *Caenorhabditis elegans*. **(a)** The picture of *C. elegans* (the picture is from Nematode.org). **(b)** Connectome of *C. elegans* (the picture is from mathinsight.org).

Research methods for the identification of neural function and problems with existing research methods

Data-driven science, which accumulates vast amounts of data with no hypotheses, has deepened our understanding of complex biological processes. For example, the application of state-of-the-art mass spectrometers and next-generation sequencers streamlines our understanding of diverse biological processes [3, 4]. However, even with such omics analysis, it is challenging to gain a deeper understanding of the individual-level biological processes produced by complex neural networks.

In recent years, a data-driven approach, which can be called structural cell omics, is bringing about a transformation. For example, serial electron microscopy combining a microtome or a focused ion beam with electron microscopy enables high-throughput acquisition of 3D structures of cellular networks with sub-nanometer resolution [9]. In addition, combining tissue clearing and various staining methods makes it possible to rapidly evaluate the cellular network structure of organs and even individuals (**Figure 3**) [10]. In neuroscience, attempts are underway to use these structural cell omics to elucidate the connectome in various animal species, from *Drosophila* to humans. Such comprehensive structural information will provide valuable insights into the function of cellular networks. In addition, new approaches, such as Brainbow and Optobow, have been developed, attempting to reveal brain structures by optical microscopy [11, 12].

Structural information is vital for understanding how the brain works, but more is needed. It is necessary to accumulate comprehensive neurophysiological and behavioral data on which parts of the brain are functioning and to construct a behavioral model based on neuroanatomical evidence. Various neurophysiological data collection methodologies are used in *C. elegans* research, including calcium imaging, optical and electrophysiological recordings, and laser ablation [13–15]. However, no methodology achieves a high-throughput, systematic, and comprehensive accumulation of neurophysiological and behavioral data on the function of specific parts of the neural network.

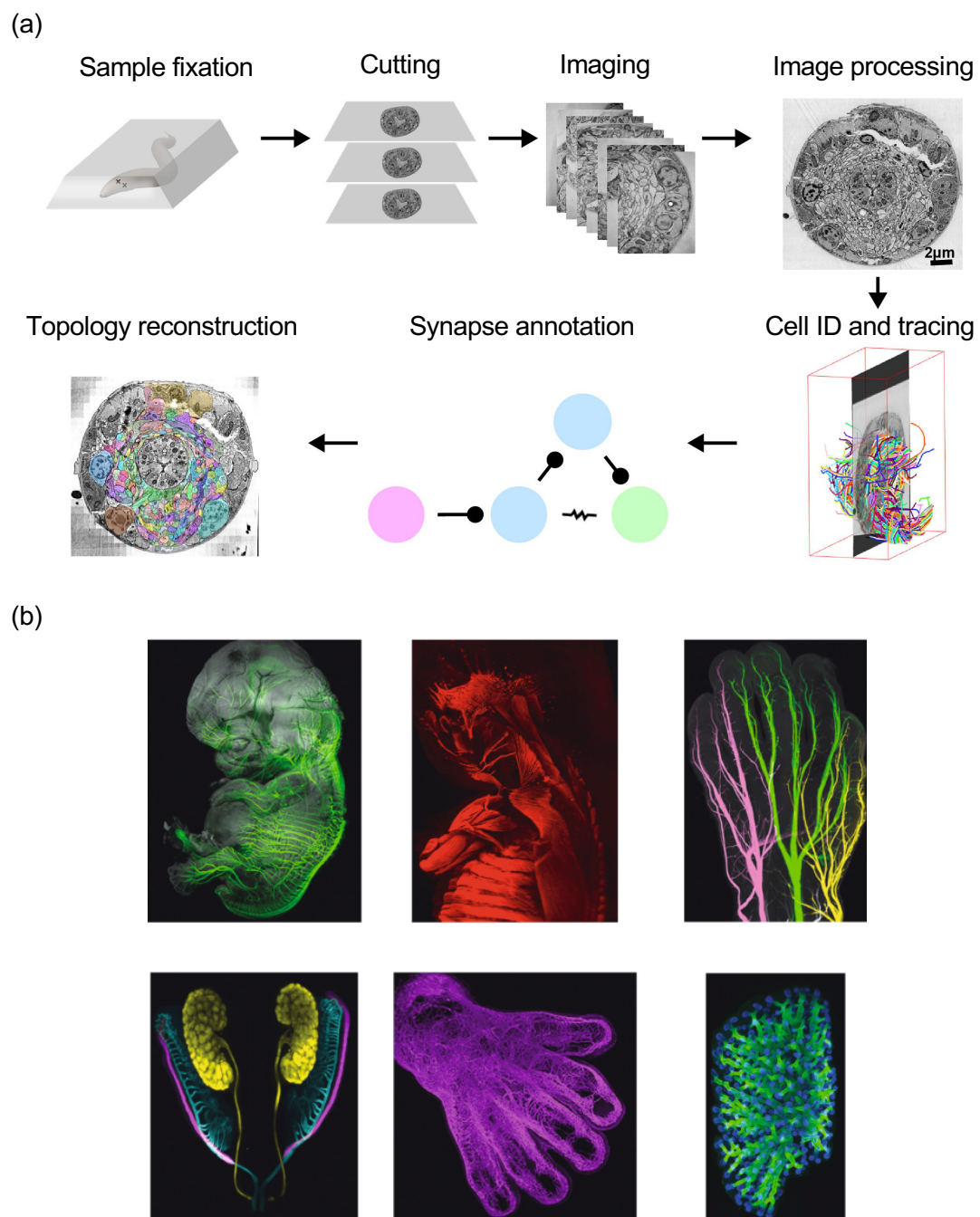


Figure 3. Technologies for descriptive cellomics approaches. **(a)** Serial electron microscopy. The figure is modified from [16]. **(b)** Tissue clearing. The figure is from [17].

Sparse labeling

Sparse labeling is a genetic technique used to label only a small number of cells in an entire population. Sparse labeling affects many research fields but is especially important in neuroscience. This is because the nervous system contains a vast number of neurons with unique morphologies, and the brain contains a huge number of neuron cells [18, 19]. In addition, the brain is densely populated with neurons and a mixture of dendrites and axonal projections, making it challenging to visualize clear morphology. In particular, paradigms that examine the properties of probabilistically selected subsets of cells of the same type are very useful because the single-cell analysis can elucidate the functional logic of neural circuits. Against this background, there is a strong need for probabilistic analysis of gene expression in small populations of cells.

Various methodologies have been developed to achieve sparse labeling. One method is to screen transgenic lines with diverse gene expressions and then use animal lines with the desired expression pattern [20–23]. Other methods are to use site-specific recombinase (SSR). SSR-based methods can control sparseness levels in several ways. First, sparse levels can be controlled by appropriate tamoxifen doses in a CreER-*lox*-mediated recombination system [24–30]. Second, low-titer viral injections into Cre driver lines can be performed [31, 32]. However, these methods require very sophisticated experimental techniques and time-consuming titration of chemical and genetic guidance conditions to limit the spatial and temporal expression of recombinase. This creates a significant problem in that it is difficult to determine the sparseness level a priori with reproducibility.

Methodologies that can control sparseness at a predictive level with high reproducibility would be beneficial. Under these circumstances, several methods have been developed in recent years. First, the MORF (mononucleotide repeat frameshift) method is a Cre expression-dependent sparse cell labeling method based on mononucleotide repeat frameshift as a stochastic translation switch. MORF can control sparseness levels with high reproducibility [33, 34]. The labeling rate is approximately 1 %–5 %, depending on the cell type of interest and the Cre line used. Second, mosaic analysis with repressive cell markers (MARCM) and mosaic analysis with dual markers (MADM) transgenic approaches have been established to sparsely label cells based on Cre-induced interchromosomal recombination that occurs during mitosis [35–37]. The labeling rate is approximate ~1 %–5 %, depending on the type of cells targeted and the Cre line used. These methods can label effectors with high reproducibility but cannot tune sparseness to desired levels because labeling rates depend on cell type and Cre line. In addition, MADM and MARCM can only be used in cells undergoing mitosis. Third, SPARC (sparse predictive activity through recombinase competition) is a method that utilizes two attP and attB target sequences that compete with PhiC31 recombinase [38]. SPARC regulates sparseness levels using three types of

progressively truncated attP sequences. SPARC-D (Dense) labeled 48 %-51 % of cells, SPARC-I (Intermediate) 17 %-22 %, and SPARC-S (Sparse) 3 %-7 %. Finally, there is stochastic gene activation with genetically regulated sparseness (STARS) derived from Brainbow [39–41]. The Brainbow system is a method for probabilistically labeling cells using two mutually exclusive *lox* sequences. STARS regulates the sparseness level of effectors by lengthening the spacer DNA sequences between *lox2272* sequences, thereby controlling the efficiency of removal by Cre. STARS transgenes with spacers of various lengths can regulate sparseness levels from 5 % to 50 % of the cell population. These methods have resulted in highly reproducible sparse labeling. However, controlling the sparsity rate to the desired level takes work. For example, SPARC can only adjust the sparseness level in three steps. STARS requires very long spacer DNA (e.g., 10 kb or more) to achieve low stochastic labeling rates, hindering the construction of transformants.

Identification of neural function at molecular level

Neurons are classified into several subclasses, each with different morphology, expression patterns, connectivity, and function [42]. For instance, two neurons called feeding neurons control sugar-induced feeding behavior in *Drosophila* feeding, and AFD (amphid finger cell D) neurons are known to control body temperature in *C. elegans* (**Figure 4**) [43–46]. Understanding the molecular mechanisms of these neuronal controls requires analysis of the molecular details of each neuronal subclass [47, 48].

The evaluation of expression patterns in neurons from mice, *Drosophila*, and *C. elegans* using single-cell RNA-seq analysis has led to the identification of new neuronal subclasses [49–52]. However, the amount of mRNA transcription does not correlate with the amount of protein [53, 54]. Post-translational modifications such as phosphorylation and ubiquitination are also crucial in regulating cellular functions [55, 56]. Therefore, it is necessary to analyze the abundance and expression patterns of proteins in neurons of the target subclass to elucidate the molecular mechanisms of neuronal function. Fluorescent reporters are commonly used to evaluate protein expression, but it is challenging to observe highly multiplexed individual protein expression patterns with this method [57].

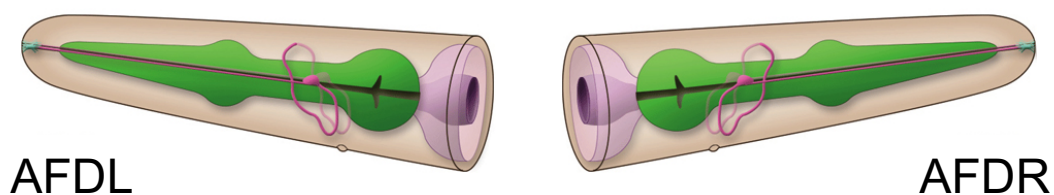


Figure 4. Schematic of AFD neurons in *C. elegans*. AFD neurons play an important role in the thermoregulation of *C. elegans* [44–47] (the picture is from wormatlas.org).

Target specific proteome

Comprehensive protein expression analysis can identify unique proteins and molecular mechanisms of neural function in target subclasses [47, 48]. Mass spectrometry-based proteome analysis generated protein expression patterns for several subclasses [43, 46, 58]. In these studies, specific neuronal subclasses or all neurons were isolated via *in vitro* differentiation, laser dissection, flow cytometric sorting, and antibody-binding microbeads [59–62]. However, these methods have several limitations. For example, neurons differentiated *in vitro* do not form neuronal networks as in the brain. Laser dissection is difficult to use to separate branched or interconnected cells. Extended sample preparation durations are needed to separate cells for procedures like flow cytometric sorting and antibody-bound microbeads. In addition, artifacts during cell isolation cannot be ignored in flow cytometry sorting.

In vivo proteome analysis with cell-selective metabolic labeling

In vivo labeling of cell-specific proteins has been developed to recover proteins from specific cells without the need for cell isolation. *In vivo* cell-selective metabolic labeling of the proteome allows comprehensive proteome analysis of targeted cells [63–66]. These methods can be divided into two groups. One is protein biotinylation, in which target cells express artificial ascorbate peroxidase or artificial *E. coli* biotin ligase [63, 67]. The first method developed was BioID, which uses the *E. coli* biotin ligase BirA with a one-residue mutation [67]. This mutation reduces substrate specificity, and proteins near BirA are labeled with biotin. The protein in the target cells can be labeled with biotin by expressing the mutant BirA only in the target cells with a cell type-specific promoter. Biotin-labeled proteins can be recovered with avidin. However, the reaction efficiency of the single residue mutant BirA was low, and a long reaction time was required for labeling. Recently, a highly active BirA enzyme (TurboID) was obtained by introducing 15 mutations into BirA through directed evolution [66]. This enzyme enables highly efficient labeling of proteins in cells in a reaction of about 3 h. The other is protein azidation, in which mutant aminoacyl-tRNA synthetases are expressed in target cells. This method is called cell-selective bioorthogonal non-normal amino acid tagging (BONCAT) (**Figure 5**) [64, 68, 69]. BONCAT uses a mutated phenylalanyl-tRNA synthetase (MuPheRS) with a single residue mutation in aminoacyl-tRNA synthetase, which mediates the binding of amino acids to tRNA. This mutation allows the addition of azide-containing unnatural amino acids to tRNA, and the protein translated in the cell is labeled with azide. Biotinylation, as the name implies, involves labeling proteins in target cells with biotin and using biotin-streptavidin interactions to recover biotinylated proteins. However, biotinylated proteins are toxic to target cells [66]. In protein

azidation, newly synthesized proteins are labeled with azide-containing amino acids, and copper-catalyzed azide-alkyne cycloaddition reactions recover these proteins. Azide-labeled proteins are known to have low toxicity to living cells [69].

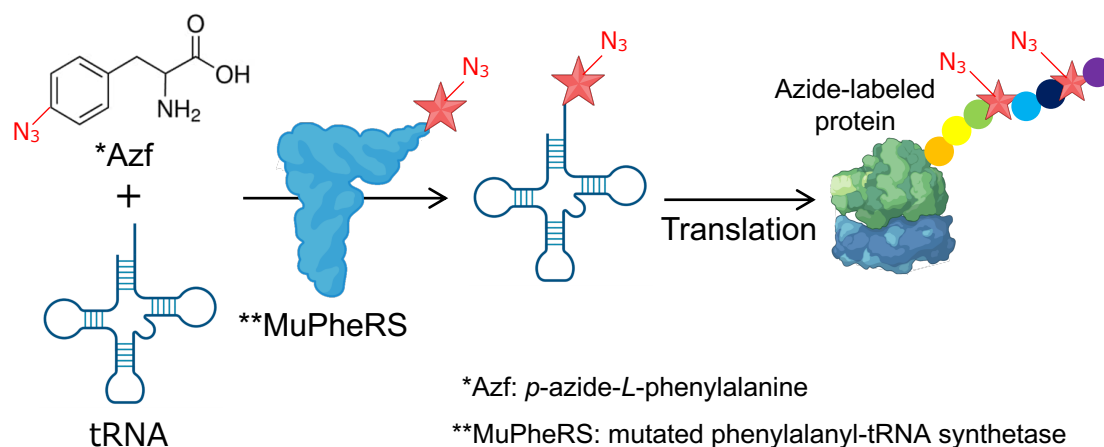


Figure 5. Schematic of cell-selective biorthogonal non-normal amino acid tagging (BONCAT). The figure is made using Biorender (<https://biorender.com/>).

References

1. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry* vol. 8 61–80 Preprint at <https://doi.org/10.1146/annurev-anchem-071114-040325> (2015).
2. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* vol. 17 333–351 Preprint at <https://doi.org/10.1038/nrg.2016.49> (2016).
3. Herculano-Houzel, S., Mota, B. & Lent, R. Cellular scaling rules for rodent brains. www.pnas.org/cgi/doi/10.1073/pnas.0604911103 (2006).
4. Herculano-Houzel, S. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences of the United States of America* vol. 109 10661–10668 Preprint at <https://doi.org/10.1073/pnas.1201895109> (2012).
5. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **314**, 1–340 (1986).
6. Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput Biol* **7**, (2011).
7. Kimata, T., Sasakura, H., Ohnishi, N., Nishio, N. & Mori, I. Thermotaxis of *C. elegans* as a model for temperature perception, neural information processing and neural plasticity. *Worm* **1**, 31–41 (2012).
8. Margie, O., Palmer, C. & Chin-Sang, I. *C. elegans* chemotaxis assay. *Journal of Visualized Experiments* (2013) doi:10.3791/50069.
9. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol* **2**, (2004).
10. Tainaka, K., Kuno, A., Kubota, S. I., Murakami, T. & Ueda, H. R. Chemical Principles in Tissue Clearing and Staining Protocols for Whole-Body Cell Profiling. *Annual Review of Cell and Developmental Biology* vol. 32 713–741 Preprint at <https://doi.org/10.1146/annurev-cellbio-111315-125001> (2016).
11. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
12. Förster, D., Dal Maschio, M., Laurell, E. & Baier, H. An optogenetic toolbox for unbiased discovery of functionally connected cells in neural circuits. *Nat Commun* **8**, (2017).
13. Nguyen, J. P. *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **113**, E1074–E1081 (2016).

14. Tye, K. M. & Deisseroth, K. Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nature reviews. Neuroscience* **13**, 251–266 (2012).
15. Fang-Yen, C., Gabel, C. v., Samuel, A. D. T., Bargmann, C. I. & Avery, L. Laser Microsurgery in *Caenorhabditis elegans*. in *Methods in Cell Biology* vol. 107 177–206 (Academic Press Inc., 2012).
16. Mulcahy, B. *et al.* A pipeline for volume electron microscopy of the *caenorhabditis elegans* nervous system. *Front Neural Circuits* **12**, (2018).
17. Ueda, H. R. *et al.* Tissue clearing and its applications in neuroscience. *Nature Reviews Neuroscience* vol. 21 61–79 Preprint at <https://doi.org/10.1038/s41583-019-0250-1> (2020).
18. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335–346 (2016).
19. Swanson, L. W. Quest for the basic plan of nervous system circuitry. *Brain Research Reviews* vol. 55 356–372 Preprint at <https://doi.org/10.1016/j.brainresrev.2006.12.006> (2007).
20. Xiao, T., Roeser, T., Staub, W. & Baier, H. A GFP-based genetic screen reveals mutations that disrupt the architecture of the zebrafish retinotectal projection. *Development* **132**, 2955–2967 (2005).
21. Feng G, Mellor RH, Bernstein M, Keller-Peck C, Nguyen QT, Wallace M, et al. Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron*. 2000;28: 41–51. pmid:11086982
22. Young, P. *et al.* Single-neuron labeling with inducible Cre-mediated knockout in transgenic mice. *Nat Neurosci* **11**, 721–728 (2008).
23. Chong, S. Y. C. *et al.* Neurite outgrowth inhibitor Nogo-A establishes spatial segregation and extent of oligodendrocyte myelination. *Proc Natl Acad Sci U S A* **109**, 1299–1304 (2012).
24. Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: A tool for temporally regulated gene activation/inactivation in the mouse. *Dev Biol* **244**, 305–318 (2002).
25. Badea, T. C., Wang, Y. & Nathans, J. *A Noninvasive Genetic/Pharmacologic Strategy for Visualizing Cell Morphology and Clonal Relationships in the Mouse*. (2003).
26. Lu, X. H. & Yang, X. W. Genetically-directed Sparse Neuronal Labeling in BAC Transgenic Mice through Mononucleotide Repeat Frameshift. *Sci Rep* **7**, (2017).
27. Ellisor, D. & Zervas, M. Tamoxifen dose response and conditional cell marking: Is there control? *Molecular and Cellular Neuroscience* **45**, 132–138 (2010).
28. Gómez-Casati, M. E., Murtie, J., Taylor, B. & Corfas, G. Cell-specific inducible gene recombination in postnatal inner ear supporting cells and glia. *JARO - Journal of the Association for Research in Otolaryngology* **11**, 19–26 (2010).

29. Wu, H., Williams, J. & Nathans, J. Morphologic diversity of cutaneous sensory afferents revealed by genetically directed sparse labeling. *Elife* **2012**, (2012).
30. Badea, T. C. *et al.* New mouse lines for the analysis of neuronal morphology using CreER(T)/*loxP*-directed sparse labeling. *PLoS One* **4**, (2009).
31. Zingg, B., Peng, B., Huang, J., Tao, H. W. & Zhang, L. I. Synaptic specificity and application of anterograde transsynaptic AAV for probing neural circuitry. *Journal of Neuroscience* **40**, 3250–3267 (2020).
32. Lin, R. *et al.* Cell-type-specific and projection-specific brain-wide reconstruction of single neurons. *Nat Methods* **15**, 1033–1036 (2018).
33. Rotolo, T., Smallwood, P. M., Williams, J. & Nathans, J. Genetically-directed, cell type-specific sparse labeling for the analysis of neuronal morphology. *PLoS One* **3**, (2008).
34. Veldman, M. B. *et al.* Brainwide Genetic Sparse Cell Labeling to Illuminate the Morphology of Neurons and Glia with Cre-Dependent MORF Mice. *Neuron* **108**, 111-127.e6 (2020).
35. Luo, L. Fly MARCM and mouse MADM: Genetic methods of labeling and manipulating single neurons. *Brain Research Reviews* vol. 55 220–227 Preprint at <https://doi.org/10.1016/j.brainresrev.2007.01.012> (2007).
36. Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L., Mosaic Analysis with a Repressible Neurotechnique Cell Marker for Studies of Gene Function in Neuronal Morphogenesis.
37. Lee T, Luo L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron*. 1999;22: 451–461. pmid:10197526
38. Isaacman-Beck, J. *et al.* SPARC enables genetic manipulation of precise proportions of cells. *Nat Neurosci* **23**, 1168–1175 (2020).
39. Wang, S. Z., Liu, B. H., Tao, H. W., Xia, K. & Zhang, L. I. A genetic strategy for stochastic gene activation with regulated sparseness (STARS). *PLoS One* **4**, (2009).
40. Ibrahim, L. A. *et al.* Sparse Labeling and Neural Tracing in Brain Circuits by STARS Strategy: Revealing Morphological Development of Type II Spiral Ganglion Neurons. *Cerebral Cortex* **31**, 2759–2772 (2021).
41. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
42. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nature Reviews Neuroscience* vol. 8 427–437 Preprint at <https://doi.org/10.1038/nrn2151> (2007).
43. Plum, S. *et al.* Combined enrichment of neuromelanin granules and synaptosomes from human substantia nigra pars compacta tissue for proteomic analysis. *J Proteomics* **94**, 202–206 (2013).

44. Komatsu, H., Mori, I. & Rhee, J.-S. *Mutations in a Cyclic Nucleotide-Gated Channel Lead to Abnormal Thermosensation and Chemosensation in C. elegans.* *Neuron* vol. 17 (1996).
45. Hobert, O. *et al.* *Regulation of Interneuron Function in the C. elegans Thermoregulatory Pathway by the ttx-3 LIM Homeobox Gene.* *Neuron* vol. 19 (1997).
46. Kobayashi, K. *et al.* *Single-Cell Memory Regulates a Neural Circuit for Sensory Behavior.* *Cell Rep* **14**, 11–21 (2016).
47. Schreiner, D., Savas, J. N., Herzog, E., Brose, N. & de Wit, J. *Synapse biology in the ‘circuit-age’—paths toward molecular connectomics.* *Current Opinion in Neurobiology* vol. 42 102–110 Preprint at <https://doi.org/10.1016/j.conb.2016.12.004> (2017).
48. Dieterich, D. C. & Kreutz, M. R. *Proteomics of the synapse - A quantitative approach to neuronal plasticity.* *Molecular and Cellular Proteomics* **15**, 368–381 (2016).
49. Zeisel, A. *et al.* *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* *Science* **347**, 1138–1142 (2015).
50. Usoskin, D. *et al.* *Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing.* *Nat Neurosci* **18**, 145–153 (2015).
51. Li, H. *et al.* *Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing.* *Cell* **171**, 1206.e22–1220.e22 (2017).
52. Cao, J. *et al.* *Comprehensive single-cell transcriptional profiling of a multicellular organism.* <https://www.science.org>.
53. Liu, Y., Beyer, A. & Aebersold, R. *On the Dependency of Cellular Protein Levels on mRNA Abundance.* *Cell* vol. 165 535–550 Preprint at <https://doi.org/10.1016/j.cell.2016.03.014> (2016).
54. Gygi, S. P. *et al.* *Correlation between Protein and mRNA Abundance in Yeast.* *MOLECULAR AND CELLULAR BIOLOGY* vol. 19 <https://journals.asm.org/journal/mcb> (1999).
55. Brüning, F. *et al.* *Sleep-wake cycles drive daily dynamics of synaptic phosphorylation.* *Science (1979)* **366**, (2019).
56. Szabó, Á. *et al.* *Ubiquitylation Dynamics of the Clock Cell Proteome and TIMELESS during a Circadian Cycle.* *Cell Rep* **23**, 2273–2282 (2018).
57. Dean, K. M. & Palmer, A. E. *Advances in fluorescence labeling strategies for dynamic cellular imaging.* *Nature Chemical Biology* vol. 10 512–523 Preprint at <https://doi.org/10.1038/nchembio.1556> (2014).
58. Biesemann, C. *et al.* *Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting.* *EMBO Journal* **33**, 157–170 (2014).
59. Molina, M. *et al.* *Enrichment of single neurons and defined brain regions from human brain tissue samples for subsequent proteome analysis.* *J Neural Transm* **122**, 993–1005 (2015).

60. García-Berrocso, T. *et al.* Single cell immuno-laser microdissection coupled to label-free proteomics to reveal the proteotypes of human brain cells after ischemia. *Molecular and Cellular Proteomics* **17**, 175–189 (2018).
61. Chai, H. *et al.* Neural Circuit-Specialized Astrocytes: Transcriptomic, Proteomic, Morphological, and Functional Evidence. *Neuron* **95**, 531-549.e9 (2017).
62. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci* **18**, 1819–1831 (2015).
63. Rhee, H.-W. *et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
64. Yuet, K. P. *et al.* Cell-specific proteomic analysis in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **112**, 2705–2710 (2015).
65. Stone, S. E., Glenn, W. S., Hamblin, G. D. & Tirrell, D. A. Cell-selective proteomics for biological discovery. *Current Opinion in Chemical Biology* vol. 36 50–57 Preprint at <https://doi.org/10.1016/j.cbpa.2016.12.026> (2017).
66. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880–898 (2018).
67. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **196**, 801–810 (2012).
68. Erdmann, I. *et al.* Cell-selective labelling of proteomes in *Drosophila melanogaster*. *Nat Commun* **6**, (2015).
69. Alvarez-Castelao, B. *et al.* Cell-type-specific metabolic labeling of nascent proteomes in vivo. *Nat Biotechnol* **35**, 1196–1201 (2017).

Chapter I

High-throughput functional annotation of *Caenorhabditis elegans* neural network using a cellomics method

In Chapter I, I established a new methodology called "functional" cellomics, which comprehensively explores the function of single neurons or subsets within a complex neural network in a given behavior. Functional cellomics requires a high-throughput, hypothesis-free, single-cell-resolved, and simple methodology to manipulate neurons and subsets and quantify their effects on behavior. I combined optogenetics, Brainbow technology, and behavioral analysis to achieve this approach. Optogenetics is a technology that allows on-demand optical control of neural activity through the expression of opsin genes [1]. To express opsin in a specific neuron, a specific promoter must be chosen in advance in traditional optogenetics [2]. Therefore, this method is hypothesis-driven, and while it is effective for accurately testing existing hypotheses, it is not suited for establishing entirely new hypotheses (**Figure 1a**).

To solve this problem, a novel optogenetic experimental scheme was devised to achieve hypothesis-free annotation of neural networks (**Figure 1b**). Specifically, I developed a system to determine whether effectors are produced in each neuron probabilistically and to obtain a *C. elegans* library in which effectors are labeled in diverse patterns. To achieve this probabilistic labeling, Brainbow's technology, based on the Cre-*lox* system, was adopted [3]. If Brainbow technology can yield a library of *C. elegans* worms with random labeling patterns of effectors, behavioral experiments under light irradiation using that library could reveal previously unknown relationships between neural networks and behavior in a high-throughput manner. This method is hypothesis-free because it first detects individuals exhibiting abnormal behavior and then identifies the neurons responsible for it. It is expected that the information on structural cell omics and the physiological and behavioral data of individual neurons obtained from functional cellomics can be comprehensively accumulated to construct a behavioral model of neural networks at the whole-brain level.

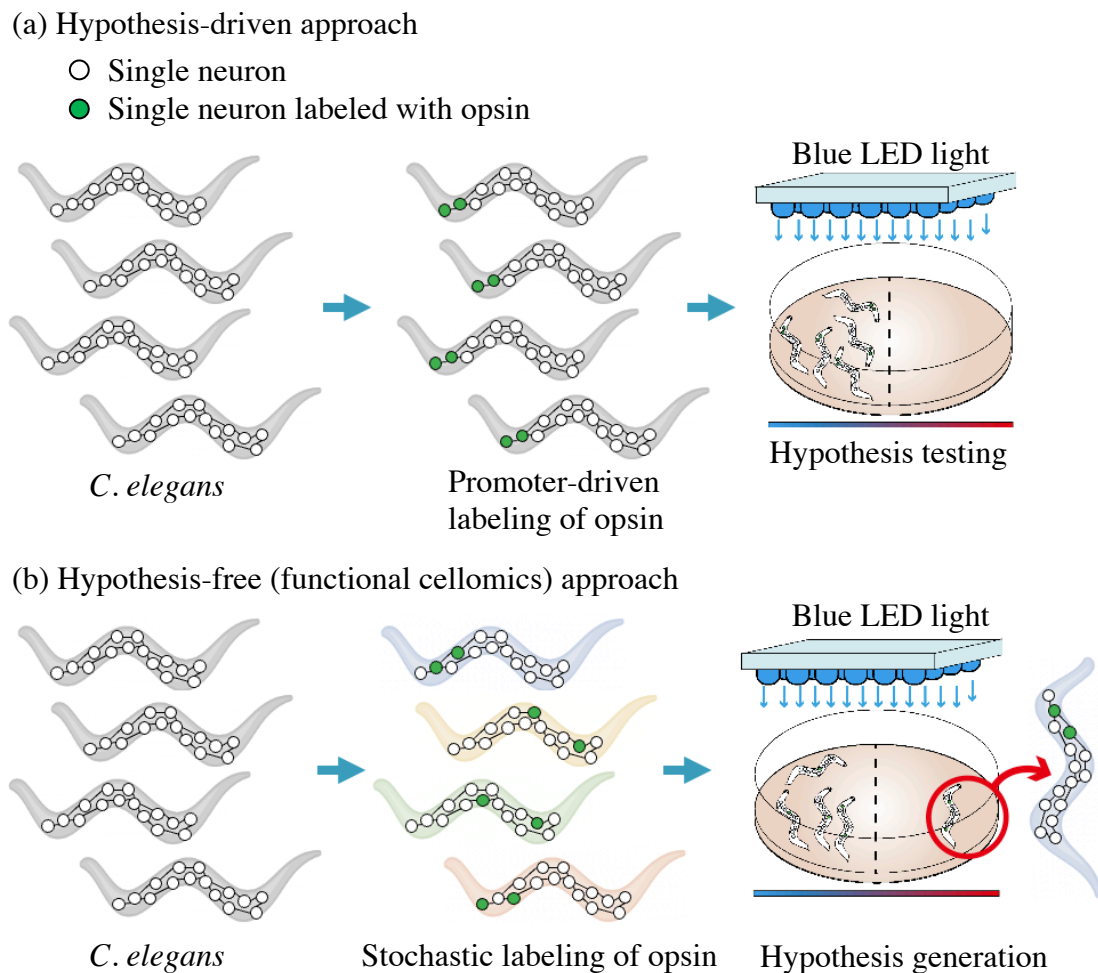


Figure 1. Comparison of two approaches to annotating neural networks: hypothesis-driven and hypothesis-free (functional cellomics) approach. **(a)** A hypothesis-driven approach involves formulating a hypothesis regarding the specific neurons or subset of neurons responsible for a particular behavior. Transgenic *C. elegans* are then engineered to express an effector molecule, such as opsin, in these neurons, and behavioral experiments are conducted to determine the accuracy of the initial hypothesis. **(b)** A hypothesis-free approach (functional cellomics) involves randomly labeling individual neurons with an effector molecule. By performing behavioral experiments on a library of *C. elegans* with various patterns of neuron labeling and identifying individuals that exhibit different behaviors from wild-type, it may be possible to discover new relationships between neural networks and behavior by isolating and examining the neurons that produced the effector.

Materials and methods

Construction of plasmids

For the construction of pCre, SV40NLS-Cre was amplified from pPGK-Cre-bpA from Klaus Rajewsky (Addgene plasmid #11543). The amplified fragment was then inserted into pPD49_78, which was deposited by Andrew Fire (Addgene plasmid #1447).

For the construction of pSTAR, a backbone plasmid containing the *lox* and QF2w sequences were synthesized (Thermo Fisher Scientific, MA, USA) [4]. In addition, a pan-neuronal promoter, F25B3.3p, was cloned from the *C. elegans* genome [5], and mCherry was subcloned from pGH8 deposited by Erik Jorgensen (Addgene plasmid #19359). Into the backbone plasmid, these two fragments were inserted.

For the construction of pQUAS_ChR2_GFP, ChR2(H134R) and GFP(S65C) were amplified from pAAV-Efla-vCreDIO hChR2(H134R)-EYFP from Karl Deisseroth (Addgene plasmid #55643) and L2680 from Andrew Fire (Addgene plasmid #1516). Furthermore, the QF2w responsive promoter sequence QUAS:: Δ pes-10 was constructed (Thermo Fisher Scientific). These three fragments were integrated into pPD49_78 [6].

For the construction of pF25B3.3p_mCherry, F25B3.3p and mCherry were subcloned from pSTAR. In pPD49_78, these fragments were inserted.

Culture Environment

The worms were grown on nematode growth medium (NGM) plates supplemented with *Escherichia coli* OP50. Specifically, OP50 plates were prepared by seeding 6 cm NGM plates with 250 μ L of OP50. The worms were maintained at 20°C, with care taken to minimize temperature shifts. To perform optogenetic experiments, 300 μ L of 500 μ M ATR (Sigma-Aldrich, MO, USA) was added to solid NGM plates containing *E. coli*. The samples were allowed to dry while protected from light by aluminum foil.

Construction of transgenic strains

The nematodes were injected using a stereomicroscope (SZX10; Olympus, Tokyo, Japan) equipped with a Femtojet 4i (5252 000.021; Eppendorf, Hamburg, Germany) and Femtotips II (1-01040; Eppendorf). By co-injecting the four plasmids constructed in this study (50 ng μ L⁻¹ each in water) into the *C. elegans* N2 background, the strain AYK338 (*aykEx338* [*hsp-16.2p::Cre,F25B3.3p::lox2272::mCherry::loxP::lox2272::QF2w::loxP*,

QUAS::ChR2::GFP, F25B3.3p::mCherry) was generated. Three mCherry expressing lines were obtained from 20 nematodes with an N2 background. Previous research showed that illuminating the N2 background with blue light did not affect egg-laying behavior [7]. For negative control experiments, I generated the strain AYK339 (*aykEx339 [F25B3.3p::lox2272::mCherry::loxP::lox2272::QF2w::loxP, QUAS::ChR2::GFP, F25B3.3p::mCherry]*, 50 ng μL^{-1} each in water) and AYK340 (*aykEx340 [hsp-16.2p::Cre, QUAS::ChR2::GFP, F25B3.3p::mCherry]*, 50 ng μL^{-1} each in water) by microinjection into the N2 background.

Inducing Cre recombinase to label ChR2 stochastically

On NGM plates with or without ATR, transgenic worms were plated. Worms were incubated at 37°C for 30 min to induce Cre recombinase by heat shock and then transferred to an incubator at 20°C. Worms were examined by egg-laying assay and fluorescence microscopy 12 h after heat shock.

Confocal laser scanning microscopy

A 5 % agarose pad (041149-05; Nacalai Tesque, Kyoto, Japan) was prepared and coated with 5 μL of 50 mM sodium azide (830011; Nacalai Tesque). *C. elegans* worms were collected and placed on sodium azide-coated agarose pads with a cover glass gently placed over them. Fluorescence was observed using confocal laser scanning microscopy (LSM 700; Carl Zeiss, Oberkochen, Germany). The 488 nm and 561 nm lasers were used to observe the fluorescence of GFP and mCherry, respectively. Zen Lite, Imaris, or ImageJ software was used to process the acquired images [8].

Egg-laying assay

Stereomicroscopes (SZX10; Olympus) were used to observe the worms. The stereomicroscope was equipped with a camera (HAS-L1; DITECT, Tokyo, Japan). To prevent ChR2-GFP from being activated in the worms during observation, the halogen lamp (410849; PHILIPS, Amsterdam, the Netherlands) of the SZX10 was equipped with an optical filter (Asahi Spectra, Tokyo, Japan) that blocks wavelengths below 600 nm. For ChR2-GFP activation, the worms were illuminated with blue light (LDL2-98X30BL2; CCS, Kyoto, Japan). The blue light was supplied by PD3-5024-4-PI (CCS). To prevent the blue light from the LED from being detected by the camera, the object lens was equipped with an optical filter (Asahi Spectra). This

filter blocks wavelengths below 570 nm. Each worm was transferred to a 6 cm agar NGM plate without *E. coli* and filmed for 30 sec for an egg-laying assay to examine light-dependent behavioral modulation. During this recording, the blue light was turned on and off at 5 sec intervals. Individuals that showed egg-laying behavior within 30 sec were defined as egg-laying individuals. Regardless of the presence or absence of *E. coli*, activation of HSNs induces egg-laying behavior [9].


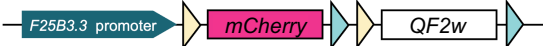


Results

Strategy design of functional cellomics

Using stochastic labeling with an effector gene that causes cell-autonomous activity allows for a high-throughput and hypothesis-free method of functional cellomics at single-cell resolution. Brainbow technologies refer to systems that can randomly determine the expression of a specific gene in a particular cell using the Cre-*lox* recombination system. In Brainbow technologies, multiple *lox* variants (such as *loxP* and *lox2272* sequences) are alternately placed downstream of one promoter, with two genes interspersed between these *lox* sequences (**Figure 2a**). If Cre recombinase is applied to this sequence, excision occurs only between *loxP* sequences or *lox2272* sequences. As a result, it is possible to determine which of these two genes is expressed in a Cre-dependent manner.

I designed four plasmids to carry out functional cellomics (**Figure 2a**). The plasmid pCre expresses Cre recombinase in reaction to heat stress. The plasmid pSTAR contains two *lox* sequences, mCherry, and a transcription factor (QF^{2w}) downstream of a pan-neuronal promoter (*F25B3.3p*). The plasmid pQUAS_ChR2_GFP expresses ChR2_GFP in a QF2w-dependent way. This study used channelrhodopsin-2 fused with GFP (ChR2::GFP) as the effector gene. Since the constructs for producing a transcription factor or an effector are modularized, it is easy to utilize a variety of effectors as well as opsin. I also constructed pF25B3.3p_mCherry, which continues to produce mCherry after Cre recombination.

When all of these plasmids are introduced into *C. elegans*, all neurons initially produce only mCherry. Upon applying a heat shock to induce Cre recombinase, QF2w is produced if excision occurs between *lox2272* sequences, and mCherry production continues from pF25B3.3p_mCherry even after Cre recombination. In QF2w-producing neurons, ChR2-GFP is produced as an effector and can be activated on demand through light illumination. The fusion of GFP to ChR2 makes it easy to identify which neurons are producing opsin after performing a behavioral experiment.

- (a) 1. pCre : Heat-shock-dependent conditional expression of Cre
- 
2. pSTAR : Brainbow-based construct for stochastic expression of a transcription factor, *QF2w*
- 
3. pQUAS_ChR2_GFP : QF2w-dependent conditional expression of an effector, ChR2::GFP
- 
4. pF25B3.3p_mCherry : Constitutive pan-neuronal expression of mCherry
- 

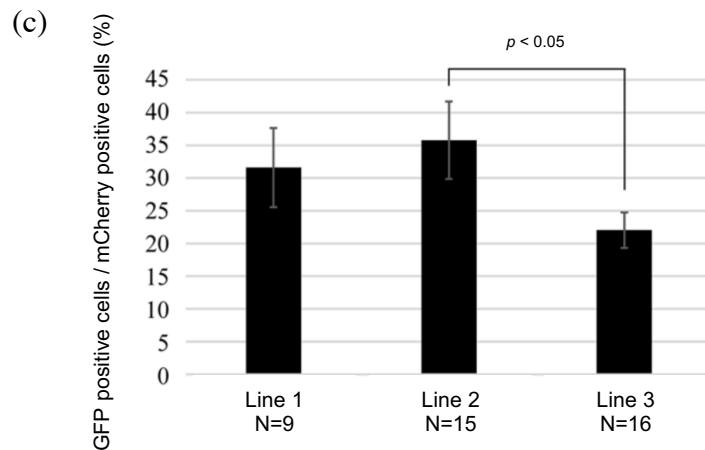
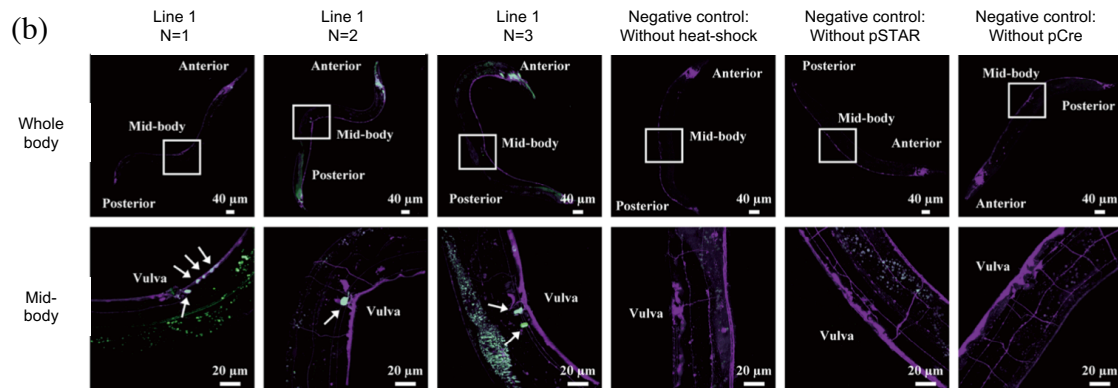


Figure 2. The application of Brainbow technologies for stochastic labeling of neurons. **(a)** The Brainbow technology utilizes two constructs, one encoding Cre recombinase under the control of a heat-shock-dependent promoter (pCre) and another encoding the transcription factor QF2w, two *loxP* sequences (*lox2272* and *loxP* sequences), and the fluorescent protein mCherry downstream of a pan-neuronal promoter (pSTAR). In the absence of Cre recombinase, mCherry is produced in all neurons by pSTAR and pF25B3.3p_mCherry. Upon Cre-mediated excision between the

lox2272 sequences, QF2w is produced from pQUAS_ChR2_GFP, and mCherry continues to be expressed from the pan-neuronal promoter, even after Cre-mediated recombination. **(b)** Stochastic labeling of neurons with ChR2-GFP is achieved by applying a brief heat shock to transgenic *C. elegans* carrying the four constructs as extrachromosomal arrays and examining the worms by confocal laser scanning microscopy 12 h later. Negative control experiments include transgenic *C. elegans* without heat shock, transgenic lines without the pSTAR plasmid, and transgenic lines without the pCre plasmid with heat shock. The whole body and mid-body section (framed in white in the whole body image) are examined at magnifications of 10x and 40x, respectively. The fluorescence of ChR2-GFP is shown in green, and the fluorescence of mCherry is shown in magenta, with cells producing both fluorescence in white and indicated by an arrow. The differences in white-dyed cells between individuals demonstrate stochastic labeling. **(c)** The percentage of GFP-positive cells out of all cells. Three transgenic lines harboring all plasmids shown in **Figure 2a** were established, and at least nine individuals from each line were used to quantify the ratio of GFP-positive cells. In each individual, I counted 12–26 fluorescent cells. In this experiment, GFP-positive cells in the mid-body and tail were counted. The data show a mean \pm standard deviation, and the significance of differences between all pairs of results was examined using Tukey's test.

Randomized labeling of neurons at single-cell level

I introduced the four plasmids mentioned above into *C. elegans* and established three lines (Lines 1–3) [hsp-16.2p::Cre, F25B3.3p::*lox2272*::mCherry::*loxP*::*lox2272*::QF2w::*loxP*, QUAS::ChR2::GFP, F25B3.3p:: mCherry]. After this *C. elegans* strain had propagated, I applied a brief heat shock to determine whether ChR2–GFP was labeled stochastically in each *C. elegans* individual. After isolating at least nine individuals, I observed their mid-body sections with low neuron density at a magnification of 40x. The results showed that ChR2-GFP was generated in all heat shock-treated individuals and that the labeling pattern of ChR2-GFP differed from individual to individual (**Figure. 2b**). In **Figure 2b**, three representative micrographs of Line 1 are shown. The average percentage of cells labeled with ChR2-GFP was quantified. Although the production of ChR2-GFP was random in all strains, there were slight differences in labeling rates due to differences in the structure of the extrachromosomal (Ex) array (**Figure 2c**). I compared the probability of ChR2-GFP labeled neurons at the mid-body region and the tail region. I discovered that transgenic *C. elegans* exhibited a GFP labeling probability of 32 ± 12 % (mean \pm standard deviation) in the caudal region and 31 ± 16 % (mean \pm standard deviation) in the mid-body region. In addition, three negative control experiments were performed: without heat shock, without pSTAR plasmid (substrate for Cre), and with pCre plasmid plus heat shock. I established

three lines in each negative control experiment and observed ten animals from each line. As a result, no green fluorescence was observed in the neurons (**Figure 2b**). These findings suggested that the induction of Cre by heat shock was necessary for the induction of ChR2-GFP.

Identifying the neurons which play an essential role in egg-laying behavior

In functional cellomics, the use of stochastic labeling of an effector gene allows for the exploration of the relationships between neural networks and behaviors in a hypothesis-free and comprehensive manner. I chose the egg-laying behavior of *C. elegans* as a model to show the feasibility of functional cellomics. The egg-laying behavior of *C. elegans* is known to be controlled by a relatively simple neural network. Two hermaphrodite-specific neurons (HSNs: HSNR and HSNL) play a central role in egg-laying behavior by directly exciting vulval and ventral C neurons (VC) [10–12]. Additionally, it is known that the activation of the HSNs by ChR2 elicits egg-laying behavior. If the HSNs can be identified by functional cellomics in a high-throughput manner, it shows that this strategy is effective.

I constructed a *C. elegans* library stochastically labeled with ChR2-GFP. The transgenic *C. elegans* individuals were photographed for 30 sec while irradiated with blue light (**Figure 3a**). Among the individuals photographed, 65 % exhibited light-dependent egg-laying behavior, while 35 % did not (**Figure 3b, c**). When a similar experiment was conducted without adding all-trans-retinal (ATR), a cofactor for ChR2, no egg-laying behavior was observed. These results suggest that the egg-laying behavior shown in this experiment is ChR2-dependent and that generating individual nematodes with the desired phenotype is possible via stochastic labeling of an effector.

Using confocal laser scanning microscopy, I examined whether ChR2-GFP was present in HSNs. By isolating egg-laying and non-egg-laying individuals and analyzing the area surrounding the vulva, I found that ChR2-GFP was present in the HSNs of all egg-laying individuals (**Figure 4a**) but not in the non-egg-laying individuals (**Figure 4b**). ChR2-GFP was observed in HSNR but not in HSNL in the representative individual shown in **Figure 4a**. The probability of GFP labeling was $31 \pm 11\%$ (mean \pm standard deviation, $N = 10$) and $29 \pm 14\%$ (mean \pm standard deviation, $N = 10$) in egg-laying and non-egg-laying animals, respectively, similar to the data in **Figure 2**, and there was no significant difference between the two groups as determined by t-test. Quantitative analysis of fluorescence intensity revealed that ChR2-GFP was detected in HSNR but not in other neurons adjacent to HSNs in the egg-laying individual (**Figure 4c**). A previous study found that destroying one HSN through laser ablation had no noticeable effect on egg-laying behavior in nematodes while destroying both HSNs resulted in a marked inhibition of egg-laying behavior [11]. Our result is consistent with this previous study in that egg-laying behavior can be sufficiently induced by activating one HSN. However, the present

results could not rule out the possibility that egg-laying behavior is activated through the Chr2-dependent activation of other neurons. In this study, the high probability of Chr2-GFP labeling made it challenging to accurately identify neurons other than HSNs that may potentially play a role in egg-laying behavior. In the future, functional cellomics should be improved by using nuclear localization signals and/or reducing the probability of Chr2-GFP labeling.

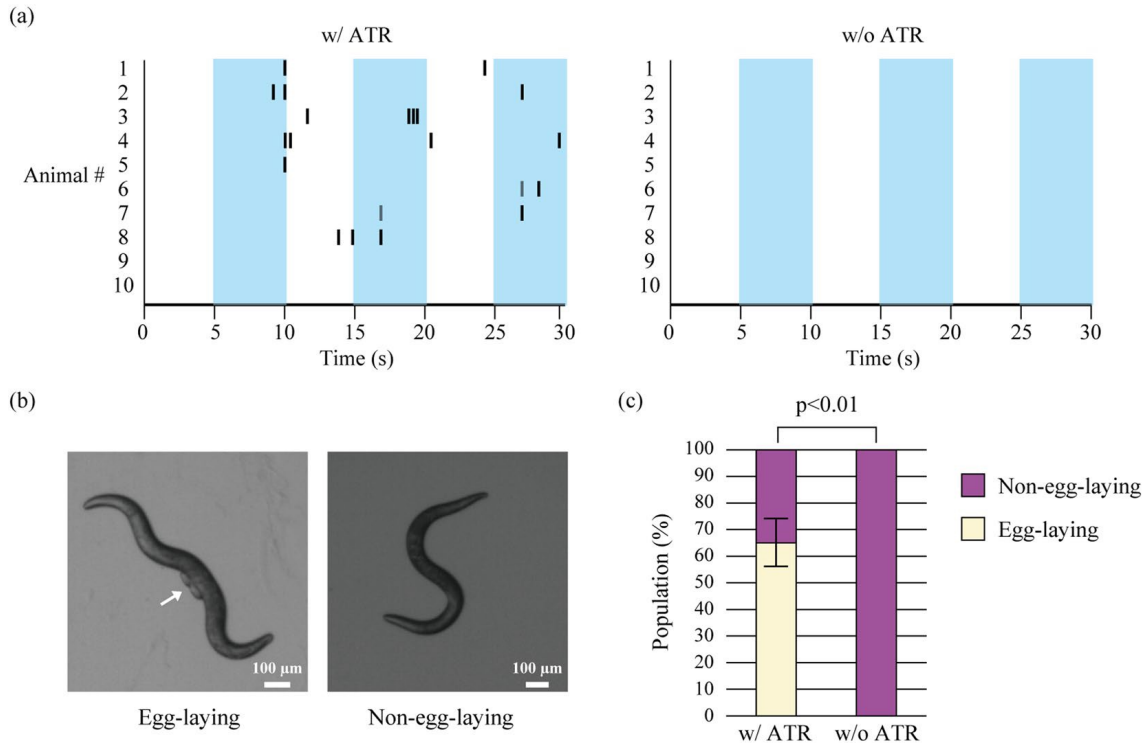


Figure 3. Obtaining individuals who exhibit light-dependent egg-laying behavior. **(a)** A raster graph displaying the results of the behavioral experiment, with black lines indicating egg-laying events. During a 30 sec firing period, blue light was turned on and off at 5 sec intervals. **(b)** Representative images from the behavioral experiment show an individual who exhibits light-dependent egg-laying behavior (left), with eggs indicated by arrows, and an individual without this behavior (right). **(c)** The percentage of individuals displaying light-dependent egg-laying behavior. 65 % exhibited this behavior in the presence of all-trans-retinal (ATR) (N=20), and none exhibited it in the absence of ATR (N=20). Fisher's exact test was used to determine the significance of differences, with error bars representing 95 % confidence intervals. In these tests, at least six individuals from each line were used.

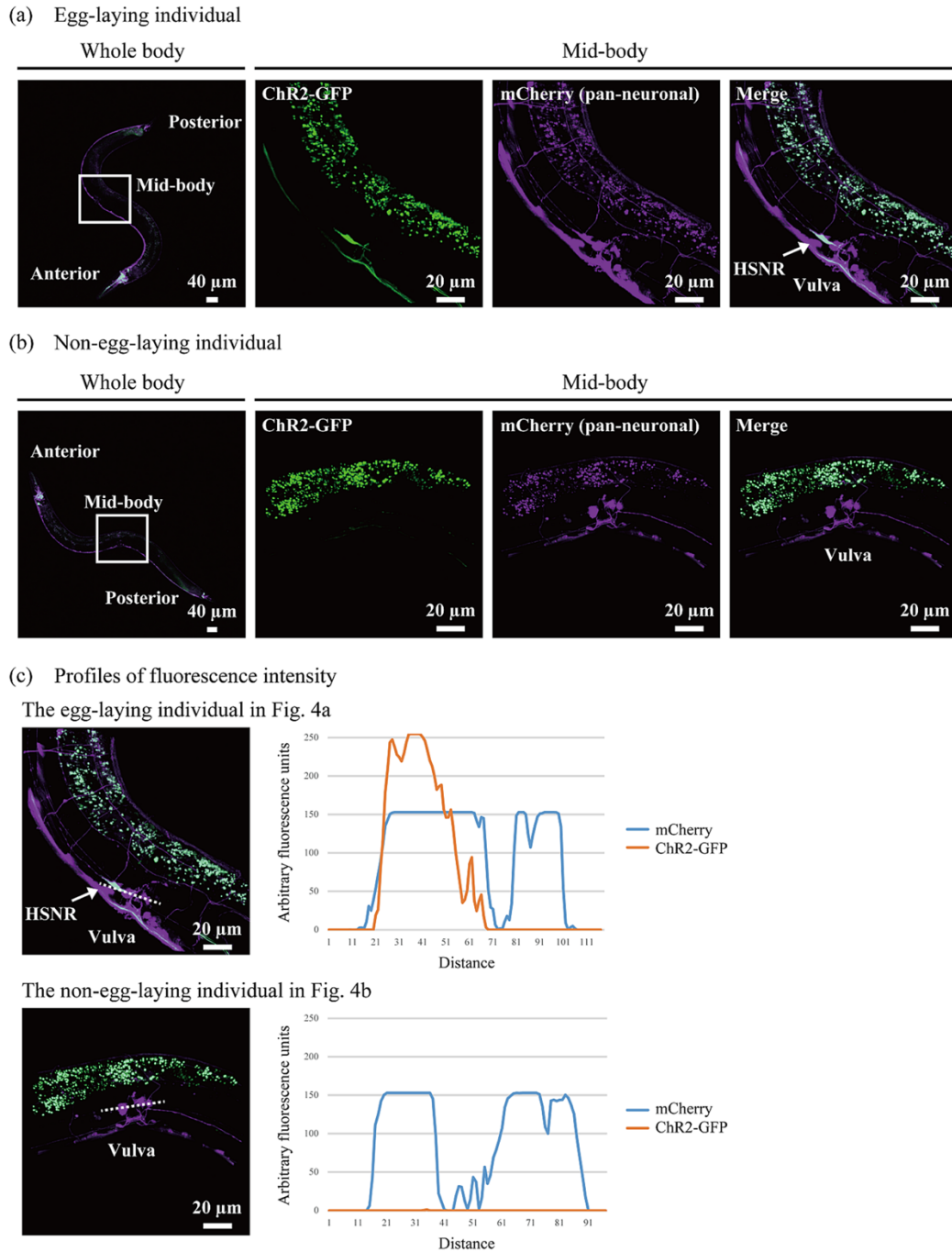


Figure 4. Detection of HSNs. The whole body and mid-body were observed using a 10x and 40x objective, respectively. Two-dimensional (2D) images generated from the maximum-intensity projection of z-stack images acquired using confocal microscopy. **(a)** Fluorescence images of an individual displaying light-dependent egg-laying behavior, with ChR2-GFP fluorescence in green and mCherry fluorescence in magenta. Cells producing both fluorescence are shown in white, and expression of ChR2-GFP in HSNR neurons is observed. **(b)** Fluorescence image of an individual not exhibiting light-dependent egg-laying behavior, with no production of ChR2-GFP observed

in neurons around the vulva. (c) Fluorescence intensity profiles. Dotted fluorescence profiles were quantified in ImageJ.

Discussion

Although numerous methodologies have been developed to study the properties of neural networks, they still need to fully meet the requirements of the conceptual framework of functional cellomics. [13–16]. Functional cellomics described in this study is the first approach that combines all necessary qualities for achieving individual-level cellomics: high-throughput, hypothesis-free, single-cell resolution, and simplicity. By employing functional cellomics on the egg-laying behavior of *C. elegans*, I successfully demonstrated the feasibility of this approach.

Functional cellomics exhibits several advantages over existing methodologies regarding throughput, resolution, and expandability. Firstly, it does not require special equipment and can be easily implemented in any laboratory. Secondly, as it has no limitations on feasible labeling patterns, it is entirely hypothesis-free, enabling easy labeling at single-cell resolution, even for bilaterally symmetrical neuron pairs with highly similar gene expression patterns. Thirdly, as one transgenic *C. elegans* individual can propagate many others with different labeling patterns, it is both simple and high-throughput. Fourthly, it can manipulate neural networks in various ways. In addition to opsin, which was used in this study, any effector can be utilized as long as it causes either loss or gain of function in neurons, allowing for various interventions, such as cell killing, suppression, activation, and gene expression control [17–22]. A recent study suggested that contradictory results may be obtained depending on the mode of intervention, highlighting the need to compare results from various intervention modalities (activation, inhibition, killing, etc.). This approach would be even more helpful if it could identify the neurons required for specific behavior. Fifthly, numerous experimental designs are possible using the *C. elegans* promoter and bipartite gene expression systems like QF2w and Gal4 [23]. For example, instead of a pan-neuronal promoter, a more "focused" functional cellomics can be performed using a promoter specific to a subset of neurons. Additionally, by simultaneously employing QF2w and Gal4, one can stochastically label multiple effectors.

Although functional cellomics holds promise, there are three points of concern to address. One is the reliability of the heat-shock promoter used in this study. While previous studies have shown that the promoter effectively induces protein production in the nervous system, it has yet to be confirmed that it works equally well in all neuronal cells. To solve this problem, it will be crucial to determine the probability of ChR2::GFP labeling in each cell type and to use a wide range of cell-type-specific pSTAR plasmids. The second point is the high probability of effector labeling, making it challenging to calculate the labeling rate accurately. In order to conduct a well-

designed experiment, it is critical to exert strict control over the probability of effector labeling, similar to the predetermined mutation rate in forward genetics. **Figure 3b** shows that a large percentage (65 %) of *C. elegans* displayed egg-laying behavior, which may be due to the high copy number of the pSTAR plasmid, increasing the likelihood of excision of the *lox2272* sequence by Cre and, therefore, a large proportion of neurons labeled by opsin. Suppose effectors label an excessive number of neurons. In that case, this results in a high level of neuronal activation, and it is challenging to identify the neurons that are responsible for the target behavior. One potential solution is concurrently using single-copy integration and *lox* variants to control the probability of labeling only a desired number of the 302 neurons in a *C. elegans* hermaphrodite. The third point is how to ensure the reproducibility of the results obtained. Even if functional cellomics suggests that a specific labeling pattern may play a role in the target behavior, it still needs to be confirmed using other methods. To reproduce the same labeling pattern, methodologies that can evoke gene expression in desired cells, such as using a pulsed infrared laser or multi-step optogenetics, may be applied to relatively quickly verify the results [15, 16]. The intersectional Cre-*lox* strategy and multiple-feature Boolean logic may also help reproduce the labeling pattern [24, 25].

In conclusion, by employing Brainbow technologies to randomize the labeling patterns of effector genes, I have demonstrated, for the first time, the feasibility of identifying neurons involved in the target behavior. While further improvements and additional data are needed to establish the utility of our approach for various behaviors, the results support the basic concept of functional cellomics, which allows for the functional annotation of the neural networks of *C. elegans* in a high-throughput, hypothesis-free, single-cell-resolution, and simple manner. *C. elegans* is particularly well-suited for functional cellomics due to its already-mapped connectome information. By superimposing the connectome information with the results of extensive intervention experiments using functional cellomics, I can construct neuroanatomically grounded models of behavior that can explain how complex neuronal networks perform computation.

References

1. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* 8, 1263–1268 (2005).
2. Tye, K. M. & Deisseroth, K. Optogenetic investigation of neural circuits underlying brain disease in animal models. *Nature Reviews Neuroscience* vol. 13 251–266 Preprint at <https://doi.org/10.1038/nrn3171> (2012).
3. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450, 56–62 (2007).
4. Riabinina, O. et al. Improved and expanded Q-system reagents for genetic manipulations. *Nat Methods* 12, 219–222 (2015).
5. Frøkjær-Jensen, C. et al. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 40, 1375–1383 (2008).
6. Wei, X., Potter, C. J., Luo, L. & Shen, K. Controlling gene expression with the Q repressible binary expression system in *Caenorhabditis elegans*. *Nat Methods* 9, 391–395 (2012).
7. Emtage, L. et al. IRK-1 potassium channels mediate peptidergic inhibition of *Caenorhabditis elegans* serotonin neurons via a Go signaling pathway. *Journal of Neuroscience* 32, 16285–16295 (2012).
8. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* vol. 9 671–675 Preprint at <https://doi.org/10.1038/nmeth.2089> (2012).
9. Leifer, A. M., Fang-Yen, C., Gershow, M., Alkema, M. J. & Samuel, A. D. T. Optogenetic manipulation of neural activity in freely moving *Caenorhabditis elegans*. *Nat Methods* 8, 147–152 (2011).
10. Schafer, W. F. Genetics of egg-laying in worms. *Annual Review of Genetics* vol. 40 487–509 Preprint at <https://doi.org/10.1146/annurev.genet.40.110405.090527> (2006).
11. Zhang, M. et al. A Self-Regulating Feed-Forward Circuit Controlling *C. elegans* Egg-Laying Behavior. *Current Biology* 18, 1445–1455 (2008).
12. Collins, K. M. et al. Activity of the *C. elegans* egg-laying behavior circuit is controlled by competing activation and feedback inhibition. *eLife* 5, <https://doi.org/10.7554/eLife.21126> (2016).
13. Fang-Yen, C., Gabel, C. v., Samuel, A. D. T., Bargmann, C. I. & Avery, L. Laser Microsurgery in *Caenorhabditis elegans*. in *Methods in Cell Biology* vol. 107 177–206 (Academic Press Inc., 2012).
14. Churgin, M. A., He, L., Murray, J. I. & Fang-Yen, C. Efficient single-cell transgene induction in *Caenorhabditis elegans* using a pulsed infrared laser. *G3: Genes, Genomes, Genetics* 3,

- 1827–1832 (2013).
15. Itoh, M., Yamamoto, T., Nakajima, Y. & Hatta, K. Multisteped optogenetics connects neurons and behavior. doi:10.1016/j.
 16. Shipley, F. B., Clark, C. M., Alkema, M. J. & Leifer, A. M. Simultaneous optogenetic manipulation and calcium imaging in freely moving *C. elegans*. *Front Neural Circuits* **8**, (2014).
 17. Qi, Y. B., Garren, E. J., Shu, X., Tsien, R. Y. & Jin, Y. Photo-inducible cell ablation in *Caenorhabditis elegans* using the genetically encoded singlet oxygen generating protein miniSOG. *Proc Natl Acad Sci U S A* **109**, 7499–7504 (2012).
 18. Xu, S. & Chisholm, A. D. Highly efficient optogenetic cell ablation in *C. Elegans* using membrane-targeted miniSOG. *Sci Rep* **6**, (2016).
 19. Pokala, N., Liu, Q., Gordus, A. & Bargmann, C. I. Inducible and titratable silencing of *Caenorhabditis elegans* neurons in vivo with histamine-gated chloride channels. *Proc Natl Acad Sci U S A* **111**, 2770–2775 (2014).
 20. Schild, L. C. & Glauser, D. A. Dual color neural activation and behavior control with chrimson and CoChR in *Caenorhabditis elegans*. *Genetics* **200**, 1029–1034 (2015).
 21. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
 22. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
 23. Wang, H. *et al.* CGAL, a temperature-robust GAL4-UAS system for *Caenorhabditis elegans*. *Nat Methods* **14**, 145–148 (2017).
 24. MacOsko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009).
 25. Fenno, L. E. *et al.* Targeting cells with single vectors using multiple-feature Boolean logic. *Nat Methods* **11**, 763–772 (2014).

Chapter II

Cre-*lox* engineering with machine learning to achieve sparse labeling at a desired sparseness level

In Chapter II, we aimed to develop a new methodology that utilizes the widely used Cre-*lox* recombination system to achieve high reproducibility and the ability to achieve sparseness level at the desired level [1–5]. In particular, we paid attention to the Brainbow system. I hypothesized that the expression rate of a gene could be regulated by mutating one of the *lox* sequences (*lox2272* or *loxP*) and reducing the rate of Cre recognition [6–17]. Therefore, to investigate the effect of introducing mutations in the *lox* sequence on recognition of the *lox* sequence by Cre, we performed random mutagenesis on the *lox* sequence and obtained mutants with a reduced recognition rate by Cre. We further hypothesized that a machine learning model using the dataset of *lox* mutants we evaluated would allow us to design mutant *lox2272* sequences that exhibit various labeling rates at a desired level. The *lox* mutants obtained in this study have the potential to label effector genes in a simple and reproducible manner and regulate gene expression in cell populations with the desired sparseness level. The method is advantageous because Cre can apply to a wide range of organisms, including mice, flies, and worms, and can be employed in post-mitotic cells [18–22].

Materials & Methods

Construction of a library of mutant *lox* sequences

PCR was used to generate a library of mutant *lox* sequences. Randomized primers were used for mutagenizing the *lox* sequences (forward primer: 5'-GTGAGCGCGCGTAATAC-3', reverse primer: 5'-AACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACATAACTTCGTATAATGTATGCTATACGAAGTTATGGCATATGGAATGACTTGGCGTTGTTTC-3'. The mutated region is shown in the underbar in the reverse primer.). Primers were commercially prepared by solid phase synthesis (Eurofin, Tokyo, Japan). Mutated bases were generated using biased randomization with an 84.7 % chance of retaining the original sequence (e.g., A means A: 84.7 %, T: 5.1 %, G: 5.1 %, C: 5.1 %). This is because this study was designed to evaluate *lox* mutants with one or two nucleotide substitutions and, to some extent, mutants with three or more nucleotide substitutions. The two nucleotide substitutions are most efficiently obtained when x =

84.7. Using the following formula, I calculated the distribution of the number of base substitutions:

$$\frac{13!}{(y!(13-y)!)} \times \left(\frac{x}{100}\right)^{13-y} \times \left(\frac{100-x}{100}\right)^y$$

In this formula, x represents the retention rate [%], and y represents the number of base substitutions. Competent *E. coli* DH5 α (F⁻, Φ 80*dlacZ* Δ M15, Δ (*lacZYA-argF*)U169, *deoR*, *recA1*, *endA1*, *hsdR17*(rK⁻, mK⁺), *phoA*, *supE44*, λ^- , *thi-1*, *gyrA96*, *relA1*) was used to transform the library. Transformed *E. coli* were grown in Luria-Bertani (LB) medium (1 % [w/v] tryptone, 0.5 % [w/v] yeast extract, and 1 % [w/v] sodium chloride) containing 100 μ g/mL ampicillin. Subsequently, plasmid extraction was performed using the FastGene Minikit (NIPPON Genetics, Tokyo, Japan, FG-90502). I generated yeast that can express Cre when exposed to β -estradiol (pRS403-CreEBD). I used the *S. cerevisiae* strain BY4741 (*MATa*, *his3 Δ 1*, *leu2 Δ 0*, *met15 Δ 0*, *ura3 Δ 0*) as host. The pRS403-CreEBD plasmid was genomically integrated into the yeast *his3 Δ 1* site. Yeast cells were transformed with the constructed plasmid using a Frozen EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA, T2001). Synthetically defined (SD) solid medium without L-histidine was used to screen the transformants. The components of the SD solid medium were 0.67 % [w/v] yeast nitrogen base without amino acids, 2 % [w/v] glucose, and 2 % [w/v] agar supplemented with appropriate amino acids and a nucleobase (0.012 % [w/v] L-leucine, 0.002 % [w/v] L-methionine, and 0.002% [w/v] uracil). The mutagenized *lox* library has been transformed into a yeast strain that carries the CreEBD transgene. The transformants were screened on a solid SD medium in the absence of L-histidine and uracil.

Inducing Cre in yeast

I generated yeast that can express Cre when exposed to β -estradiol (pRS403-CreEBD). I used the *S. cerevisiae* strain BY4741 (*MATa*, *his3 Δ 1*, *leu2 Δ 0*, *met15 Δ 0*, *ura3 Δ 0*) as host. The pRS403-CreEBD plasmid was genomically integrated into the yeast *his3 Δ 1* site. Yeast cells were transformed with the constructed plasmid using a Frozen EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA, USA, T2001). Synthetically defined (SD) solid medium without L-histidine was used to screen the transformants. The components of the SD solid medium were 0.67 % [w/v] yeast nitrogen base without amino acids, 2 % [w/v] glucose, and 2 % [w/v] agar supplemented with appropriate amino acids and a nucleobase (0.012 % [w/v] L-leucine, 0.002 % [w/v] L-methionine, and 0.002 % [w/v] uracil). The mutagenized *lox* library has been transformed into a yeast strain that carries the CreEBD transgene. The transformants were screened on a solid SD

medium in the absence of L-histidine and uracil. In total, more than 10,000 colonies were collected to prevent the loss of diversity in the *lox* library. Colonies were pre-incubated for 24 h at 30°C, 250 rpm on liquid SD medium. The pre-incubated yeast was inoculated into 10 mL of liquid SD medium containing D-galactose as a carbon source with an OD600 of 1. Yeast cells were grown at 30°C for 24 h at 250 rpm.

Preparing DNA samples for analysis by the Illumina sequencing

The ZymoprepTM Yeast Plasmid Miniprep II Kit (ZYMO RESEARCH, D2004) was used to extract the *lox* library after Cre induction. PCR amplification was performed on the extracted samples. The following cycling parameters were used; followed by five cycles at 98°C for 10 s/55°C for 5 s/68°C for 30 s, 68°C for 7 min, and final hold at 4°C. To minimize PCR bias, the number of PCR cycles was set to five. The prepared samples were sequenced on the paired end using the Novaseq6000 (Illumina, San Diego, CA, USA). The services of MacroGen Japan (Tokyo, Japan) were used for sample quality control, the addition of adaptor sequences, the addition of index sequences, and Illumina sequencing runs.

Construction of Gaussian process regression model

Gaussian process regression modeling method applied in this study is based on previous research [23–25]. The 13 bp of recombinase binding element (RBE) of the *lox* sequence was one-hot encoded. The regression model was trained to predict the cleavage rate by Cre. The mean and standard deviation of all training data were normalized to be 0 and 1, respectively. Gaussian process regression model requires kernel functions that measure the similarity between DNA sequences, and optimizing learning requires improving the form of the kernel and its hyperparameters. The Matérn kernel was used in this study based on previous research. The mutant *lox2272* sequence evaluated by Illumina sequencing was initially used to train the regression model, and the sequence information evaluated by qPCR was added to train the model after optimization. An open-source module in the SciPy ecosystem was used to perform regression [26–28].

qPCR (quantitative polymerase chain reaction)

I performed a validation experiment using quantitative polymerase chain reaction (qPCR) to confirm the accuracy of the prediction accuracy of Gaussian process regression model. I used PCR to construct plasmids that carried the desired mutation. Next, three sets of primers

were prepared for qPCR: (1) primers for quantification of cleavage rates between *lox2272* sequences (forward primer: 5'-GTCAATCGTATGTGAATGCTGGTCGC-3', reverse primer: 5'-CGCTGCTGTCTCCCCGACATTC-3'), (2) primers for quantification of cleavage rates between *loxP* sequences (forward primer: 5' - GCCACTGAGGTTCTTCTTTTCATATACTTCCTT-3', reverse primer: 5' - CTGCAGCCGCTGCTGTCTC-3'), and (3) primers for quantification of non-cleavage rates (forward primer: 5'-TGCCCAGATGCGAAGTTAAGTGCG-3', reverse primer: 5' - AGTTTTTAAACACCAAGAAGTACTTAGTTTCGACGG-3'). In addition, I prepared Δ *loxP* plasmid and Δ *lox2272* plasmid to generate a calibration curve. The calibration curve was generated using a 7-point dilution series in 10-fold increments from 1.0×10^7 copies/ μ L to 1 copy/ μ L. Using these calibration curves, the non-cleavage rate, the cleavage rate between *loxP*, and the cleavage rate between *lox2272* of individual mutant *lox* sequences were calculated. The PCR mix included: 17 μ L distilled water, 25 μ L FastStart SYBR Green Master (without ROX) (Roche, Basel, Switzerland, 04673484001), 1.5 μ L forward primer at 10 pmol/ μ L, 1.5 μ L reverse primer at 10 pmol/ μ L, and 5 μ L template plasmid. PCR was performed using the following cycling conditions for all primer sets: 95°C for 10 min; followed by 40 cycles of 95°C for 15 s, 60°C for 30 s, and 1 cycle of 95°C for 15 s, 60°C for 1 min and 95°C for 10 s on a StepOnePlus™ instrument (Thermo Fisher Scientific, USA).

Generation of stable HEK293 cell lines with single-copy *lox* variants integration

Stable HEK293 cell lines with a single copy of *lox* variant constructs were generated using the Flp-In™ system (Invitrogen, Massachusetts, USA). In brief, the cells were grown in a temperature- and humidity-controlled incubator at 37°C in DMEM with 10 % FBS and 1 % penicillin/streptomycin. Sub-confluent cells were co-transfected with pOG44, which express FLP: pcDNA/FRT/*lox*, which has mutant *lox* sequences at a 9:1 ratio using Xfect™ Transfection Reagent (TAKARA, Tokyo, Japan). After 48 h, the cells were split into 10 cm tissue culture dishes and selected with 100 μ g/ml hygromycin until colonies were visible. Individual colonies were screened for fluorescence expression and selected to expand until they reached subconfluence for transfection. As the Flp-In system inserts the transgene into a single Frt site in the genome of Flp-In™HEK293 cells using Flp recombinase, the generated clones are expected to be isogenic and have a single copy integration.

Results

Procedure to achieve recombination efficiency at the desired rate

In the Brainbow system, Cre-mediated recombination is mutually exclusive between one of two identical *lox* site pairs (a pair of *loxP* sequences and a pair of *lox2272* sequences) [29]. Furthermore, excision by one recombination event removes the *lox* sites necessary for the occurrence of the other recombination event. In a genetic circuit in which two pairs of *lox* sequences are alternately inserted, with gene A located between *loxP* sequences and gene B situated between *lox2272* sequences, the expression of gene A and gene B becomes a stochastic and mutually exclusive process (**Figure 1**). The recognition efficiency of the two pairs of *lox* sequences is comparable [29], resulting in similar expression rates for gene A and gene B. I initially examined the reaction kinetics of Cre-*lox*-mediated intrachromosomal recombination to devise a methodology for stochastically activating gene expression with a desired sparseness level. I hypothesized that the affinity of Cre for the mutagenized *lox* sequence could be reduced compared to its affinity for another *lox* sequence, thus enabling the regulation of the expression rates of gene A and gene B through manipulation of the mutagenized *lox* sequences.

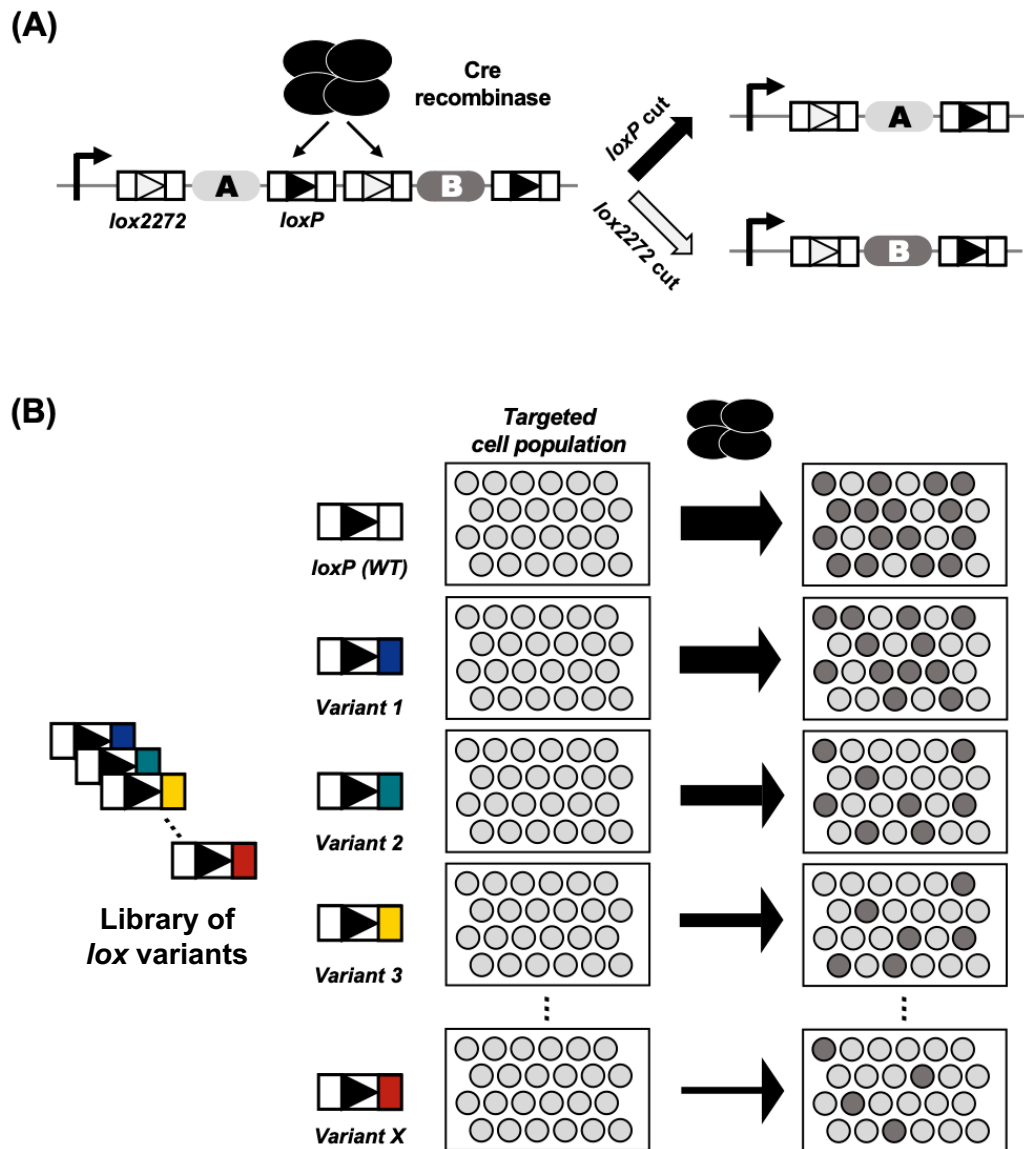


Figure 1. Schematic overview of experimental strategy. **(A)** The stochastic removal of *lox* sequences by Cre. Gene A is removed when Cre cleaves *lox2272* sequences, and gene B is removed when Cre cleaves *loxP* sequences. Cre cleaves one of two pairs of *lox* sequences (either a pair of *lox2272* sequences or a pair of *loxP* sequences) in a mutually exclusive manner. Cre-mediated excision of *lox* sequences leading to the exclusive expression of either gene A or gene B. **(B)** The regulation of the sparseness level using mutant *lox* sequences. Modulating the sparseness level through mutant *lox* sequences, which may alter the affinity of the *lox* sequence for Cre, potentially renders *lox* sequences less likely to be cleaved by Cre.

Procedures for creating and evaluating a library of mutant *lox* sequences

I developed a strategy for high-throughput analysis of the impact of mutations on the excision rate by Cre (**Figure 2**). First, one of the recombinase binding elements (RBE) of the *lox* sequence was mutated by PCR to construct a library of mutant *lox* sequences (**Figure 2A, B**). This library was then cloned into a centromere-type plasmid and introduced into *Saccharomyces cerevisiae* in conjunction with a CreEBD system, a β -estradiol-inducible Cre expression construct [30]. Cre was then induced in yeast carrying the mutant *loxP* library (**Figure 2C**). The library of mutant *lox* sequences was extracted from the yeast following Cre induction (**Figure 2D**). The library of mutant *lox* sequences was then subjected to Illumina sequencing to quantify the cleavage rate (**Figure 2E**). Finally, I used machine learning to build a model that can predict the cleavage rate of any mutant *lox* sequence based on data acquired by Illumina sequencing (**Figure 2F**).

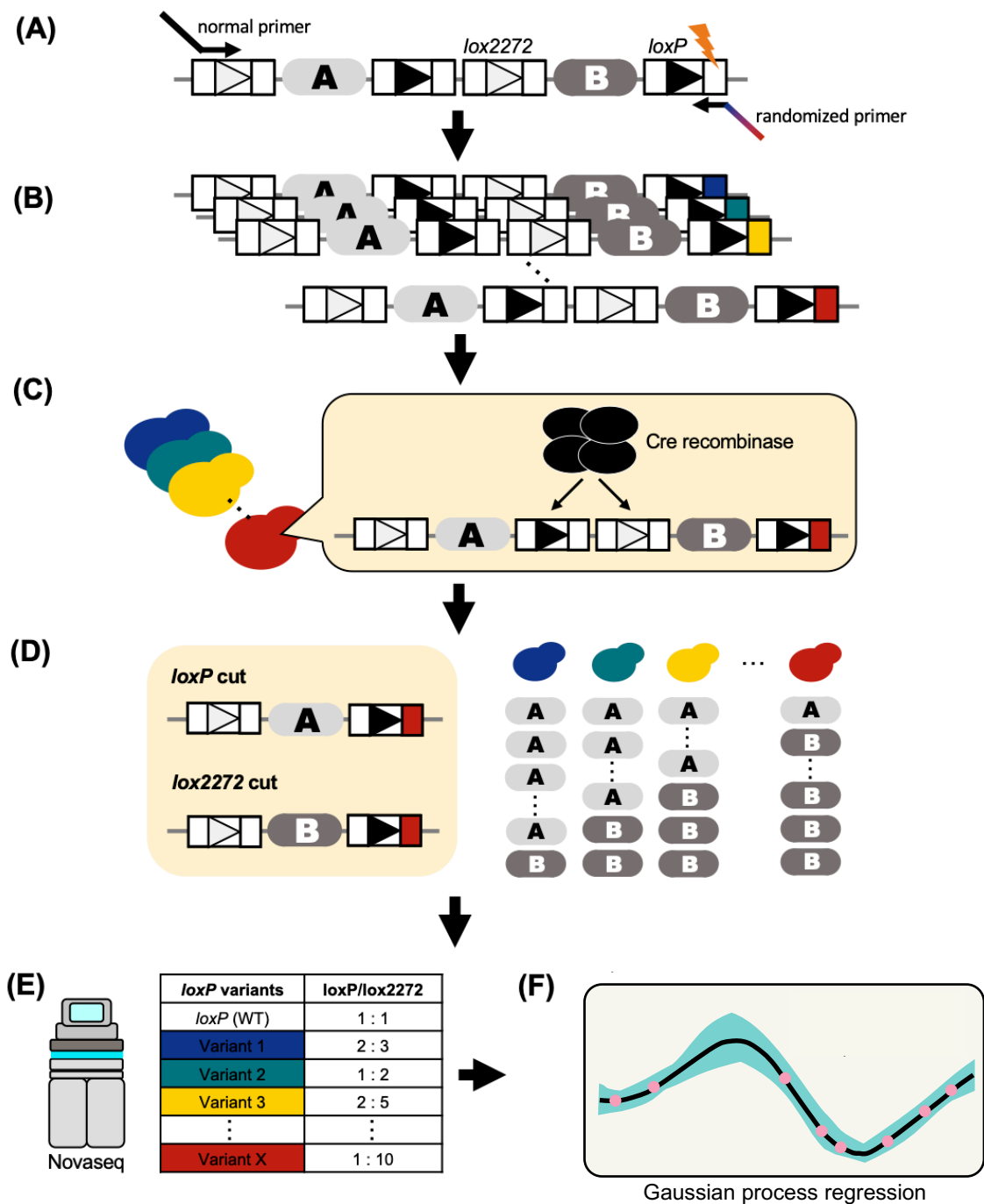


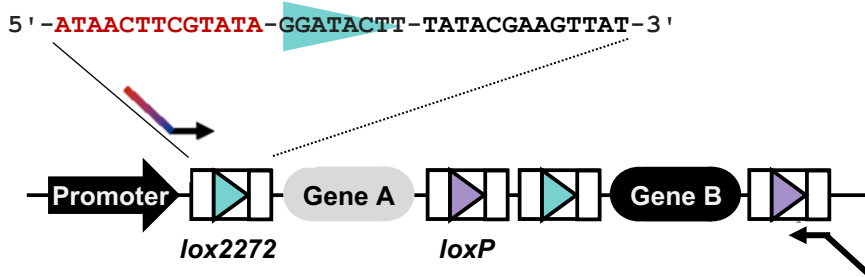
Figure 2. Schematic overview of the experimental procedure. The experimental procedure comprised six steps: **(A, B)** Introducing mutation into *lox* sequence. I mutated one of the RBEs of the *lox* sequence to create a library of mutant *lox* sequences. **(C)** Cre induction. The library of mutant *lox* sequences was introduced into *Saccharomyces cerevisiae*, and Cre was induced using the CreEBD system. **(D)** Extracting the library from *S. cerevisiae*. I extracted the library from *S. cerevisiae* after Cre induction. **(E)** Analysis. I employed Novaseq to analyze the excision rate of the mutant *lox* sequences. **(F)** Constructing a machine learning model. Utilizing data on

mutated *lox* sequences and their cleavage rates obtained using Novaseq, I developed a machine-learning model capable of predicting the Cre-mediated cleavage rate of any *lox* sequence.

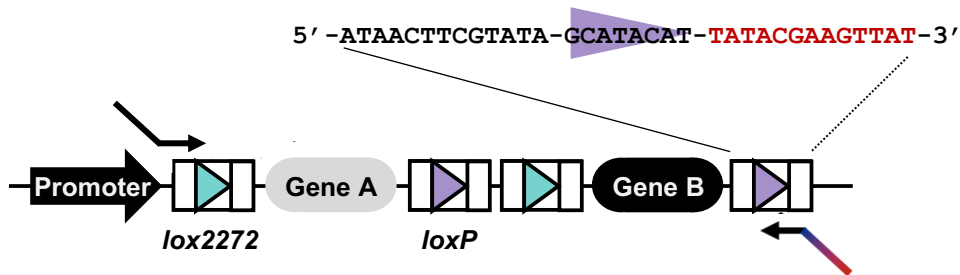
Detailed design of a library of mutant *lox* sequences

In the Cre-*lox* recombination process, Cre binds to the *lox* sequences by recognizing inverted repeat sequences comprising 13 base pairs on either side of the *lox* sequences, referred to as recombinase binding elements (RBEs) [5]. In this study, a mutation was introduced at 13 bp (5'-ATAACTTCGTATA-3') of the left RBE of the *lox2272* sequence and the right RBE of the *loxP* sequence (**Figure 3**). An increase in the number of substitutions was expected to decrease the affinity of the *loxP* mutant for Cre. It was difficult to obtain sufficient coverage for substitutions of three or more bases because the number of combinations increased exponentially. Therefore, this research aimed to evaluate the majority of mutants with one or two nucleotide substitutions, as well as a fraction of mutants with substitutions of three or more nucleotides. The RBE sequence was designed to preserve the original base with 84.7 % probability of obtaining as many 2-base substitutions as possible. For instance, if the original base was A, it was planned so that 84.7 % of it would remain A, 5.1 % would become T, 5.1 % would become G, and 5.1 % would become C.

(A) *lox2272* variants



loxP variants



(B) Original base

	mutated rate			
base	[A]	[T]	[G]	[C]
[A]	84.6 (%)	5.1 (%)	5.1 (%)	5.1 (%)
[T]	5.1 (%)	84.6 (%)	5.1 (%)	5.1 (%)
[G]	5.1 (%)	5.1 (%)	84.6 (%)	5.1 (%)
[C]	5.1 (%)	5.1 (%)	5.1 (%)	84.6 (%)

Figure 3. Design of primers to establish a library of *lox* variants. (A) I designed primers for mutagenizing the *lox* sequence, introducing a mutation into the left arm (13bp) of the *lox2272* sequence (shown in blue) and the right arm of the *loxP* sequence (shown in red). The mutation rate of both primers for *lox2272* and *loxP* was set at 15.3 % to obtain as many 2-base substitutions as possible. (B) I validated the distribution of the number of base substitutions using Sanger sequencing, with gray representing the theoretical value (mutation rate = 15.3 %) and green representing the results of Sanger sequencing.

Illumina sequencing analysis of the library of mutant *lox* sequences

Following a cleavage reaction performed by inducing Cre in yeast using the CreEBD system, the library of mutagenized *lox* sequences was subjected to Illumina sequencing. The workflow of this analysis is depicted in **Figure 4**. Initially, Illumina sequencing reads were categorized based on the type of mutation present in the *lox* sequences (**Figure 4(1), (2)**). Subsequently, I analyzed the cleavage rate of individual *lox* mutant sequences by combining the Illumina sequencing reads acquired using a 5' primer and 3' primer (**Figure 4(3)**). Analysis showed that the average non-cleavage rate was 3.5 %, confirming that Cre was induced.

Prior to conducting a comprehensive analysis of individual mutant *lox* sequences, I verified the quality of the library of mutagenized *lox* sequences. First, the distribution pattern of the number of base substitutions revealed by Illumina sequencing was almost identical to the theoretical one (**Figure 5A**). The coverage of each base substitution was also examined. The results showed that 100 % (39/39 variants) of single-base substitutions and 92.6 % (650/702 variants) of the 2-base substitutions in the library of mutant *lox2272* sequences and that 100 % (39/39 variants) of the single-base substitutions and 60.5 % (425/702 variants) of the 2-base substitutions in the library of mutant *loxP* sequences (**Figure 5B**). Third, I evaluated the nucleotide bias at each of the 13 positions within the left RBE of the *lox2272* sequence and the right RBE of the *loxP* sequence. The results indicated no bias in the type of bases introduced at each of the positions within the RBE (**Figure 5C**).

Subsequently, the cleavage rates of individual *lox2272* and *loxP* sequences were analyzed. As shown in **Figure 6**, mutant *lox2272* and *loxP* sequences with reduced recognition efficiency by Cre were obtained at various rates. This result supports the hypothesis that sparseness can be controlled by mutating the RBE of the *lox2272* and *loxP* sequences.

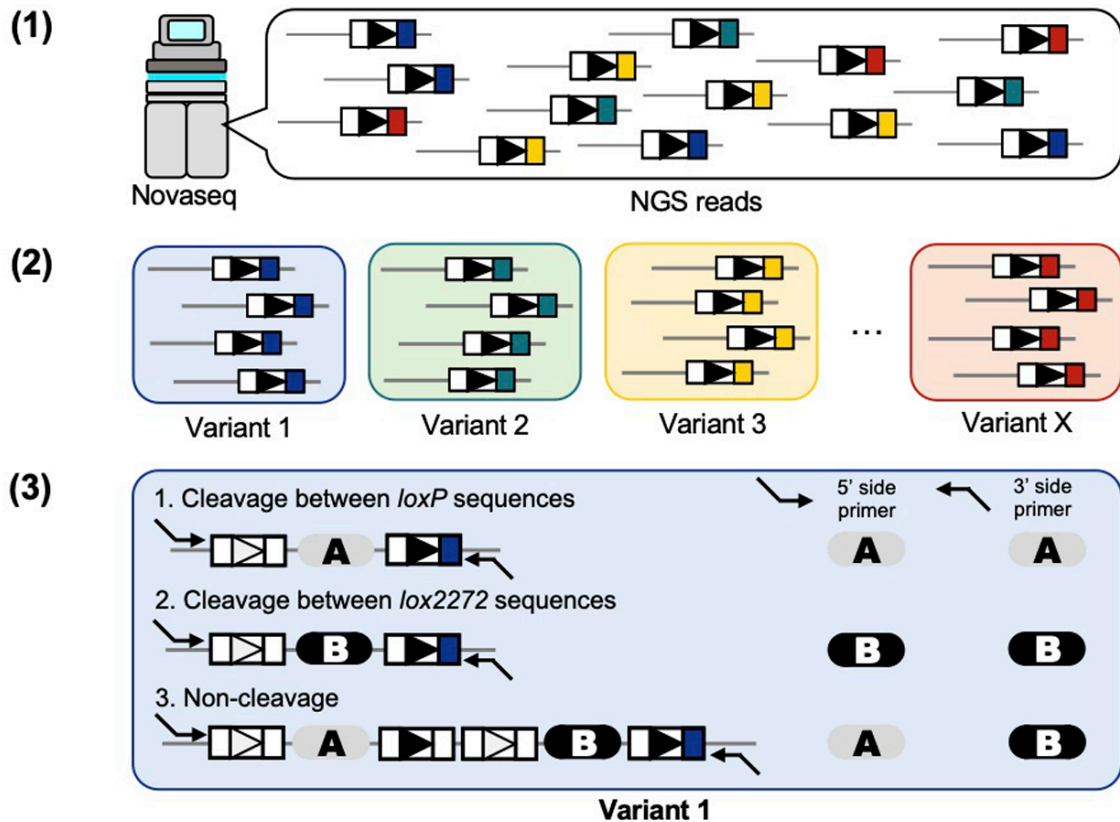


Figure 4. Schematic diagram of the Illumina sequencing analysis. Illumina sequencing data were analyzed in three steps. Firstly, DNA sequences with the sequence of the *lox* variant were extracted. Secondly, all paired-end Illumina sequence reads were classified according to the type of *lox* variants. Lastly, utilizing the paired-end Illumina sequencing reads generated using primers at the 5' and 3' sides, I determined the patterns of recombination of Illumina sequencing reads for each variant. If the Illumina sequence read had sequence A, the plasmid was determined to have been cleaved between *loxP* sequences by Cre. If the Illumina sequence read had sequence B, the plasmid was determined to have been cleaved between *lox2272* sequences by Cre. If the read had both sequences A and B, the plasmid was determined that no cleavage had occurred.

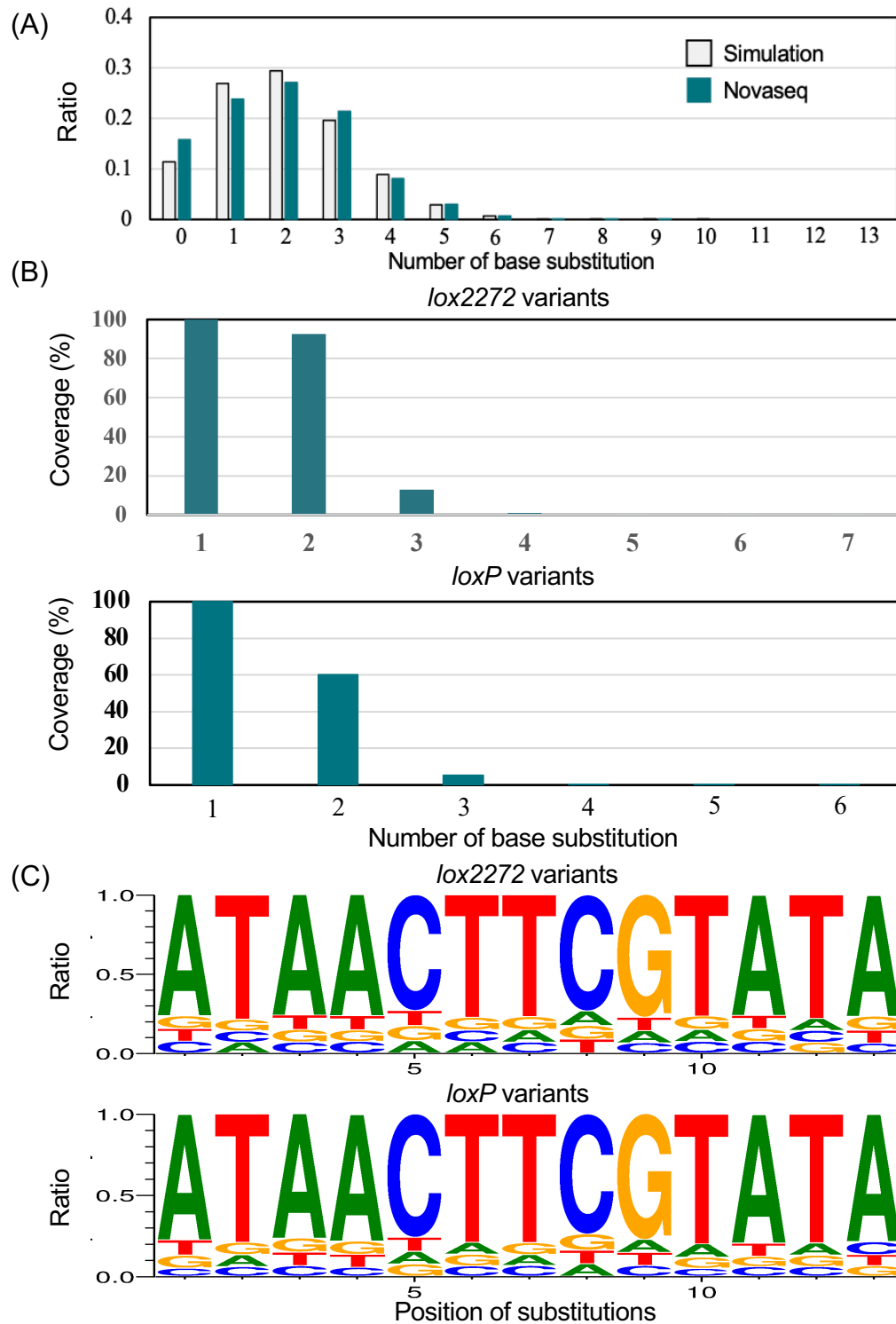


Figure 5. Evaluation of characteristics of mutant *lox* sequence library by Novaseq. (A) The distribution of the number of substituted bases between experimental and theoretical values is compared. The theoretical values were represented by a light-gray color (substitution rate =

15.3 %). The experimental values were represented by green color. (B) Analysis of the coverage rate (%) of substitution. (C) Percentage of base substitutions at each position in the sequence.

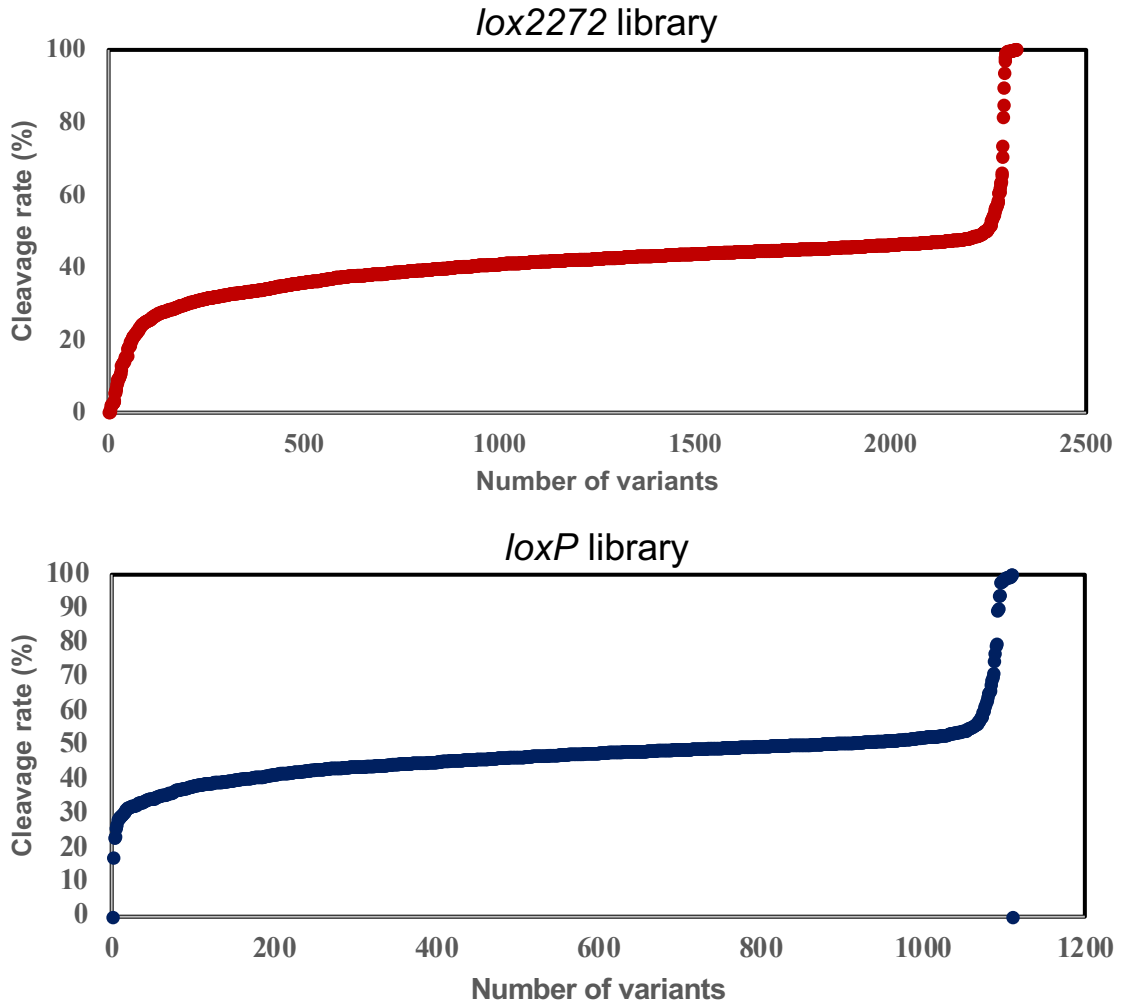


Figure 6. Analysis of cleavage rate of mutant *lox2272* and *loxP* variants. The distribution of all *lox2272* and *loxP* variants was evaluated. The vertical axis denotes the cleavage rate. Variants with decreased recognition rate by Cre were arranged in ascending-order of the cleavage rate between *loxP* sequences. Some overlap between plots was observed.

Construction of a Gaussian process regression model capable of predicting the cleavage rate of *lox2272* sequence at a desired sparseness level

Subsequently, I engineered a machine learning model that can predict the cleavage rate by Cre of any mutant *lox2272* sequence to manipulate the labeling rate freely. The model I selected was a Gaussian process regression (GPR) model. The GPR model was chosen because (1) it can be used even when linear regression cannot be used for fitting, and (2) it can represent the uncertainty in estimation because the function can be obtained as a distribution. Initially, the library of mutant *lox2272* sequences obtained from Illumina sequencing analysis was randomly divided into 9:1 ratio for model training and model test data, respectively (**Figure 7A**). Next, we performed one-hot encoding of the 13 bp DNA sequence for the training and test data, respectively, using 5-digit binary notation. I tested whether a model could accurately predict the cleavage rate of the test data by training on the dataset based on the position of mutation and the effect of that mutation on the cleavage rate. Then, I tested whether the GPR could accurately predict the cleavage rate by Cre of an evaluated *lox2272* sequence. First, I evaluated the prediction performance of the constructed model from two perspectives, 20-fold cross-validation, and test data. Then, I examined the correlation between the values predicted by the model and the measured values. The correlation coefficient $R = 0.82$ for 20-fold cross-validation and $R = 0.89$ for test data (**Figure 7B**). Next, I predicted the unevaluated mutant *lox2272* sequence using the constructed Gaussian process regression model. I examined the correlation with the predicted values and measured values by qPCR and found the correlation coefficient $R=0.92$, indicating that the cleavage rate was predicted with very high accuracy (**Figure 7C**). On the other hand, there was a discrepancy between the predicted and measured values when I focused on only the mutants with a low cleavage rate. Therefore, I added new mutants evaluated by qPCR as training data to optimize the Gaussian process model and re-trained the model. The 20-fold cross-validation correlation coefficient for the model after optimization was $R=0.88$, and the correlation coefficient for the test data was $R=0.92$, indicating an improvement in prediction accuracy over pre-optimization (**Figure 7D**). I also evaluated the cleavage rate of new mutant *lox2272* sequences that have never been evaluated and assessed the prediction accuracy of the Gaussian process regression models before and after optimization. The results showed that the pre-optimized model was less accurate in predicting the cleavage rate of *lox* sequences with the low cleavage rate. In contrast, the optimized model was relatively accurate in predicting the low cleavage rate. The overall correlation was also higher in the post-optimized model (**Figure 7E**). Based on these findings, I succeeded in constructing a Gaussian process regression model that can predict the cleavage rate of any *lox2272* sequence by Cre.

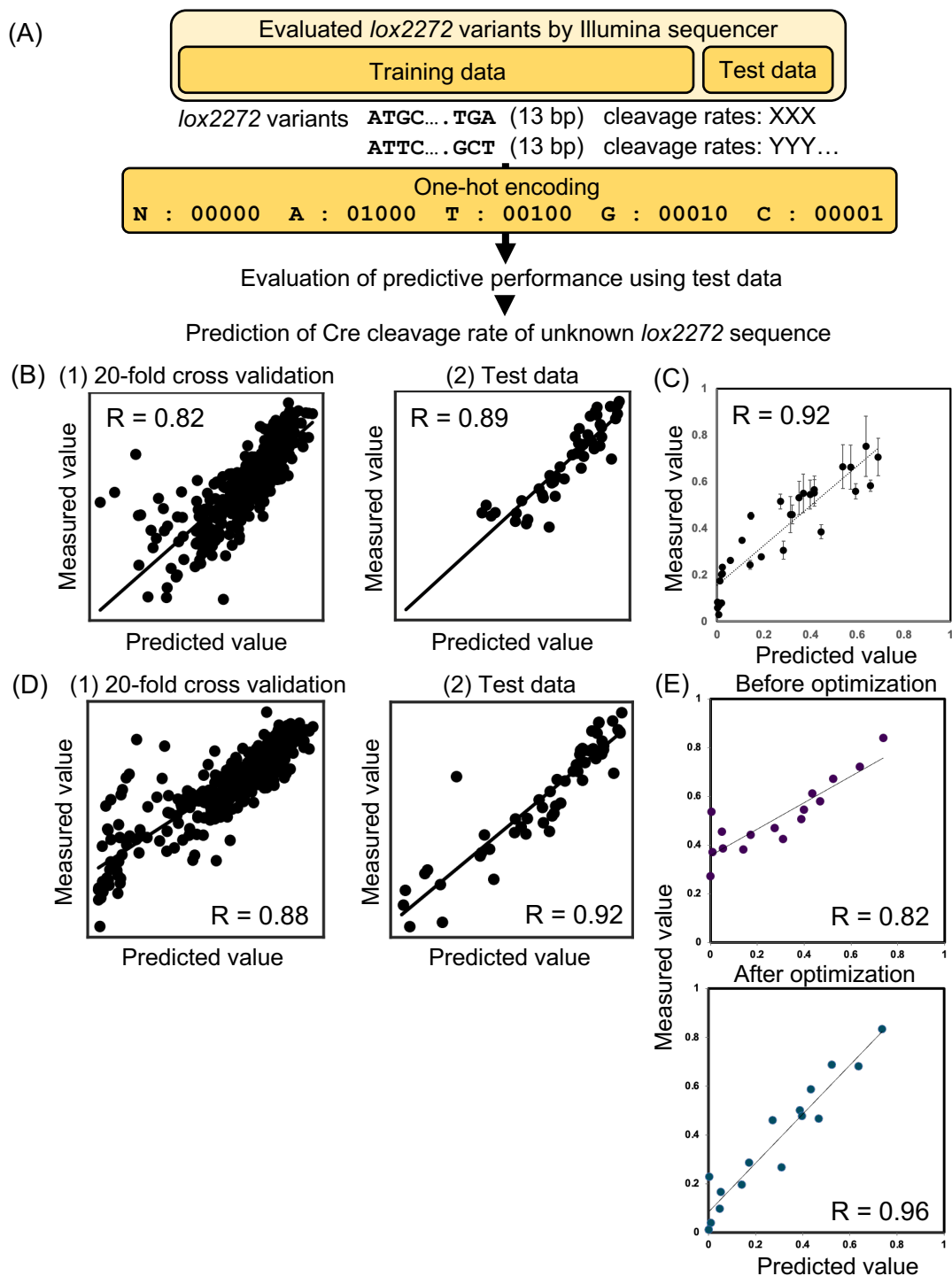


Figure 7. Gaussian process regression modeling. (A) Schematic of the workflow of constructing a Gaussian process regression model. (B) Evaluating the predictive accuracy of gaussian process regression models (1) 20-fold cross-validation (2) test data. (C) Validation of model accuracy by qPCR using unevaluated *lox2272* sequences. (D) Reassessing the predictive accuracy of Gaussian process regression models after optimization. (E) Comparison of model performance before and after optimization by qPCR.

Evaluation of the scalability of the Gaussian process regression model using cultured cells

The Gaussian process regression model has been constructed up to this point using data evaluated in yeast to facilitate high-throughput evaluation. However, in the future, it is desirable that the mutant *lox2272* sequence predicted by the constructed gaussian process regression model can be applied to other eukaryotic organisms, such as *Caenorhabditis elegans* used in the functional cellomics. To confirm the applicability of the constructed Gaussian process model to other eukaryotic organisms, validation was performed using human cultured cells. Seven randomly selected sequences from the mutant *lox2272* sequences evaluated in yeast were integrated into the genome of human cultured cells, and the genome was extracted following Cre induction via the transient expression of the Cre-expressing gene. The extracted genome was then used as a template, and the cleavage rate of each mutant *lox2272* sequence was quantified by qPCR. As a result, the correlation coefficient between the *lox2272* cleavage rate by qPCR in yeast and that by qPCR in cultured cells was $R = 0.93$ (**Figure 8**). This result suggests that the Gaussian process regression model constructed in yeast could be widely used in other eukaryotes, such as cultured cells.

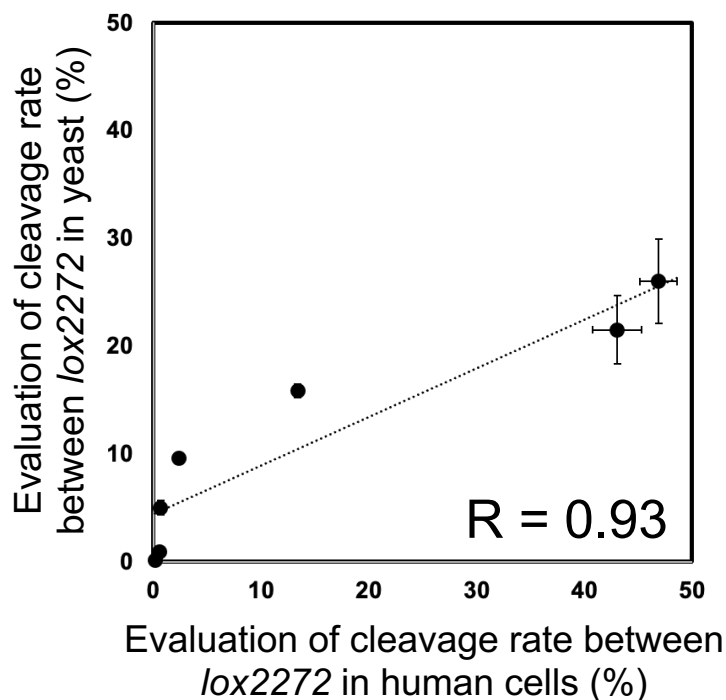


Figure 8. Comparison of cleavage rates of *lox2272* variants in cultured human cells and yeast.

Discussion

One of the problems with sparseness labeling identified in previous studies was that the sparseness level of the effector was less regulatable and reproducible. In this research, I devised a novel technique of sparse labeling that overcame the deficiencies of previous studies. I aimed to obtain mutant *lox* sequences by incorporating random mutation in the RBE sequence based on the hypothesis that *lox* variants with diminished recognition efficiency by Cre can regulate the sparseness level in a desirable manner. I examined 2321 of *lox2272* variants and 1111 of *loxP* variants, and succeeded in identifying mutants with recognition efficiencies ranging from 0.05 % to 100 %. These results confirm the hypothesis that the recognition efficiency of *lox* sequences by Cre can be controlled by introducing precise mutations in the arms of the *lox* sequence.

Additionally, experiments verified that as the number of substitutions introduced into the RBE of the *lox* sequence increased, the efficiency of *lox* recognition by Cre decreased. I also discovered that some mutants with fewer base substitutions than other mutants with more base substitutions have lower cleavage rates. This result implies that the evaluation of individual mutant *lox* sequences is crucial to achieving sparse labeling at the desired rates. Furthermore, using the data obtained from Illumina sequencing, I have successfully constructed a Gaussian process regression model that can predict the mutant *lox2272* sequence, allowing sparse labeling at any desired labeling rate. The mutant *lox2272* sequence obtained in this study, which exhibits a labeling rate of 5.7×10^{-4} %, is the sparsest among the previous studies [6–17]. Furthermore, the mutant *lox2272* sequence showed a high correlation in the cleavage rate by Cre between yeast and human cultured cells, suggesting that the established Gaussian process regression model may be adaptable in various eukaryotes.

As stated in the introduction, several previous studies have assessed the impact of the mutation on the cleavage rate of *lox* sequences by Cre. Hartung, M. & Kisters-Woike, B. examined the impact of the mutation on Cre recombinase [15]. However, it is difficult to precisely control the sparseness level by introducing mutations in Cre recombinase. In contrast, our methodology allows precise control of the sparseness level by selecting an appropriate mutant *lox* sequence while utilizing existing Cre lines. Then, Missirlis, P. I., *et al.* incorporated mutations in the spacer region of *lox* sequences [8]. Given that the spacer sequence determines specificity, many mutants are expected to completely lose recognition by Cre, making it an inappropriate approach for regulating the sparseness level. Sheren J *et al.* examined the effects of introducing mutations in the spacer region and RBE of the *loxP* sequence, respectively [9]. In experiments, the mutant *lox2272* sequences compete with the *loxP* sequence (and mutant *loxP* sequences compete with the *lox2272* sequence). On the other hand, Sheren J *et al.* evaluated the mutant *loxP* sequence alone. Furthermore, Cre cannot cleave the other *lox* sequence if it cleaves one *lox*

sequence, whereas if the mutant *lox* sequence is present alone, Cre can always cleave it. Therefore, the cleavage rate of mutant *lox* sequences in the presence of other competitive *lox* sequences remains unclear. In addition, the cleavage rates evaluated in this study for more than 2000 mutant *lox2272* sequences and more than 1000 mutant *loxP* sequences allow us to adjust the sparseness level very strictly. Furthermore, the cleavage rate by Cre of the exhaustive mutant *lox2272* sequence can be predicted by the Gaussian process regression model constructed in this study. In conclusion, this study provides the first comprehensive dataset for calculating the cleavage rate of mutant *lox* sequences in the presence of other *lox* sequences.

The capacity to adjust the sparseness level at the desired rate with high reproducibility is a significant advantage of this method. One potential application of this method is in combination with the cellomics approach established in chapter I, which employs the Cre-*lox* system to stochastically label opsin in a small population of neural networks of *Caenorhabditis elegans*. By utilizing mutant *lox* sequences in the cellomics approach, it may be possible to intervene in the activity of a smaller number of neurons. Other examples are the generation of genetic mosaics to analyze population doses of genes involved in sporadic genetic diseases or the application of this method to promote cancer development. The method may be applied to mimic the cell-competitive environment when a minority of cancer cells are present in a healthy cell population by progressively transforming healthy cells into cells expressing cancer-inducing genes at desired levels [31–36]. Overall, as a purely genetic method, it has the potential to be adapted to a wide range of research fields.

References

1. Sternberg, K. & Asd, *Bacteriophage P1 Site-specific Recombination I. Recombination Between loxP Sites. I. Mol. Biol.* vol. 150 (1981).
2. Hoess, R. H., Ziese, M. & Sternberg, N. *P1 site-specific recombination: Nucleotide sequence of the recombining sites (dyad symmetry/lox site/BAL-31 deletion/plasmid recombination) A. Proc. Natl Acad. Sci. USA* vol. 79 <https://www.pnas.org> (1982).
3. van Duyne, G. D. *A STRUCTURAL VIEW OF Cre-loxP SITE-SPECIFIC RECOMBINATION.* www.annualreviews.org (2001).
4. van Duyne, G. D. Cre Recombinase. (2015) doi:10.1128/microbiolspec.
5. Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T. & Buchholz, F. Cre Recombinase and Other Tyrosine Recombinases. *Chem Rev* **116**, 12785–12820 (2016).
6. Lee, G. & Saito, I. *Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene* vol. 216 (1998).
7. Thomson, J. G., Rucker, E. B. & Piedrahita, J. A. Mutational analysis of *LoxP* sites for efficient cre-mediated insertion into genomic DNA. *Genesis (United States)* **36**, 162–167 (2003).
8. Missirlis, P. I., Smailus, D. E. & Holt, R. A. A high-throughput screen identifying sequence and promiscuity characteristics of the *loxP* spacer region in Cre-mediated recombination. *BMC Genomics* **7**, (2006).
9. Sheren, J., Langer, S. J. & Leinwand, L. A. A randomized library approach to identifying functional *lox* site domains for the Cre recombinase. *Nucleic Acids Res* **35**, 5464–5473 (2007).
10. Siegel, R. W., Jain, R. & Bradbury, A. Using an in vivo phagemid system to identify non-compatible *loxP* sequences. *FEBS Lett* **505**, 467–473 (2001).
11. Hoess, R. H., Wierzbicki, A. & Abremski, K. *The role of the loxP spacer region in P1 site-specific recombination. Nucleic Acids Research* vol. 5 <https://academic.oup.com/nar/article/14/5/2287/2382062> (1986).
12. Sauer, B. *Multiplex Cre/lox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. Nucleic Acids Research* vol. 24 (1996).
13. Albert, H., Dale, E. C., Lee, E. & Ow, D. W. Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *The Plant Journal* **7**, 649–659 (1995).
14. Langer, S. J., Ghafoori, A. P., Byrd, M. & Leinwand, L. *A genetic screen identifies novel non-compatible loxP sites.*
15. Hartung, M. & Kisters-Woike, B. *Cre Mutants with Altered DNA Binding Properties*.*

- <http://www.jbc.org> (1998).
16. Rüfer, A. W. & Sauer, B. *Non-contact positions impose site selectivity on Cre recombinase. Nucleic Acids Research* vol. 30 (2002).
 17. Eroshenko, N. & Church, G. M. Mutants of Cre recombinase with improved accuracy. *Nat Commun* **4**, (2013).
 18. Hadjiconomou, D. *et al.* Flybow: Genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. *Nat Methods* **8**, 260–266 (2011).
 19. Hampel, S. *et al.* *Drosophila* Brainbow: A recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat Methods* **8**, 253–259 (2011).
 20. Sethi, S. & Wang, J. W. A versatile genetic tool for post-translational control of gene expression in *Drosophila melanogaster*. doi:10.7554/eLife.30327.001.
 21. Kage-Nakadai, E. *et al.* A conditional knockout toolkit for *Caenorhabditis elegans* based on the Cre/loxP recombination. *PLoS One* **9**, (2014).
 22. Aoki, W. *et al.* Cellomics approach for high-throughput functional annotation of *Caenorhabditis elegans* neural network. *Sci Rep* **8**, (2018).
 23. Bedbrook, C. N. *et al.* Machine learning-guided channelrhodopsin engineering enables minimally invasive optogenetics. *Nat Methods* **16**, 1176–1184 (2019).
 24. Bedbrook, C. N., Yang, K. K., Rice, A. J., Gradinaru, V. & Arnold, F. H. Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization. *PLoS Comput Biol* **13**, (2017).
 25. Romero, P. A., Krause, A. & Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proc Natl Acad Sci U S A* **110**, (2013).
 26. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng* **13**, 22–30 (2011).
 27. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering* **9**: 90–95. (2007)
 28. Oliphant TE. Python for Scientific Computing. *Computing in Science and Engineering* **9**: 10–20. (2007)
 29. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
 30. Jia, B. *et al.* Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nat Commun* **9**, (2018).
 31. Deepak Muzumdar, M., Luo, L., Zong, H. & Howard, § †. *Modeling sporadic loss of heterozygosity in mice by using mosaic analysis with double markers (MADM)*. www.pnas.org/cgi/content/full/ (2007).
 32. Muzumdar, M. D. *et al.* Clonal dynamics following p53 loss of heterozygosity in Kras-driven

- cancers. *Nat Commun* **7**, (2016).
33. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
 34. Gonzalez, P. P. *et al.* p53 and NF 1 loss plays distinct but complementary roles in glioma initiation and progression. *Glia* **66**, 999–1015 (2018).
 35. Tian, A. *et al.* Oncogenic State and Cell Identity Combinatorially Dictate the Susceptibility of Cells within Glioma Development Hierarchy to IGF1R Targeting. *Advanced Science* **7**, (2020).
 36. Yao, M. *et al.* Astrocytic trans-Differentiation Completes a Multicellular Paracrine Feedback Loop Required for Medulloblastoma Tumor Growth. *Cell* **180**, 502-520.e19 (2020).

Chapter III

Neuronal cell subclass-specific proteomic analysis of the *Caenorhabditis elegans*

The cellomics approach developed in Chapter I has enabled the high-throughput identification of the cell or cells from the cellular networks that generate biological functions. The next step is to develop a methodology to investigate which molecular mechanisms in the identified cells contribute to the biological functions at the network level. Thus, I attempted to obtain expression information related to cell function by using cell-selective BONCAT to analyze the proteome of a single-cell subtype in *Caenorhabditis elegans*. So far, proteome analysis of neural cells by in vivo cell-selective labeling has only been reported in mice and *Drosophila* [1, 2]. BONCAT has rarely been applied in *C. elegans*. Only the pharyngeal muscle of *C. elegans* has undergone in vivo cell-selective labeling for proteome analysis [3]. Another example is the proteome analysis of newly produced proteins utilizing protein azidation techniques [4]. Therefore, in this study, I attempted cell-selective BONCAT using monolith nano LC-MS/MS for the first time to analyze the proteome of selected subclasses of neural cells in *C. elegans*.

Materials and methods

Maintenance of nematode

The *Caenorhabditis* Genetics Center (CGC) provided the *Caenorhabditis elegans* N2 (Bristol) strain and the *Escherichia coli* OP50-1 strain (*ura⁻*, *strR*). The worms were grown and maintained on nematode growth medium (NGM) plates containing *Escherichia coli* OP50-1 in 3 cm or 6cm dishes [5].

Plasmids prepared in this research

pGH8 was provided by Erik Jorgensen (Addgene plasmid # 19359; <https://n2t.net/addgene:19359>; RRID: Addgene_19359) [6]. pKPY197 was provided by David Tirrell (Addgene plasmid # 62599; <https://n2t.net/addgene:62599>; RRID: Addgene_62599) [3]. pCFJ104 (Pmyo-3::mCherry::unc-54 3'UTR) was provided by Erik Jorgensen (Addgene plasmid # 19328; <https://n2t.net/addgene:19328>; RRID: Addgene_19328) [6]. pKPY514 was provided by David Tirrell (Addgene plasmid #

62598; <https://n2t.net/addgene:62598>; RRID: Addgene_62598) [3]. pHW394 (15xUAS::GFP::let-858 3'UTR) was provided by Paul Sternberg (Addgene plasmid # 85584 ; <https://n2t.net/addgene:85584>; RRID:Addgene_85584) [7]. pF25B3.3p::mcherry was previously constructed [8].

For the construction of pKPY197-Prab3 and pKPY197-Pgcy-8, pKPY197 was digested with Sall. The *rab-3* and *gcy-8* promoters were cloned from pGH8 and *C. elegans* genomes, respectively [3]. Each fragment was integrated into the digested pKPY197 plasmid. Primers prepared in this study are shown in Table 1. For the construction of pF23B2.10p::GFP, the 5' regulatory region of *F23B2.10* was cloned from the *C. elegans* genome and inserted into pCFJ90. pCFJ90 with the 5' regulatory region of *F23B2.10* was linearized by PCR amplification and the mCherry region was omitted. GFP (S65C) was amplified from pHW394. These fragments were joined together.

Table 1

Sequence (5' - 3')	Description
ttgcatgcctgcaggatcttcagatgggagcagt	Cloning from pGH8
atcctctcatgctgatgctttttgtacaaactgtc	Cloning from pGH8
ttgcatgcctgcaggagccaattttaacgggg	Cloning from <i>C. elegans</i> genome
atcctctcatgctgattgatgtgaaaaggtagaatc	Cloning from <i>C. elegans</i> genome
tgtatagaaaagttgCGCCAGTTACAACAGAGAGTG	Cloning from <i>C. elegans</i> genome
gtacaaactgtcatTGTGCCGAAATTTAAATTTAAATGA	Cloning from <i>C. elegans</i> genome
taccagcttctgtacaaagtgggtg	Cloning from pCFJ90 with F23B2.10
ttgtacaaactgtcatTGTGCCG	Cloning from pCFJ90 with F23B2.10
acaagttgtacaaaaagcaggcttaATGAGTAAAGGAGAAGAATTTTCACTGGAGTTG	Cloning from pHW394
caagaaagctgggtaCTATTGTATAGTTCATCCATGCCATGTGTAATCCC	Cloning from pHW394

Primers used in this study are listed.

Constructing transgenic nematode strains

Using a stereomicroscope (SZX10; Olympus, Tokyo, Japan) equipped with a Femtojet 4i (5,252,000.021; Eppendorf, Hamburg, Germany) and Femtotips II (1,501,040; Eppendorf), injections were performed on an N2 background. By co-injection of two plasmids [10 ng/μL of pKPY197-Prab3 and 90 ng/μL of pUC19 in water] into *C. elegans*, the strain TG1 (*TGIs1[Prab-3::frs-1(Thr412Gly)::fib-1/rps-16::gfp(S65C, synIVS)::unc-54 3'UTR]*) was generated. By UV irradiation, Ex arrays of three GFP-expressing strains were incorporated into the *C. elegans* genome [9]. By co-injection of two plasmids [45 ng/μL of pKPY197-Pgcy-8 and 5 ng/μL of pCFJ104 in water] into *C. elegans*, the strain TG2 (*TGIs2[Pgcy-8::frs-1(Thr412Gly)::fib-1/rps-16::gfp(S65C, synIVS)::unc-54 3'UTR, Pmyo-3::mcherry]*) was generated. By UV irradiation, Ex arrays of three GFP-expressing strains were incorporated into the *C. elegans* genome. By co-injection of two plasmids [35 ng/μL of pF23B2.10p::GFP, 50 ng/μL of pF25B3.3p::mcherry and

15 ng/ μ L of pUC19 in water] into *C. elegans*, the strain TG3 (*TGEx3[Pf23b2.10::GFP::unc-54 3'UTR, Pf25b3.3::mcherry]*) was generated.

Labeling of *C. elegans* with azido phenylalanine (Azf)

Escherichia coli KY33 [pKPY514], donated by David Tirrell, is a strain that is auxotrophic for arginine, lysine, and phenylalanine [3]. *Escherichia coli* KY33 was labeled with azido-phenylalanine according to the methods of a previous report [3]. Worms were pre-cultured with 5 mL S medium supplemented with 15 mg/mL *E. coli* KY33 grown on phenylalanine at 20°C, 250 rpm [5]. Pre-cultured worms were centrifuged at 300 g for 5 min at room temperature to form pellets. The pellets were washed with 1 mL of S medium. This process was repeated three times. The *C. elegans* pellet was incubated for 24 h at 20°C and 250 rpm in 5 mL S medium supplemented with 15 mg/mL azido-phenylalanine-cultured *E. coli* KY33. Labeled nematodes were collected using a 20 μ m nylon filter (pluriStrainer 20 μ m; pluriSelect, Leipzig, Germany). The nematodes were washed with 5 mL M9 buffer (0.6 % w/v Na₂HPO₄ (Nacalai Tesque, Kyoto, Japan), 0.3 % KH₂PO₄ (Nacalai Tesque), 0.5 % NaCl (Nacalai Tesque)). Nematodes were collected by centrifuging at 300 g for 5 min and processed according to the following procedures.

Nematode fixation and TAMRA staining

As described in a previous report, nematodes were fixed and labeled with dibenzocyclooctyne-PEG4-Fluor 545 (TAMRA-DBCO; Sigma-Aldrich, St. Louis, MO, USA) [3].

Fluorescence Microscopy and Imaging

A 2 % agarose pad was prepared and coated with 5 μ L 1mM levamisole (Tokyo Chemical Industry Co., Ltd., Tokyo, Japan) in M9 buffer. Worms were picked up and placed on the agarose pad with levamisole. Then, a coverslip was gently placed over the worms. Using confocal laser scanning microscopy (LSM 700; Carl Zeiss, Oberkochen, Germany), fluorescence was observed. The 488 nm, 555 nm, and 561 nm lasers were used to observe the fluorescence of GFP, TAMRA, and mCherry, respectively. Image processing was performed with Zen Lite and ImageJ [10].

Preparing a sample for neuronal subclass-selective proteomics

The labeled worms were given lysis buffer (8 M Urea (Nacalai Tesque), 4 % CHAPS (Dojindo, Kumamoto, Japan) and 1 M NaCl, in 200 mM Tris-HCl pH 8.0 (Nacalai Tesque)), and a probe sonicator with a 1/8" probe was used to sonicate the mixture on ice (10 s on, 10 s off, 1,000 J). Sonicated samples were centrifuged at 10,000 g for 5 min. I selectively enriched azide-phenylalanine (Azf)-labeled proteins using a Click-iT™ Protein Enrichment Kit for click chemistry capture of azide-modified proteins (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. On an alkyne-agarose column, the enriched samples were trypsinized. Samples were desalted with MonoSpinC18 (GL sciences, Osaka, Japan) and vacuum centrifuged. The desiccated peptides were dissolved in 25 µL 0.1 % formic acid (Wako, Osaka, Japan).

Nano LC–MS/MS analysis

Proteomic analysis was conducted as described in a previous report [11]. Briefly, 5 µl samples were injected, and peptides were separated using a liquid chromatography (LC, Ultimate 3,000; Thermo Fisher Scientific) –tandem mass spectrometry (MS/MS, LTQ Orbitrap Velos Mass Spectrometer; Thermo Fisher Scientific) system equipped with a long monolithic silica capillary column (490 cm, 75 µm internal diameter) at a flow rate of 280 nL/min. A gradient was achieved by changing the ratio of two eluents: eluent A, 0.1 % (v/v) formic acid and eluent B, 80 % acetonitrile containing 0.1 % (v/v) formic acid. The gradient began with 5 % B, increased to 45 % B for 750 min, further increased to 95 % B to wash the column for 140 min, then returned to the initial condition and held for re-equilibration. The separated analytes were detected using a mass spectrometer with a full scan range of 350–1,500 m/z (resolution, 60,000), followed by ten data-dependent collision-induced dissociation MS/MS scans. The temperature of the ion transfer tube was set to 280 °C, and dynamic exclusion was 180 s. The electrospray ionisation voltage was set at 2.3 kV.

Data analysis

Proteome Discoverer 2.1 (Thermo Fisher Scientific) was used for data analysis. Protein identification was performed with a precursor mass tolerance of 20 ppm and a fragment ion mass tolerance of 0.8 Da using MASCOT (Matrix Science, London, UK) against the *C. elegans* UniProt protein database (27/08/17). Carbamidomethylation of cysteine was designated as a fixed modification, and Oxidation of methionine and acetylation of protein N-terminals were

designated as dynamic modifications. The data were filtered using a cut-off criterion of ≤ 0.01 (q value), corresponding to a false discovery rate of 1 % at the spectral level. WormBase.org was used for tissue enrichment analysis (TEA) [12,13].

Results

Experimental strategy

The experimental strategy for the neuronal subclass-specific proteome analysis is depicted in **Figure 1**. Transgenic *C. elegans* strains producing MuPheRS and GFP in a specific neuronal subclass were generated. MuPheRS involves the substitution of Thr412 in PheRS with Gly, enabling more efficient incorporation of Azf. This substitution allows for the labeling of MuPheRS-expressing cells with Azf [4]. To label the proteins with Azf in the desired neuronal subclass, the transgenic *C. elegans* strains were grown with the Azf-labeled *E. coli*. Monolithic nano LC-MS/MS can be used for proteome analysis to disclose the proteomic composition of interest cells following the enrichment of azide-labeled proteins using alkyne-agarose.

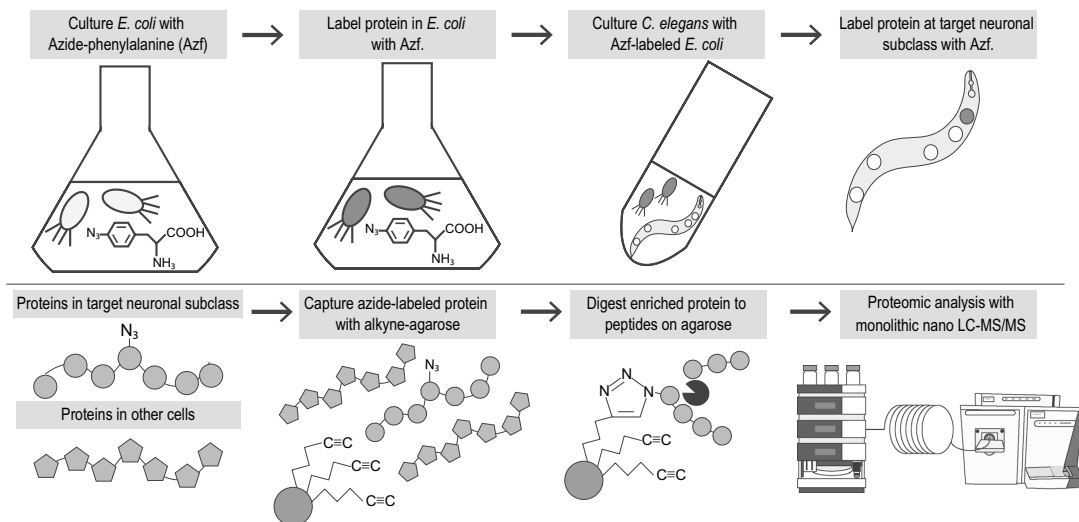


Figure 1. Experimental procedure of this research. *E. coli* was cultured with azide-phenylalanine (Azf) to label proteins within the bacteria with Azf. Subsequently, *C. elegans* was cultured with the Azf-labeled *E. coli* to label proteins in a targeted subset of neurons with Azf. Azide-modified proteins within the targeted subset of neurons were enriched using alkyne agarose after protein extraction. These enriched proteins were digested with trypsin and subjected to proteomic analysis using monolithic nano LC-MS/MS.

Confirmation of Azf incorporation by MuPheRS in target cells

Transgenic *C. elegans* strains producing MuPheRS and GFP under the *rab-3* promoter [Prab-3::frs1(Thr412Gly)::fib-1/rps-16::gfp(S65C, synIVS)::unc-54] or *gcy-8* promoter [Pgcy-8::frs1(Thr412Gly)::fib-1/rps-16::gfp(S65C, synIVS)::unc-54, Pmyo-3::mcherry] were generated [14–16]. The *rab-3* promoter is known to drive expression in all neurons and the *gcy-8* promoter was known to be expressed in AFD neurons [15, 16]. AFD neurons, essential for thermotaxis, were chosen as a model for MuPheRS production in a single neuronal subclass [17]. Fluorescence microscopy was used to observe the transgenic strain, and GFP fluorescence was confirmed in targeted cells (**Figure 2**). To confirm Azf incorporation, the transgenic strain was then fixed with TAMRA-DBCO. As a result, only when the transgenic strain was co-cultured with Azf-labeled *E. coli* KY33 and stained with TAMRA-DBCO, TAMRA fluorescence successfully was identified in neurons. No strong TAMRA fluorescence signal was observed in the transgenic strain cultured with Azf-labeled *E. coli* KY33 and not stained with TAMRA-DBCO. Both the wild-type N2 strain cultivated with Azf- or Phe-labeled *E. coli* KY33 and stained with TAMRA-DBCO showed strong background fluorescence signals across the entire body. TAMRA fluorescence was also successfully detected in a transgenic line expressing MuPheRS downstream of the *gcy-8* promoter. TAMRA fluorescence was seen surrounding the anatomical site of AFD neurons, with no strong TAMARA fluorescence signals in other cells. These findings support that transgenic strains produced MuPheRS and integrated Azf into the targeted cells.

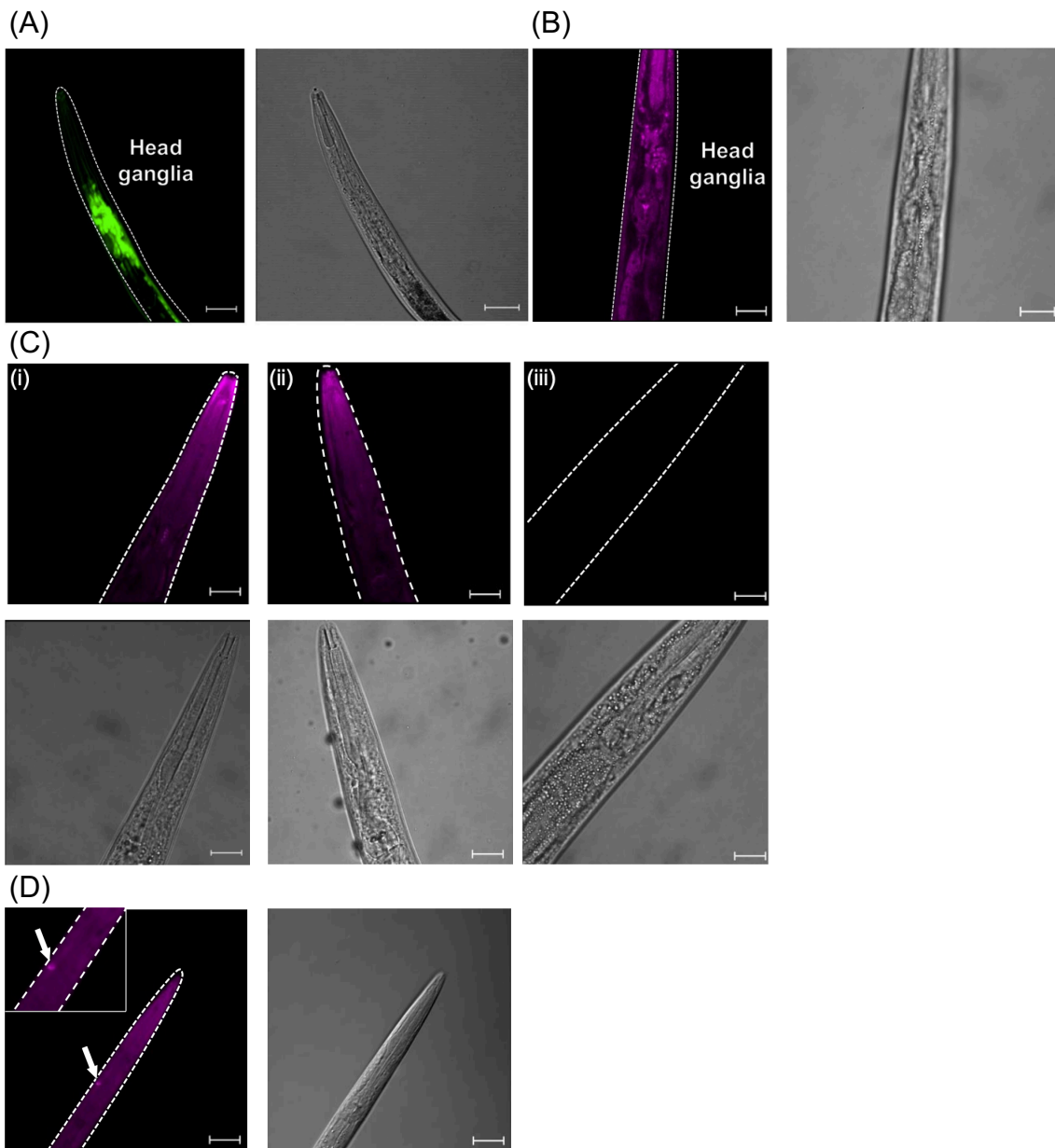


Figure 2. Verification of MuPheRS synthesis and activity in the TG1 strain. **(A)** illustrates the presence of MuPheRS production in targeted cells of the TG1 strain (TGIs1 [Prab-3::frs1(Thr412Gly)::fib-1/rps-16::gfp(S65C, synIVS)::unc-54 3'-UTR]), with green fluorescent protein expression controlled by the rab-3 promoter in all neurons. The body shape of *C. elegans* is indicated by a white dotted line, with a scale bar representing 20 μ m. **(B)** Azf incorporation in targeted cells, with azide-proteins stained using dibenzocyclooctyne-PEG4-Fluor 545 (TAMRA-DBCO). The TG1 strain was cultured with Azf-labeled *E. coli* KY33 and subsequently stained with TAMRA-DBCO, with the size of the body indicated by a scale bar representing 20 micrometers. **(C)** shows negative control experiments. (i) Wild type N2 strain cultured with Azf-labeled *E. coli* KY33 and stained with TAMRA-DBCO, (ii) TG1 strain cultured with Phe-labeled

E. coli KY33 and stained with TAMRA-DBCO, and (iii) TG1 strain cultured with Azf-labeled *E. coli* KY33 strain but not stained with TAMRA-DBCO. The size is indicated by scale bars representing 20 μm . **(D)** presents the confirmation of Azf incorporation in AFD neuronal cells of the TG2 strain cultured with Azf-labeled *E. coli* KY33, with azide-labeled protein stained using dibenzocyclooctyne-PEG4-Fluor 545 (TAMRA-DBCO). The size of the cells is indicated by a scale bar representing 20 μm , with the arrow indicating the AFD neuronal cell.

Neuron-specific proteome analysis by all neuron selective Azf incorporation

Proteomic analysis of a *C. elegans* strain expressing MuPheRS in all neurons was used to determine the molecular composition of neurons. Proteomic analysis of the N2 strain cultured with Azf- or Phe-labeled *E. coli* KY33, followed by enrichment of azide-labeled proteins with alkyne-agarose was used to assess non-specific incorporation of Azf into cells and non-specific adsorption to agarose as control experiments. Proteome analysis was performed with monolithic nano LC-MS/MS [18]. I identified 3,461 proteins in the Azf-labeled transgenic strain (average of three biological replicates). I identified 687 proteins in the N2 strain co-cultured with Phe-labeled *E. coli* KY33 and 968 proteins in the N2 strain co-cultured with Azf-labeled KY33 (**Figure 3A**). It is estimated that the number of proteins identified through non-specific adsorption to agarose was approximately 687, and the number of proteins identified through Azf incorporation into non-specific cells was approximately 281 proteins. In cell-selective BONCAT, little Azf-incorporation into non-specific cells was observed. This result further supports that the TAMRA fluorescence observed in the control experiment is caused by non-specific absorption of TMSR-DBCO rather than non-specific incorporation of Azf.

Next, proteomic samples of N2 and the transgenic strain expressing MuPheRS cultured with Azf-labeled *E. coli* KY33 without azide-labeled protein enrichment and proteomic analysis were performed. The protein compositions of the transgenic strain co-cultured with Azf-labeled KY33 with azide-protein enrichment were compared with that of N2 and transgenic strain without processing of azide-protein enrichment (**Figure 3B**). As a result of proteomic analysis, 1,397 proteins were identified only in the transgenic strain. G protein-coupled receptors (GPCRs; for example, 12 neurotransmitter receptors including SRX-29 and SRU-6) and representative neuronal proteins such as RAB-3 (Ras-related protein, *Rab-3*) were included in identified proteins [14]. Then, tissue enrichment analysis was performed to evaluate the quality of neuronal protein identification [12]. The localization of identified proteins was assessed with proteins identified only in the Azf-labeled transgenic strain (**Figure 3B**, the black-filled section). Then, the amount of detected protein was compared based on two parameters derived from protein sequence information: molecular weight and isoelectric point to check the presence of nonspecific

incorporation during the enrichment process. The results showed no bias between the enriched and non-enriched proteome analysis (**Figure 4**). These results strongly imply that our proteomic analysis enriched neuronal proteomes and successfully identified neuron-specific proteins.

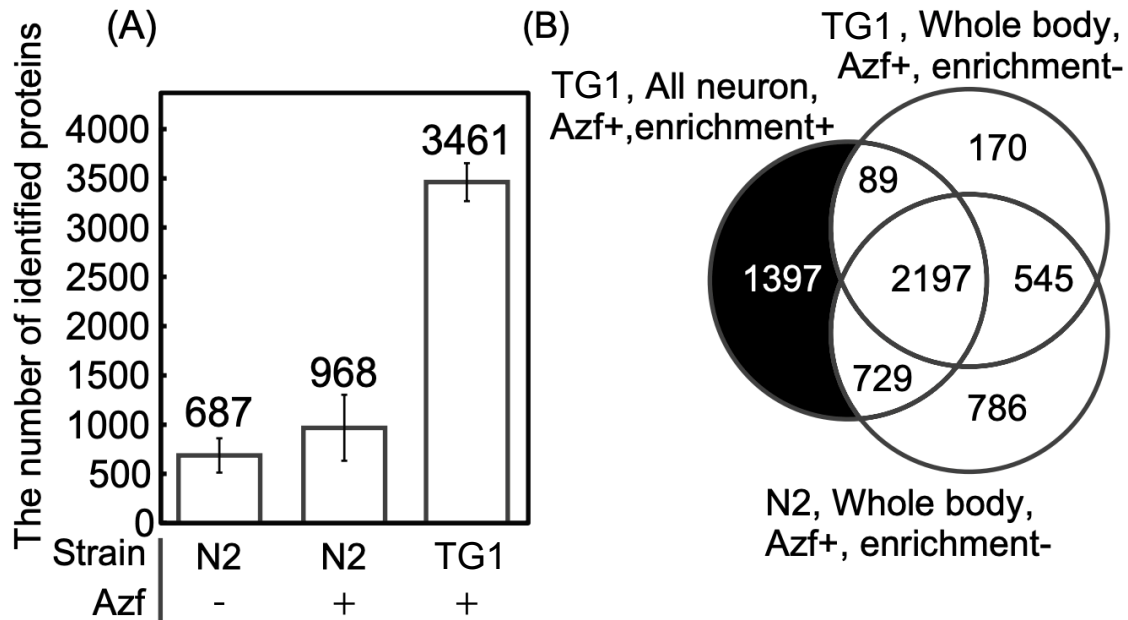


Figure 3. A comparison of the number of identified proteins. **(A)** displays the average number of identified proteins in Phe-cultured N2, Azf-cultured N2, and Azf-cultured TG1, following azide-protein enrichment, with data resented as mean \pm standard error of the mean (N=3). **(B)** presents a Venn diagram comparing the constitution of proteins identified in nematodes fed on KY33 strain cultured with Azf. The numbers of the Venn diagram represent the number of proteins found in each sample at least once (N2 or TG1). Tissue enrichment analysis was conducted on the proteins identified only in the TG1 strain cultured in the Azf-labeled KY33 strain with azide-labeled protein enrichment (the black filled area), as listed in **Table 2**.

Table 2.

Term	Enrichment Fold Change	FDR-adjusted P value
PVD	1.4	3.00E-10
outer labial sensillum	1.4	8.40E-10
IL2 neuron	1.7	2.60E-03
lateral ganglion	1.3	2.60E-03
intestine	1.1	2.60E-03
SDQL	3.2	2.60E-03
AWB	2.4	2.60E-03
AIY	2.3	2.60E-03
pharyngeal interneuron	1.4	2.60E-03
CEM	1.6	3.20E-03
ASE	1.3	3.60E-03
hook sensillum	1.6	3.60E-03
ventral nerve cord	1.3	5.30E-03
SDQR	2.7	9.80E-03
inner labial sensillum	1.5	1.10E-02

Tissue enrichment analysis was conducted on proteins detected only in the TG1 strain cultured in the Azf-labeled KY33 strain with the azide-protein enrichment procedure using WormWeb.org.

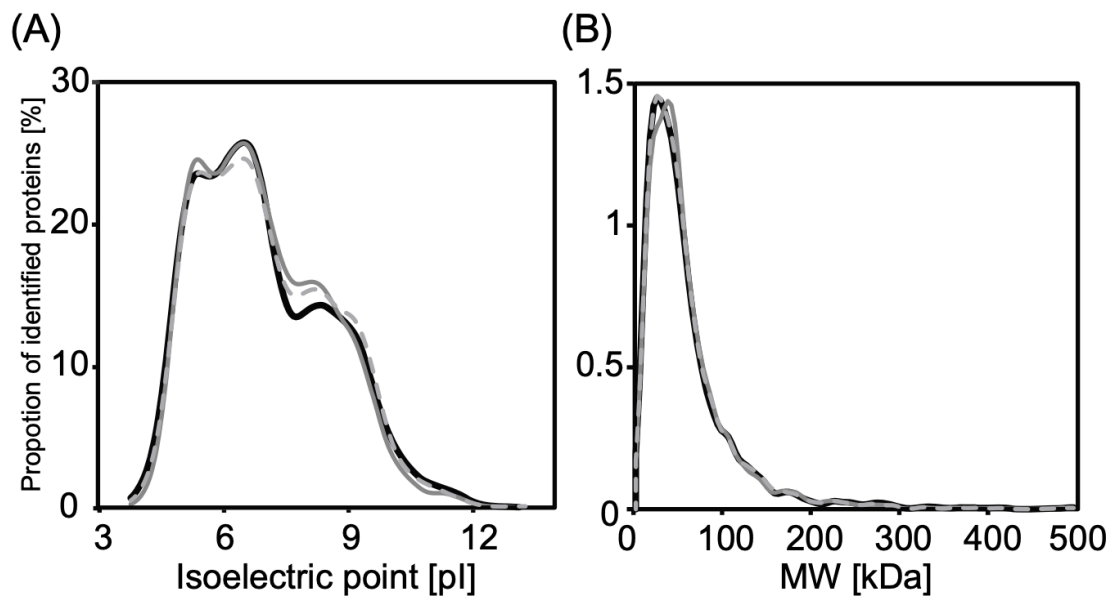


Figure 4. Distribution of identified proteins in terms of isoelectric point (A) and molecular weight (B). The lines represent the isoelectric point and molecular weight distribution of the detected proteins. The data from the N2 strain cultivated with *E. coli* KY33 that has been azide-labeled but not with further azide-protein enrichment are represented by the black line. Data from the TG1 strain cultivated with the azide-labeled KY33, either without or with azide-protein enrichment, are represented by the light grey dashed line and the dark grey line.

Confirmation of expression localization of newly identified neuron-specific proteins

In order to confirm the accuracy of our analysis, I looked for neuron-specific proteins found in our proteomic analysis that did not have an expression pattern description in Wormbase [13]. Several such proteins fall into this category (**Table 3**). To confirm that these proteins are expressed in neurons, I selected *F23B2.10* and cloned the 5' regulatory region of this gene from the *C. elegans* genome, and constructed the transgenic strain [Pf23b2.10::GFP, Pf25b3.3::mcherry]. In this strain, I confirmed that *Pf23b2.10* drives the production of GFP in neurons (**Figure 5**).

Table 3

Gene name	Accession	Description
abt-6	O01842	ABC Transporter family
B0432.14	Q3HKB5	Uncharacterized protein
B0546.3	U4PBQ7	Uncharacterized protein
C06C3.12	B7WN68	Uncharacterized protein
C12D5.5	Q17924	Uncharacterized protein
C15C7.4	Q18010	Uncharacterized protein
C18G1.7	O61918	Uncharacterized protein
C35E7.6	O61770	Uncharacterized protein
C36B1.9	Q93343	Uncharacterized protein
CELE_F23B2.10	O45400	Uncharacterized protein
CELE_F35E8.6	O45452	Uncharacterized protein
CELE_F38H12.5	O16345	Uncharacterized protein
CELE_F39C12.1	O61201	Uncharacterized protein
CELE_F41C6.6	Q20271	Uncharacterized protein
CELE_F48C1.11	G4RU19	Uncharacterized protein
CELE_F57G4.11	D9N150	Uncharacterized protein
CELE_K08D9.10	A0A0S4XR86	Uncharacterized protein
CELE_R03G8.4	Q21673	Aminopeptidase
CELE_T08E11.1	O76644	Uncharacterized protein
CELE_T26C5.5	Q7YWS4	Uncharacterized protein
CELE_W02F12.4	H1ZUX4	Uncharacterized protein
CELE_W04G5.9	O18173	Uncharacterized protein
CELE_Y26E6A.2	O62419	Uncharacterized protein
CELE_Y54G2A.42	Q4R167	Uncharacterized protein
CELE_ZC8.6	Q23083	Uncharacterized protein
clcc-82	H2L0R0	C-type LECTin
fbxb-99	Q9TYK5	F-box B protein
flp-28	Q7YWT6	FMRF-Like Peptide
gcp-2.3	Q58A98	Glutamate CarboxyPeptidase 2 homolog
glct-3	Q9XU73	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase
gpb5	A7DT38	Putative glycoprotein hormone-beta5
irl-2	Q19110	Insulin/EGF-Receptor L Domain protein
marc-6	Q400N4	MARCH (Membrane-Associated Ring finger (C3HC4)) homolog
nduo-4	P24892	NADH-ubiquinone oxidoreductase chain 4
npax-1	Q19677	N-terminal PAX (PAI domain only) protein
pals-29	O45859	Protein containing ALS2cr12 (ALS2CR12) domain
pcp-4	Q9GRV9	Prolyl Carboxy Peptidase like
pho-11	Q09451	Putative acid phosphatase 11
R07B7.12	Q21802	Glycosyltransferase family 92 protein R07B7.12
srw-66	O76620	Serpentine Receptor, class W
srx-29	O45662	Serpentine Receptor, class X
str-180	H2L036	Seven TM Receptor
Y18D10A.3	Q9XW15	NAD(P)H-hydrate epimerase

Proteins detected only in the neuron-enriched fraction without localization information in tissue enrichment analysis were further analyzed. The 5' regulatory regions of the target gene (indicated in the bold letter) were cloned from the *C. elegans* genome and used to construct promoter::GFP fusion plasmids.

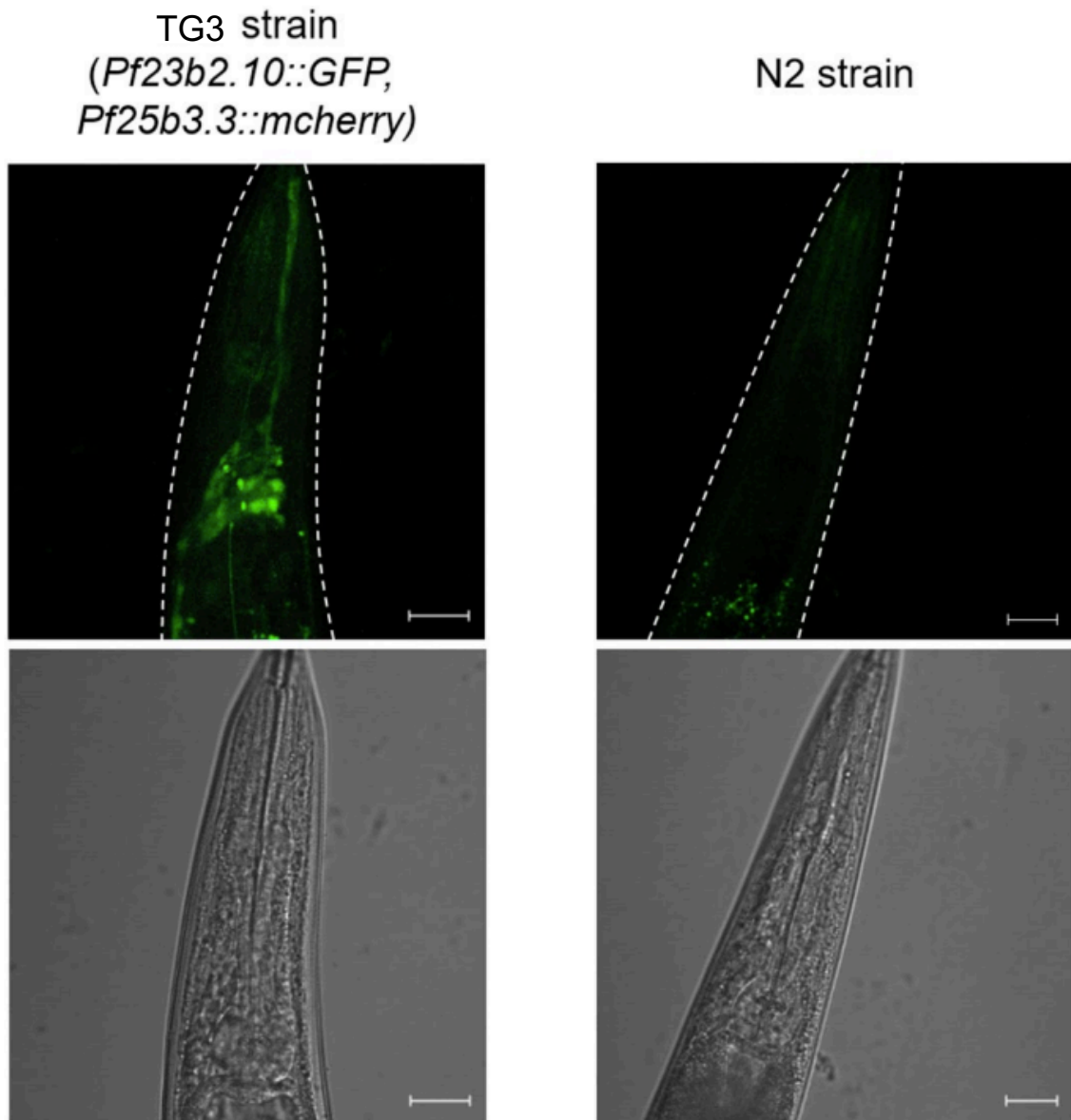


Figure 5. GFP expression is controlled by the 5' regulatory region of F23B2.10. Confocal imaging of the head ganglia was conducted in the TG3 strain (TGEx3[*Pf23b2.10::GFP*, *Pf25b3.3::mcherry*]) and N2 strain. GFP fluorescence was observed in only the subset of neurons in the TG3 strain. The scale bar indicates a length of 20 μ m.

***Gcy-8* promoter based AFD subclass neuron-selective proteomic analysis**

Using the *gcy-8* promoter, I specifically generated MuPheRS in AFD subclass neurons to conduct a single neuronal subclass-selective proteomic analysis. The molecular mechanism underlying thermosensation has yet to be fully understood [16, 17, 19, 20]. In previous research, Kobayashi and colleagues have performed a phosphoproteomic analysis of in vitro differentiated AFD subclass neuron. However, the protein composition of these cells has yet to be established. The protein composition in AFD neuronal cells may provide insights into the molecular mechanisms underlying thermotaxis in *C. elegans* [17]. Azf-labeled KY33 was used to cultivate the transgenic strain, and a proteome analysis was performed on the Azf-labeled protein.

AFD subclass neuron selective proteomic analysis revealed 1,834 proteins and identified TAX-6, which is known to be produced in AFD [22]. The proteome analysis of the transgenic strain expressing MuPheRS only in AFD was compared with that of the transgenic strain expressing MuPheRS in all neurons cultured with Azf-labeled *E. coli* KY33 (**Figure 6**). 183 proteins that were unique to the AFD neuron-enriched proteomic analysis were identified through this comparison. Additionally, 1,143 proteins were identified in the single-cell transcriptome analysis of AFD neurons among the 1,834 proteins identified in our proteomic analysis [23]. The match between single-cell transcriptome analysis and our proteome analysis suggested that I successfully enriched proteins produced in AFD subclass neurons, given that 687 proteins were identified as non-specific adsorption of proteins to alkyne-agarose (**Figure 3**).

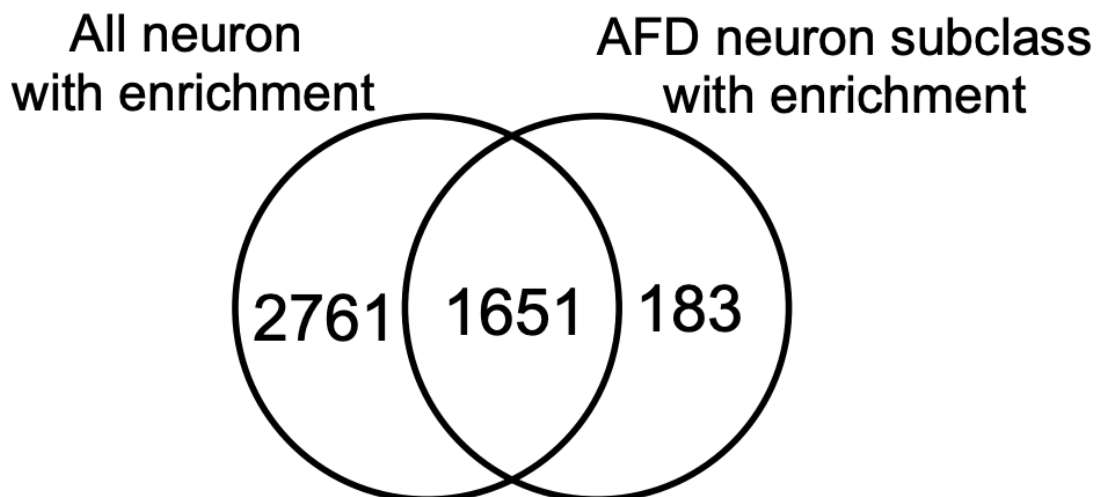


Figure 6. Venn diagram showing the comparison of the protein constitutions between azide-enriched TG1 and TG2 strains fed on the Azf-labeled KY33. The numbers in the Venn diagram represent the number of proteins that were detected in each fraction at least once.

Discussion

For the first time, I present proteomic analyses of all neurons and AFD subclass neurons using cell-selective BONCAT, identifying 4,412 and 1,834 proteins, respectively (**Figures 3 and 6**). Proteome analysis for the whole body identified 4,257 proteins, comparable to the number identified in all neurons (**Figure 3B**). The proteome coverage from all neuronal cell-selective BONCAT is as comprehensive as that of the whole-body proteome, making it sufficient for discovering new marker proteins and elucidating molecular dynamics. Among the identified proteins from all neurons or AFD subclass neurons, neuron marker proteins such as RAB-339 or TAX-643, as well as GPCRs and neuropeptides, were identified. The results of tissue enrichment analysis confirmed the accuracy of neuronal protein identification.

In addition, the localization of a novel protein encoded by F23B2.10 in neurons was also confirmed (**Figure 5**). These results indicate that protein enrichment by cell-selective BONCAT works well. Further analysis of the identified proteins, including GPCRs, is expected to provide greater insight into neuronal function. In this research, I did not employ quantitative proteomic approaches. In future studies, quantitative proteomic data will be necessary to identify proteomic variations in the target neurons. In addition, multiple promoters must be used to identify proteins specifically expressed in neurons.

In this study, the proteomic analysis could not identify all neuronal and AFD neuronal marker proteins present in trace amounts. These markers include proteins encoding GCY-8, GCY-18, GCY-27, FLP-6, NLP-7, NLP-21, and UNC-1, which are localized by a fluorescent reporter [16]. It is known that neuronal and AFD marker proteins contain some membrane proteins with high molecular weights and complex structures. Some studies have indicated that membrane proteins are not easily digested, and optimization of extract conditions and digestion may be necessary to identify membrane proteins [24]. Using more worms in the proteomic analysis may improve the identification rate.

Additionally, miniaturizing the sample preparation volume may help reduce protein loss. Sample preparation methods based on solid-phase extraction (SPE), such as miniaturized filter-aided sample preparation and in-StageTip methods, may be potentially suitable approaches as they allow for all sample processing, including cell lysis, reductive alkylation, digestion with trypsin, and elution of purified peptides, to be performed in a single enclosed reaction chamber [25, 26]. By combining solid-phase extraction with alkyne groups for sample processing, it may be possible to recover azide-labeled proteins in a single reaction chamber with minimal surface losses. A single-pot solid-phase-enhanced sample preparation may be a preferable method with high recovery efficiency [27].

In conclusion, I successfully carried out a proteomic analysis on all neurons and a

subclass of AFD neurons using cell-selective BONCAT, identifying 4,412 proteins and 1,834 proteins, respectively (**Figure 3, 6**). These results demonstrate the feasibility of performing subclass-selective proteomic analysis and have the potential to elucidate proteomic differences among neurons using subclass-selective BONCAT.

References

1. Alvarez-Castelao, B. et al. Cell-type-specific metabolic labeling of nascent proteomes in vivo. *Nat Biotechnol* 35, 1196–1201 (2017).
2. Li, J. et al. Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. *Cell* 180, 373–386.e15 (2020).
3. Yuet, K. P. et al. Cell-specific proteomic analysis in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 112, 2705–2710 (2015).
4. Ullrich, M. et al. Bio-orthogonal labeling as a tool to visualize and identify newly synthesized proteins in *Caenorhabditis elegans*. *Nature Protocols* vol. 9 2237–2255 Preprint at <https://doi.org/10.1038/nprot.2014.150> (2014).
5. Stiernagle, T. Maintenance of *C. elegans*. *WormBook: the online review of C. elegans biology* 1–11 Preprint at <https://doi.org/10.1895/wormbook.1.101.1> (2006).
6. Frøkjær-Jensen, C. et al. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 40, 1375–1383 (2008).
7. Wang, H. et al. CGAL, a temperature-robust GAL4-UAS system for *Caenorhabditis elegans*. *Nat Methods* 14, 145–148 (2017).
8. Aoki, W. et al. Cellomics approach for high-throughput functional annotation of *Caenorhabditis elegans* neural network. *Sci Rep* 8, (2018).
9. Mitani, S. Genetic regulation of mec-3 gene expression implicated in the specification of the mechanosensory neuron cell types in *Caenorhabditis elegans*. *Dev Growth Differ* 37, 551–557 (1995).
10. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* vol. 9 671–675 Preprint at <https://doi.org/10.1038/nmeth.2089> (2012).
11. Aburaya, S., Aoki, W., Kuroda, K., Minakuchi, H. & Ueda, M. Temporal proteome dynamics of *Clostridium cellulovorans* cultured with major plant cell wall polysaccharides. *BMC Microbiol* 19, (2019).
12. Angeles-Albores, D., Raymond, R. Y., Chan, J. & Sternberg, P. W. Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* 17, (2016).
13. Harris, T. W. et al. WormBase: A modern Model Organism Information Resource. *Nucleic Acids Res* 48, D762–D767 (2020).
14. Mahoney, T. R. et al. Regulation of synaptic transmission by RAB-3 and RAB-27 in *Caenorhabditis elegans*. *Mol Biol Cell* 17, 2617–2625 (2006).
15. Coates, J. C. & de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*. *Nature* 419, 925–929 (2002).

16. Inada, H. *et al.* Identification of guanylyl cyclases that function in thermosensory neurons of *Caenorhabditis elegans*. *Genetics* **172**, 2239–2252 (2006).
17. Aoki, I. & Mori, I. Molecular biology of thermosensory transduction in *C. elegans*. *Current Opinion in Neurobiology* vol. 34 117–124 Preprint at <https://doi.org/10.1016/j.conb.2015.03.011> (2015).
18. Aburaya, S., Aoki, W., Kuroda, K., Minakuchi, H. & Ueda, M. Temporal proteome dynamics of *Clostridium cellulovorans* cultured with major plant cell wall polysaccharides. *BMC Microbiol* **19**, (2019).
19. Komatsu, H., Mori, I. & Rhee, J.-S. *Mutations in a Cyclic Nucleotide-Gated Channel Lead to Abnormal Thermosensation and Chemosensation in C. elegans*. *Neuron* vol. 17 (1996).
20. Hobert, O. *et al.* Regulation of Interneuron Function in the *C. elegans* Thermoregulatory Pathway by the *ttx-3 LIM Homeobox Gene*. *Neuron* vol. 19 (1997).
21. Kobayashi, K. *et al.* Single-Cell Memory Regulates a Neural Circuit for Sensory Behavior. *Cell Rep* **14**, 11–21 (2016).
22. Kuhara, A., Inada, H., Katsura, I. & Mori, I. Negative regulation and gain control of sensory neurons by the *C. elegans* calcineurin TAX-6. *Neuron* **33**, 751–763 (2002).
23. Lockhead, D. *et al.* The tubulin repertoire of *Caenorhabditis elegans* sensory neurons and its context-dependent role in process outgrowth. *Mol Biol Cell* **27**, 3717–3728 (2016).
24. Moore, S. M., Hess, S. M. & Jorgenson, J. W. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. *J Proteome Res* **15**, 1243–1252 (2016).
25. Zhang, Z., Dubiak, K. M., Huber, P. W. & Dovichi, N. J. Miniaturized Filter-Aided Sample Preparation (MICRO-FASP) Method for High Throughput, Ultrasensitive Proteomics Sample Preparation Reveals Proteome Asymmetry in *Xenopus laevis* Embryos. *Anal Chem* **92**, 5554–5560 (2020).
26. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319–324 (2014).
27. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **10**, 757 (2014).

Acknowledgements

The author submitted this thesis to Kyoto University for a Ph.D. degree in Agriculture. The studies presented here have been carried out under the direction of Professor Mitsuyoshi Ueda, Assistant professor Kouichi Kuroda, and Professor Kenji Sugase in the Laboratory of Biomacromolecular Chemistry, Division of Applied Life Science, Graduate School of Agriculture, Kyoto University, during 2017-2023.

I want to express my most significant appreciation to Mitsuyoshi Ueda, Kouichi Kuroda, Wataru Aoki, and Kenji Sugase, who offered insightful comments, meaningful discussions, and continuous support throughout the course of my study.

I am also grateful to all my colleagues in the laboratory for wide-ranging discussion, encouragement, and support. I want to thank the financial support from the research fellowship of the Japan Society for the Promotion of Science for Young Scientists.

Yuji Yamauchi

Laboratory of Biomacromolecular Chemistry
Division of Applied Life Sciences
Graduate School of Agriculture
Kyoto University

Publications

Chapter I

Wataru Aoki, Hidenori Matsukura, Yuji Yamauchi, Haruki Yokoyama, Koichi Hasegawa, Ryoji Shinya, Mitsuyoshi Ueda
Cellomics approach for high-throughput functional annotation of *Caenorhabditis elegans* network
Scientific Reports 8:10380 (2018)

Chapter II

Yuji Yamauchi, Hidenori Matsukura, Keisuke Motone, Mitsuyoshi Ueda, Wataru Aoki
Evaluation of a library of *loxP* variants with a wide range of recombination efficiencies by Cre
PLOS One 7(10): e0276657 (2022)

Yuji Yamauchi, Hidenori Matsukura, Yusuke Shuto, Keisuke Motone, Tetsuya Kadonosono, Naoki Honda, Kenji Sugase, Mitsuyoshi Ueda, Wataru Aoki
Development of a novel sparse labeling method by machine learning-guided engineering of Cre-*lox* recombination
In preparation

Chapter III

Shunsuke Aburaya, Yuji Yamauchi, Takashi Hashimoto, Hiroyoshi Minakuchi, Wataru Aoki, Mitsuyoshi Ueda
Neuronal subclass-selective proteomic analysis in *Caenorhabditis elegans*
Scientific Reports 10:13840 (2020)