

Reviewed Preprint

v1 • March 26, 2026

Not revised

Reviewed Preprint

v2 • May 27, 2026

Revised by authors

✉ For correspondence:

me2839@columbia.eduor38@columbia.edu

Competing interests: No

competing interests declared

Funding: See [page 21](#)

Reviewing editor: Douglas

Portman, University of Rochester,
United States

© 2026, Eroglu & Hobert. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

A pilot study for whole proteome tagging in *C. elegans*

Matthew Eroglu , Oliver Hobert 

Columbia University, Department of Biological Sciences, Howard Hughes Medical Institute, New York, United States

eLife Assessment




The nematode *C. elegans* is an ideal model in which to achieve the ambitious goal of a genome-wide atlas of protein expression and localization. In this paper, the authors develop a rational and **useful** strategy for at-scale tagging of all protein coding genes with fluorescent markers, providing **solid** evidence that it could be a feasible foundation for a large-scale, community-wide project.


<https://doi.org/10.7554/eLife.110717.2.sa5>

Abstract

Tagging all proteins encoded by an animal genome with a fluorescent tag would open many windows to the discovery of unexpected patterns of protein expression and localization. To scale such an approach, it would be beneficial to introduce multiple, spectrally distinct fluorophore tags in parallel. As a first step in this direction, we undertook a pilot study in the nematode *C. elegans*, in which we set out to tag 30 different genetic loci with three different fluorophores, with 3 tags being introduced at a time. By choosing essential genes, predicted based on transcriptomics to cover a range of expression levels, we explore issues relating to disrupting gene function and visibility of tagged proteins. We demonstrate that such a tagging approach is highly efficient and indeed reveals unanticipated patterns of cellular sites of expression, as well as subcellular protein localization. We hope that this pilot study will motivate attempts to scale this tagging approach to more loci and, ultimately, the whole genome.

Introduction

Gene transcript-based expression atlases have provided invaluable insights into animal biology. The next frontier lies in generating whole animal protein expression atlases with single cell resolution that capture the cellular and subcellular specificity of protein expression and localization and visualize the dynamics of these parameters during development and during changes in internal and external states. The establishment of protein expression atlases is particularly urgent in light of the generally appreciated discordance between transcript detection – often used to infer cellular functions – and protein expression, a reflection of multifarious layers of posttranscriptional gene regulation ([Liu et al., 2016](#) ; [Schwanhäusser et al., 2011](#) ; [Vogel and Marcotte, 2012](#) 

One means to acquire protein expression and localization on a single cell level with whole proteome coverage lies in tagging all genes in the genome with a fluorescent reporter. This approach has been pioneered in yeast ([Huh et al., 2003](#) ), where chromosomally GFP-tagged strains provided the first cellular map covering ~75% of the yeast proteome, assigning localization for 70% of previously unlocalized proteins while uncovering the compartmental logic underlying protein function and interactions. Tagging proteins in multicellular organisms provides a host of new insights since the differentiation of cells into distinct types and their assembly into tissues deploys many more proteins and encompasses the establishment and dynamic control of more complex protein expression and localization patterns.

We envision the tagging of the entire proteome of a multicellular organism to be most feasible in the nematode *C. elegans*. CRISPR/Cas9 genome engineering is now routinely used to tag individual genetic loci in *C. elegans* with a fluorophore. So far, around 1,500 genes have been tagged already by individual labs (Leyhr et al., 2025 [↗](#)). With recent procedural advances (Eroglu et al., 2023 [↗](#); Ghanta and Mello, 2020 [↗](#); Paix et al., 2015 [↗](#)), such engineering is highly effective and rapid, given the short generation time of *C. elegans*. The transparency of the organism facilitates imaging of fluorophores in live animals. Advances in light-sheet fluorescence microscopy, such as dual-view inverted selective plane illumination microscopy (diSPIM), have enabled high-resolution, 3-dimensional live imaging of dynamic fluorophore localization patterns throughout all stages of development (Kumar et al., 2014 [↗](#); Wu et al., 2017 [↗](#)). Super-resolution imaging approaches such as image scanning microscopy (ISM) (Huff, 2015 [↗](#)) and stimulated emission depletion (STED) microscopy (Rankin et al., 2011 [↗](#)) enable increasingly finer localization of proteins in live worms. Further resolution can also be achieved in fixed worms with expansion microscopy, without the requirement for specialized microscopy equipment (Yu et al., 2022 [↗](#)). Extensive anatomical atlases are either already available (Altun et al., 2002 [↗](#)) or in the process of being generated (Cook et al., 2019 [↗](#); Santella et al., 2022 [↗](#); Witvliet et al., 2021 [↗](#)), allowing the mapping of fluorophore localization patterns onto anatomical features at electron microscopical scales (Vergara et al., 2021 [↗](#)).

There is also ample precedent in the *C. elegans* literature that reporter tagging results in unanticipated insights. To cite two recent examples from our own lab, we found that Golgi-resident proteins displayed cell type-specific expression patterns not anticipated by transcriptomic analysis (Leyva-Díaz et al., 2025 [↗](#)). As another example, the subcellular localization pattern of a DEG/ENaC channel in enteric neurons indicated unexpected functional insights into the protein (Bayer et al., 2025 [↗](#)). It is also notable that the widely appreciated discordance of mRNA and protein expression levels is evident in *C. elegans* (Grün et al., 2014 [↗](#); Reilly et al., 2022 [↗](#), 2020 [↗](#)), attesting to the need to extend currently existing transcriptome atlases (Gao et al., 2024 [↗](#); Ghaddar et al., 2023 [↗](#); Packer et al., 2019 [↗](#); Roux et al., 2023 [↗](#); Taylor et al., 2025 [↗](#), 2021 [↗](#)) with protein expression atlases. Individual labs have also begun to tag entire classes of extracellular matrix proteins, yielding novel insights into tissue organization and differential localization patterns of proteins within the same class (Ragle et al., 2025 [↗](#)).

Currently, around 1554 proteins representing 8% of the proteome are estimated to have been endogenously tagged (Leyhr et al., 2025 [↗](#)). However, at current rates with individual labs tagging specific proteins of interest without coordination, covering the proteome is projected to take around 100 years and, as past experience has shown, likely involve numerous duplicate attempts on a small number of commonly studied proteins (Leyhr et al., 2025 [↗](#)). It will thus be crucial for the field (a) to coordinate tagging efforts and (b) to scale up tagging protocols to enable coverage of the entire genome at a reasonable timescale and cost. Since the number of injections is a major time-limiting factor, pooling multiple injections into one would at minimum cut tagging time by a factor of 3. In *C. elegans*, screening for novel CRISPR/Cas9-induced genomic edits is already facilitated either by use of co-injection markers (i.e., plasmids that form extrachromosomal arrays) that yield phenotypes or fluorescence in progeny of successfully injected worms, or co-editing well characterized loci using established and highly efficient reagents which likewise yield visible phenotypes. In the latter approach, termed “co-CRISPR”, worms edited at the marker locus are most likely to also carry the intended edit (Arribere et al., 2014 [↗](#)); edits at F1, often being heterozygous, enable independent segregation of coCRISPR targets. Recent methods for CRISPR/Cas9 mediated genomic insertions have also pushed efficiencies to sufficient levels to simultaneously insert multiple fluorophores (e.g., mNeonGreen and mScarlet) as well as a co-CRISPR marker (*dpy-10*) at three independent loci in a single injection (Eroglu et al., 2023 [↗](#); Paix et al., 2015 [↗](#)). These attempts pooled reagents previously established to work efficiently and targeted genes that were known to yield functional fusion proteins when tagged. Thus, while in principle current methods could allow tagging of at least 3 independent loci in one injection if a co-CRISPR marker is omitted, it is not known to what extent such an approach could be generalized across the genome with previously unvalidated reagents (i.e., guides and repair template homology arms) at novel loci to yield functional tags.

As a proof of concept for scaling such a CRISPR/Cas9-engineered project to the whole genome, we report here the results of a pilot study in which we aimed to CRISPR/Cas9-engineer reporter tags into 30 genomic loci, generated through simultaneous tagging of three genes at a time. A subset of these 30 genes reveals unexpected patterns of expression and localization, illustrating how protein tagging approaches provide a route for new discoveries. We discuss here several aspects of such tagging approaches, highlighting promises, as well as limitations.

Results

Pipeline for pooled high throughput tagging by CRISPR/Cas9

Endogenously tagged genes can be detected by fluorescence stereomicroscopy if the fluorophores are expressed at sufficiently high levels and hence are readily visible as heterozygotes in F1 by visual screening without co-CRISPR markers. We reasoned that ubiquitously and highly expressed genes could be used as insertion markers and pooled with low or modestly expressed genes, with a first round of visual screening of the highly expressed genes followed by targeted PCR or higher power imaging of the lower expressed genes (Fig. 1A [↗](#)), thus eliminating the need for additional co-CRISPR markers.

To facilitate screening by fluorescence stereomicroscopy and co-imaging, we chose the three currently brightest fluorescent proteins in the blue, green and red emission spectra: mTagBFP2 (Subach et al., 2011 [↗](#)), mStayGold(J) (Ando et al., 2024 [↗](#)) and mScarlet3 (Gadella et al., 2023 [↗](#)), respectively (Fig. 1B [↗](#)). These tags are significantly brighter than first generation fluorophores, thereby allowing the detection of even very lowly expressed proteins. Autofluorescence in *C. elegans* is highest in blue and green wavelengths, and lowest in red. Furthermore, we noted that nematode growth medium (NGM) plates seeded with OP50 also contribute to background fluorescence with conventional blue and green filters on a fluorescent stereomicroscope (Fig. 1C [↗](#)). In contrast, we observed minimal background with yellow and red emission filters from worms as well as bacteria. We also noted from existing tagged strains that mStayGold could be observed brightly with YFP filters, which, combined with the minimal background, yielded a greater contrast with untagged worms enabling easier screening of mStayGold positive worms

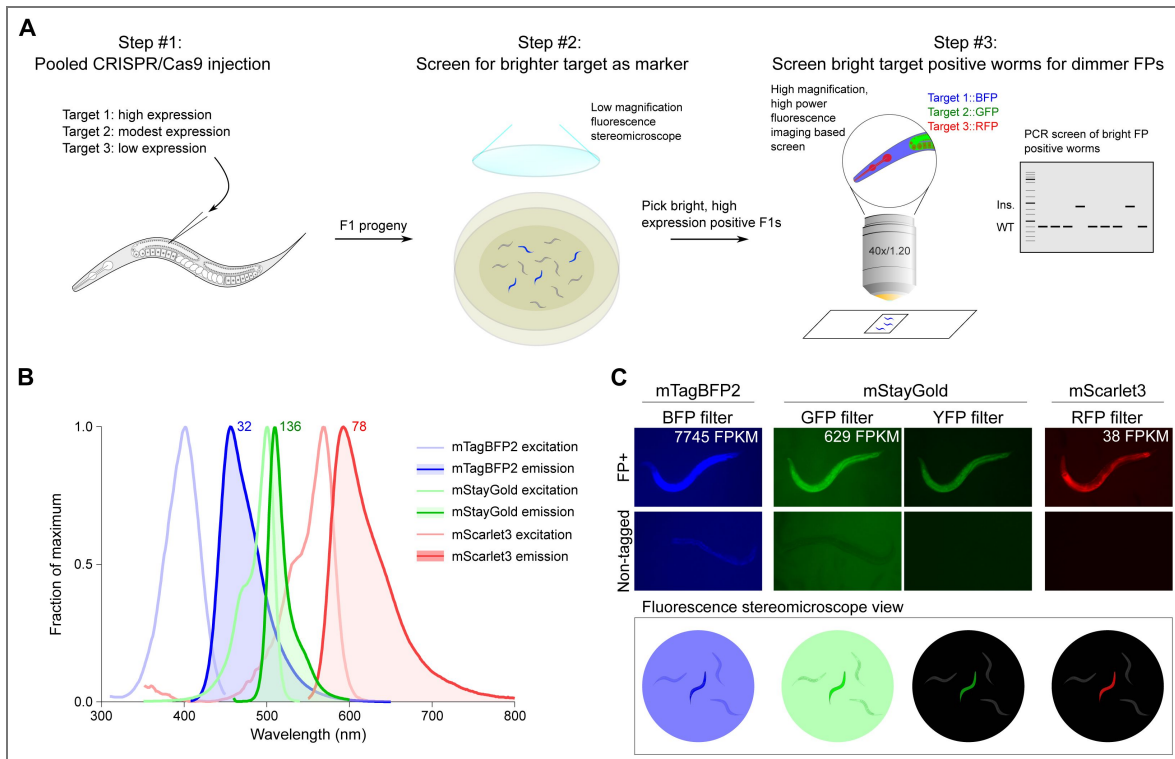


Figure 1. Pipeline for triple pooled CRISPR-Cas9 mediated gene tagging.

(A) Schematic of injection and screening strategies. FP, fluorescent protein. (B) Excitation and emission spectra of selected fluorescent proteins used in the pilot study. Theoretical brightness of the fluorophores, calculated as a product of the extinction coefficient and quantum yield, is annotated as numbers adjacent to the respective emission spectra. The spectra and brightness values were adapted from FPbase (Lambert, 2019). (C) Practical visibility of fluorophores observed under a fluorescent stereomicroscope with different filter sets in live worms on NGM plates. Top, widefield images of worms tagged with the indicated fluorophore (FP) imaged through the indicated filter contrasted to non-tagged worms and background. Bottom, schematic illustration of observed view by eye under fluorescent stereomicroscope during screening. mRNA expression of the tagged protein in young adult worms is indicated in FPKM.

compared to conventional GFP filters (Fig. 1C). However, given the suboptimal excitation and emission of visualizing mStayGold with a YFP filter, mScarlet3 in practice yielded the greatest signal-to-noise ratio for visual screening. Thus, to balance fluorescence signal-to-noise ratios with predicted gene expression, we assigned the highest expressed genes to be tagged with mTagBFP2, modestly expressed genes with mStayGold, and lowest expressed genes with mScarlet3.

In the pilot experiment, we sought to establish three major principles: (1) the endogenous expression ranges at which fluorophores could be utilized for visual screening; (2) success rate of triple pooled tagging using novel guides and targets; (3) functionality of tagged proteins. We thus selected 30 genes across a variety of bulk transcript expression ranges which are predicted to be ubiquitously expressed based on molecular function or single cell RNA sequencing (scRNA-seq) data (Table 1, Fig. 2A, B) (Taylor et al., 2021). The proteins are generally predicted to be broadly expressed and cover a very wide range of predicted cellular functions and predicted subcellular localization patterns (nuclear, cytosolic, transmembrane). 26 of 30 genes are annotated to display lethal, sterile, or other phenotypes by RNAi, enabling the assessment of tagged protein functionality.

The *C. elegans* genome encodes around 20,000 protein-coding genes (Spieth et al., 2014). Approximately 6,700 marker genes would need to be sufficiently highly expressed to be used as visible markers to group with the entire genome in sets of three. In bulk RNA-seq of N2 worms at L4 stage, 6,804 genes are expressed above 20 fragments per kilobase of transcript per million mapped reads (FPKM) (Eroglu et al., 2024; Gerstein et al., 2010; Harris et al., 2020). Based on our prior experience comparing bulk transcript FPKM with visibility, we divided these genes into three subsets to assess the detection limit of the three fluorophores: ≥ 500 FPKM for mTagBFP2; 100-500 FPKM for mStayGold; and 18-100 FPKM for mScarlet3.

We tagged genes either at their N- or C-terminus, primarily aiming to capture as many predicted isoforms as possible while retaining a potential guide RNA target sequence within 10 nucleotides of the start or stop codon. Where both termini satisfied these conditions, we selected the C-terminus over the N-terminus, as previous studies have suggested that C-terminal tags are less likely to disrupt protein localization and expression (Davidi et al., 2019; Palmer and Freeman, 2004) and that genome wide tagging attempts in yeast have successfully used C-termini (Huh et al., 2003) with high concordance to prior annotated data. However, we acknowledge that subclasses of proteins such as peroxisomal proteins, C-tail- or glycoposphatidylinositol-anchored proteins, lipidated proteins and proteins with retrieval motifs, may only show correct localization when tagged at the N-terminus (Yofe et al., 2016).

Efficient tagging across loci, expression levels and target pools

We co-injected the 30 selected targets as 10 pools (Table 1), with each pool being injected into 5 worms. Overall, we successfully isolated 8/10 mTagBFP2 tags, 9/10 mStayGold tags, and 7/10 mScarlet3 tags (Fig. 3A, B). All but one tag (*cox-6B::mTagBFP2*) was visible in the F1 generation of injected P0 animals, and these were subsequently isolated among F2 worms positive for the other tags in the pool. In 4 pools, we isolated all 3 tags within the same worm, and in an additional experiment all 3 tags but in pairs. The remaining tags were isolated as pairs or singles, and all tags could also be isolated individually or segregated separately from heterozygous animals. Even when one or two tags in the pool failed, the remaining tags in the same pool could be independently isolated. Overall, 24 of 30 tags were successfully isolated, all of which were visible with fluorescence stereomicroscopy.

During our screening, we made several observations which enabled us to assess whether failures were due to technical or biological reasons. Our first attempt with pools 4, 6 and 10 yielded no positive progeny for any of the 3 targets in the respective pools despite the injected worms showing a healthy number of viable offspring (Fig. 3C). We reasoned that, given overall efficiency for individual tags, this was due to technical failures rather than the impossibility of functionally tagging all three genes within the pool. Indeed, upon reconstitution of the injection mixes with additional quality control steps (i.e., ensuring that repair templates were of sufficient purity as assessed by proper A260/230 and A260/280 ratios; see Methods), we successfully obtained

Pool	Locus	Ontology	Tagging	Bulk RNA (FPKM)				Selected scRNA-seq top tissues				Go et al. young adult expression		Ghaderi et al. young adult expression		Observed expression
				FP	Success?	RNAi phen.	SH visible?	CHNGEN Fy01 et al. young adult expression (scaled TPMZ score)	Go et al. young adult expression (Sourat average expression)	Ghaderi et al. young adult expression (log2TPM)	Exp.	Exp.	Exp.			
1	Y2E5BR.3	III ATP synthase membrane subunit c	N	ntBFP2	No	Let	Yes	5494	7295	3002	Pharynx (marginal cell)	7.7	Spermatheca	17.9	Muscle (pharynx)	13.4
	ctb-1	III citrate synthase	C	m5G	Yes	Let	Yes	937	466	590	Muscle (body wall)	3.2	Pharynx (marginal cell)	2.8	Muscle (pharynx)	12.9
2	cat-4	X carbonic anhydrase	C	m5C3	Yes	Let	Yes	209	105	140	Epidermis	6.6	Epidermis	33	Excretory cells	10.2
	eat-1A.1	III elongation factor 1 alpha 1	C	ntBFP2	Yes	Let	Yes	3491	6059	7745	Epidermis (seam)	5.3	Sperm	90	Neurons (RB)	9.4
3	ZKX2.9	III Nuclear protein 1	N	m5G	Yes	Stop	Yes	537	561	629	Pharynx (marginal cell)	4.1	Germline	2.1	Pharynx (marginal cell)	10.3
	acdh-10	X acyl-CoA dehydrogenase	C	m5C3	Yes	Site	Yes	135	74	38	Epidermis (tail)	5.6	Sperm	1	Muscle (body wall)	6.3
4	par-5	IV tyrosine 3-monooxygenase/hydroxyphenyl 5-monooxygenase activation protein	C	ntBFP2	Yes	Let	Yes	1593	3293	1568	Somatic gonad	5.3	Germline	4.4	Neurons (AWC highest)	10.9
	Y22D74L.10	III heat shock protein family E	C	m5G	Yes	Let	Yes	426	364	375	Neurons (RB highest)	3.1	Sperm	4.0	Germline	10.5
5	hsl-1	I glucokinase	C	m5C3	Yes	Let	Yes	54	48	117	Pharynx (laccase cell)	3.4	Sperm	1.5	Excretory cells	9.3
	act-4	X actin beta	C	ntBFP2	No	Let	Yes	1895	1662	863	Neurons (DB highest)	4.0	Somatic gonad (sheath)	2.6	Various neurons (DB or V8 highest)	9.8
6	par-1	II carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase	C	m5G	Yes	Let	Yes	17	59	15	Somatic gonad (sheath)	5.9	Somatic gonad (sheath)	4.5	Somatic gonad (sheath)	9.7
	fat-2	I fatty acid synthase	C	m5C3	No	Let	Yes	32	38	43	Epidermis (seam)	7.9	Somatic gonad (sheath)	21.9	Somatic gonad (sheath)	11.2
7	cox-6B	III cytochrome c oxidase subunit 6B	C	ntBFP2	Yes	Let	Dim FL, F2 yes	932	1650	857	Pharynx (marginal cell)	3.9	Sperm	5.2	Muscle (pharynx)	11.2
	tdp-2	III tryptophan 2,3-dioxygenase	C	m5G	Yes	Gro	Yes	132	201	89	Somatic gonad (sheath)	3.4	Pharynx	4.6	Pharynx (marginal cell)	10.7
8	F54C8.1	III hydroxyacyl-CoA dehydrogenase	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory duct	8.8
	pab-1	I poly(A) binding protein	C	ntBFP2	Yes	Let	Yes	2000	1017	952	Neurons (PDB highest)	5.3	Coelomocyte	14.7	Germline	11.8
9	atf-3	IV dihydroorotase dehydrogenase	C	m5G	Yes	Let	Yes	191	200	202	Epidermis (head)	3.9	Sperm	2.2	Neurons (MA)	8.7
	atf-1	III lysine methyltransferase	C	m5C3	No	Let	Yes	33	27	58	Muscle (body wall)	3.1	Coelomocyte	1.9	Pharynx (epithelial)	8.6
10	hsp-1	IV heat shock protein family A	C	ntBFP2	Yes	Let	Yes	376	999	1377	Neurons (PQR highest)	3.7	Germline	7.3	Neurons (ASB highest)	16.5
	hsp-2	II triosephosphate isomerase	C	m5G	Yes	Yes	Yes	262	180	248	Somatic gonad (DTC)	3.5	Coelomocyte	6.9	Neurons (ASB highest)	10.2
11	fat-1	II fatty acid desaturase	C	m5C3	No	Let	Yes	18	24	30	Neurons (AWB highest)	9.7	Sperm	0.6	Excretory cell	8.8
	vha-2	III ATPase H+ transporting V0 subunit c	C	ntBFP2	Yes	Let	Yes	1565	873	799	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
12	atf-2	III histone acetyltransferase 1	C	m5C3	Yes	Let	Dim	25	157	268	Neurons (PDA highest)	6.3	Germline	1.1	Germline	6.0
	atf-3	III histone acetyltransferase 2	C	m5C3	Yes	Let	Yes	33	27	58	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
13	atf-4	III histone acetyltransferase 3	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-5	III histone acetyltransferase 4	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
14	atf-6	III histone acetyltransferase 5	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-7	III histone acetyltransferase 6	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
15	atf-8	III histone acetyltransferase 7	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-9	III histone acetyltransferase 8	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
16	atf-10	III histone acetyltransferase 9	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-11	III histone acetyltransferase 10	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
17	atf-12	III histone acetyltransferase 11	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-13	III histone acetyltransferase 12	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
18	atf-14	III histone acetyltransferase 13	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-15	III histone acetyltransferase 14	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
19	atf-16	III histone acetyltransferase 15	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-17	III histone acetyltransferase 16	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
20	atf-18	III histone acetyltransferase 17	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-19	III histone acetyltransferase 18	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
21	atf-20	III histone acetyltransferase 19	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-21	III histone acetyltransferase 20	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
22	atf-22	III histone acetyltransferase 21	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-23	III histone acetyltransferase 22	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
23	atf-24	III histone acetyltransferase 23	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-25	III histone acetyltransferase 24	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
24	atf-26	III histone acetyltransferase 25	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-27	III histone acetyltransferase 26	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
25	atf-28	III histone acetyltransferase 27	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-29	III histone acetyltransferase 28	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
26	atf-30	III histone acetyltransferase 29	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-31	III histone acetyltransferase 30	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
27	atf-32	III histone acetyltransferase 31	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-33	III histone acetyltransferase 32	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
28	atf-34	III histone acetyltransferase 33	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-35	III histone acetyltransferase 34	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
29	atf-36	III histone acetyltransferase 35	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-37	III histone acetyltransferase 36	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
30	atf-38	III histone acetyltransferase 37	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-39	III histone acetyltransferase 38	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
31	atf-40	III histone acetyltransferase 39	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-41	III histone acetyltransferase 40	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
32	atf-42	III histone acetyltransferase 41	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-43	III histone acetyltransferase 42	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
33	atf-44	III histone acetyltransferase 43	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-45	III histone acetyltransferase 44	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
34	atf-46	III histone acetyltransferase 45	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-47	III histone acetyltransferase 46	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
35	atf-48	III histone acetyltransferase 47	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-49	III histone acetyltransferase 48	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
36	atf-50	III histone acetyltransferase 49	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-51	III histone acetyltransferase 50	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
37	atf-52	III histone acetyltransferase 51	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-53	III histone acetyltransferase 52	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
38	atf-54	III histone acetyltransferase 53	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-55	III histone acetyltransferase 54	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
39	atf-56	III histone acetyltransferase 55	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis	3.3	Excretory cell	7.3
	atf-57	III histone acetyltransferase 56	C	m5C3	Yes	Let	Yes	1	34	16	Epidermis	10	Epidermis			

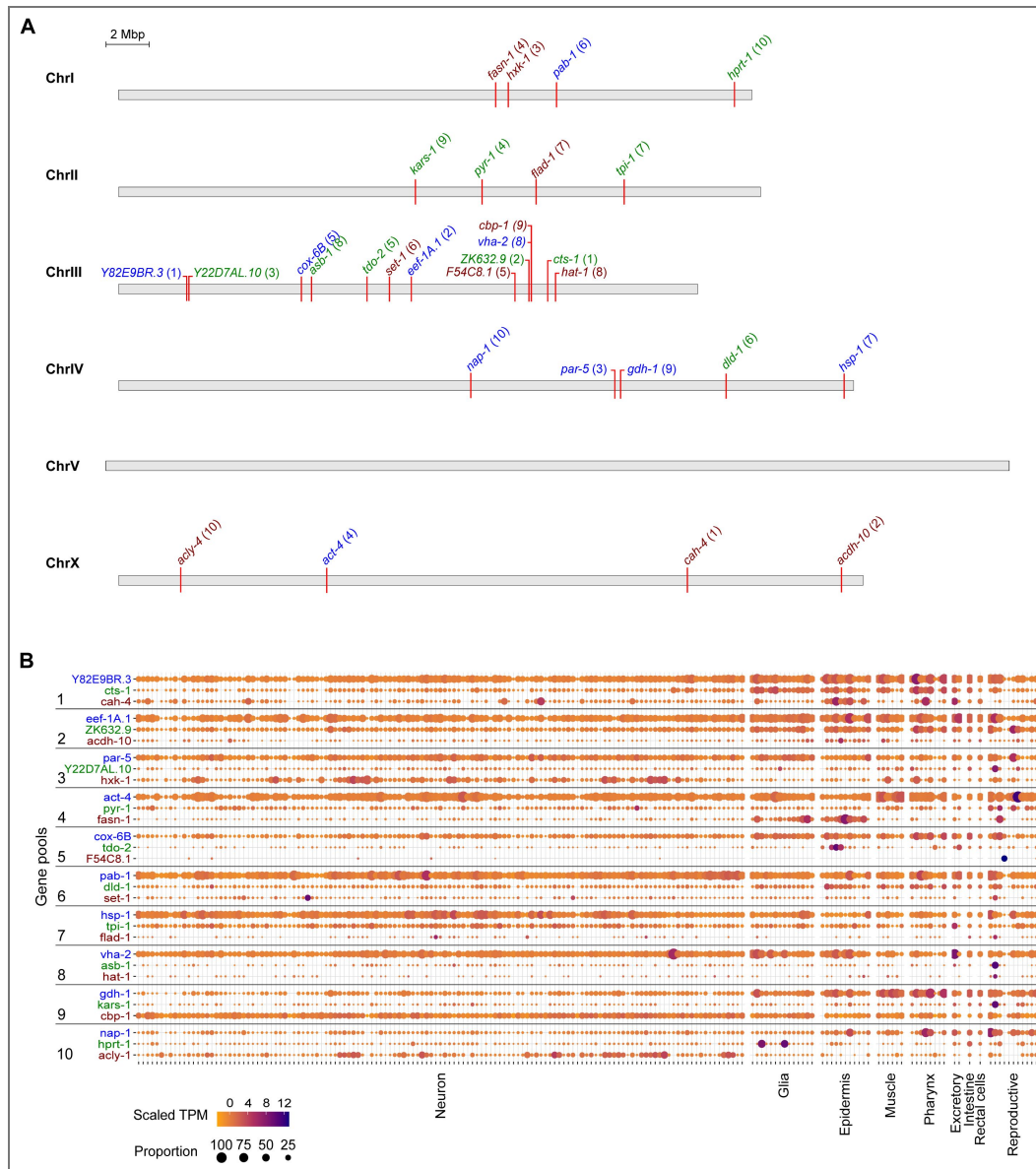


Figure 2. Selection of genes and predicted transcript expressions across tissues.

(A) Location of selected genes on *C. elegans* chromosomes. Numbers in parentheses indicate the injection pool the gene was assigned to. Color of text indicates the fluorophore assigned to the gene: blue, mTagBFP2; green, mStayGold; red, mScarlet3. (B) Transcript expression levels of the selected genes across major tissues of *C. elegans* as measured by single cell RNA sequencing of adult hermaphrodites. Data was adapted from the CeNGEN adult hermaphrodite set (Taylor et al., 2021).

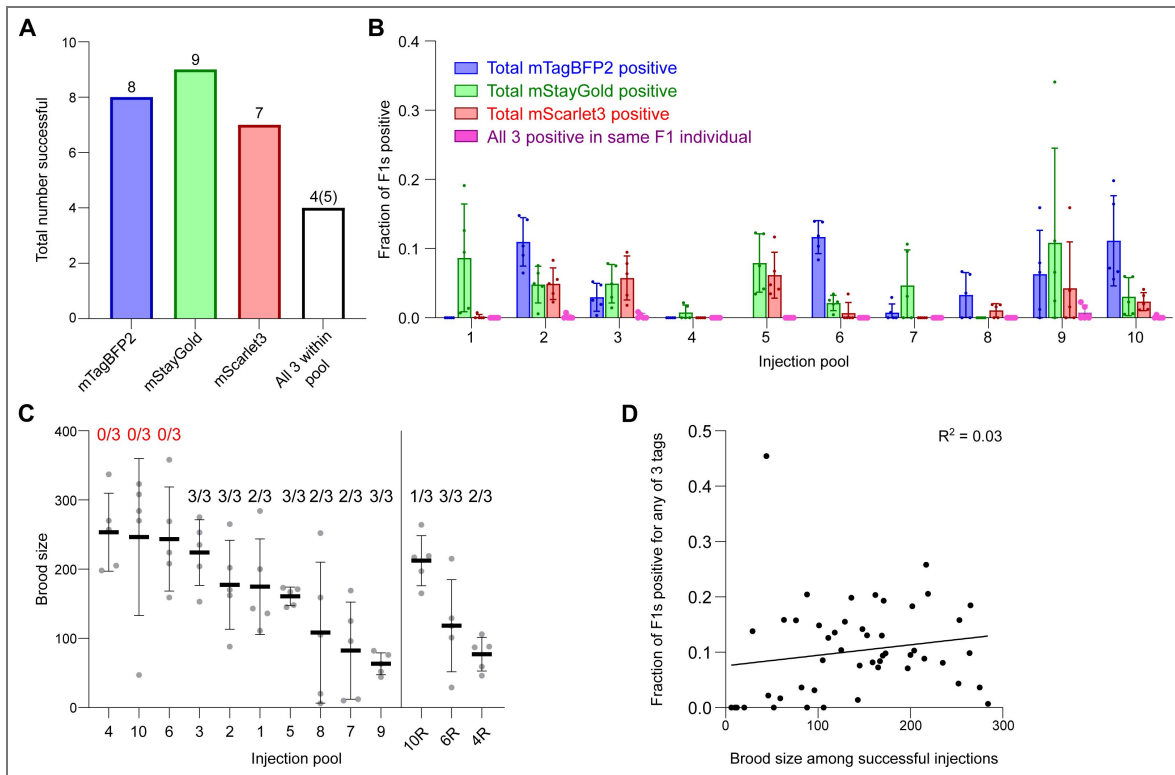


Figure 3. Successful tagging across loci, fluorophores and gene pools.

(A) Total number of targets successfully tagged. White bar denotes all 3 genes within the pool were successfully tagged in any combination. In four pools, all 3 targets were isolated in a single, triple-tagged individual worm. In one pool, we obtained all 3 targets but in doubles. (B) Total insertion efficiencies per locus for the indicated targets across injection pools. Dots, fraction of F1s positive for indicated fluorophore per injected P0. Five worms were injected per pool except for pools 4, 10 and 6 which were repeated after initial failures as shown in panel C. Data shown in panel B only represents the repeat runs and omits the original failed injections. For group 5 including *cox-6B::mTagBFP2*, worms with detectable levels of mTagBFP2 fluorescence were not recovered in the F1 generation but were isolated among progeny of F1s positive for mStayGold and mScarlet3; we were thus unable to quantify efficiency for this locus at F1. Bars, mean ± SD. (C) Viable brood size of injected P0 worms represented as individual points. Lines, mean ± SD. Numbers above plots, number of successfully isolated tags out of 3 within the pool. 4R, 6R and 10R are re-injections of groups 4, 6 and 10. (D) Correlation of viable brood of P0s with fraction of F1 worms positive for any fluorophore tags. Goodness of fit is represented as an R^2 value from a simple linear regression. Slope was not significantly greater than zero ($P = 0.2497$, F-test on Prism).

6/9 of the originally failed tags (Fig. 3C). In contrast, failures in obtaining one or two of the tags in the pool were generally interpreted as loss of function tags, and in these cases, we frequently observed fluorescent but dead embryos on the plates. While the combination of high brood size with 0/3 successful tags was a good indicator to reattempt individual pools, overall brood size across successful injections did not correlate with efficiency of insertion (Fig. 3D).

Unanticipated cell type specific enrichment of targets

We assessed the expression patterns of the protein fusions by confocal microscopy in adult worms and compared them to expected tissue type specific abundances predicted by stage-matched scRNA-seq datasets (Gao et al., 2024; Ghaddar et al., 2023; Taylor et al., 2021). Based on molecular functions in fundamental biological processes, all proteins were expected to be ubiquitously expressed. Likewise, scRNA-seq indicated that transcript abundance was ubiquitous and without strong tissue-specific enrichment with few exceptions.

In contrast, we found notable examples of mismatch between expected and observed expression patterns. The translation elongation factor *EEF-1A.1* showed strong germline enrichment whereas it was predicted to be enriched in various cell types including germ cells, epidermis, intestinal and excretory cells by scRNA-seq (Fig. 4). The acyl-CoA dehydrogenase *ACDH-10* was somatically broadly and evenly expressed but strongly depleted in the germline, while we expected enrichment in intestinal and epidermal tissues given its role in fatty acid metabolism and transcript expression in scRNA-seq datasets (Fig. 4). Strikingly, the glucokinase *HXX-1* showed strong enrichment in the gonadal sheath, whereas we expected ubiquitous expression with potential enrichment in glycolysis upregulated tissues such as muscle and neurons (Fig. 5). The lysyl-tRNA synthetase *KARS-1* showed germline enrichment, while it was expected to be evenly distributed across tissues given its role in translation, but was correctly predicted by scRNA-seq to be germline enriched (Fig. 6A-C). The nucleosome assembly protein *NAP-1* was highest expressed in the germline, whereas it was predicted to be highest in the distal tip cell and pharynx (Fig. 7A). In contrast, the phosphoribosyl transferase *HPRT-1* and ATP citrate lyase *ACLY-1* were enriched in subsets of glia and neurons, respectively, consistent with scRNA-seq datasets (Fig. 7A, B). The cytochrome c oxidase subunit 6B *COX-6B* was highest enriched in the distal gonad and intestine (Fig. 8A), while scRNA-seq predicted highest expression in the pharynx. The tryptophan 2,3-dioxygenase *TDO-2* showed highly enriched or specific expression in epidermis, whereas its expression predicted by scRNA-seq was ubiquitous with modest epidermal enrichment (Fig. 8A, B). The hydroxyacyl-CoA dehydrogenase *F54C8.1*, which we independently isolated from the same pool as *COX-6B* and *TDO-2*, showed high enrichment in sperm (Fig. 8B, C) whereas it was predicted to be enriched in the intestine based on function but correctly suggested by scRNA-seq as highly enriched in or specific to sperm. We note that the gut granule signal in *F54C8.1::mScarlet3* tagged worms was comparable to that observed in untagged N2 worms, indicating that it is background autofluorescence (Fig. 8D). The heat shock protein *HSP-1* was enriched in the germline while predicted to be ubiquitous and evenly expressed by scRNA-seq (Fig. 9). Similarly, the triosephosphate isomerase *TPI-1* was observed to be enriched in neurons, which was not predicted from scRNA-seq data (Fig. 9). Overall, numerous genes with ubiquitous and evenly distributed transcript expression showed tissue-specific enrichment not predicted by transcript levels with a general trend towards underestimation of relative protein levels in the germline.

Unanticipated subcellular localization patterns of proteins

Given that our tags were designed to be translationally fused to the proteins of interest, we assessed their subcellular localizations and compared them to localizations annotated experimentally or predicted by homology (Harris et al., 2020). Most notably, we found two proteins, *GDH-1* and *KARS-1*, detected in mitochondria by mass spectrometry (Guo et al., 2025; Li et al., 2009), and co-tagged in the same worm, to display mitochondrial-like localizations; however, they did not overlap (Fig. 6B, C). This finding suggests that subsets of mitochondria in

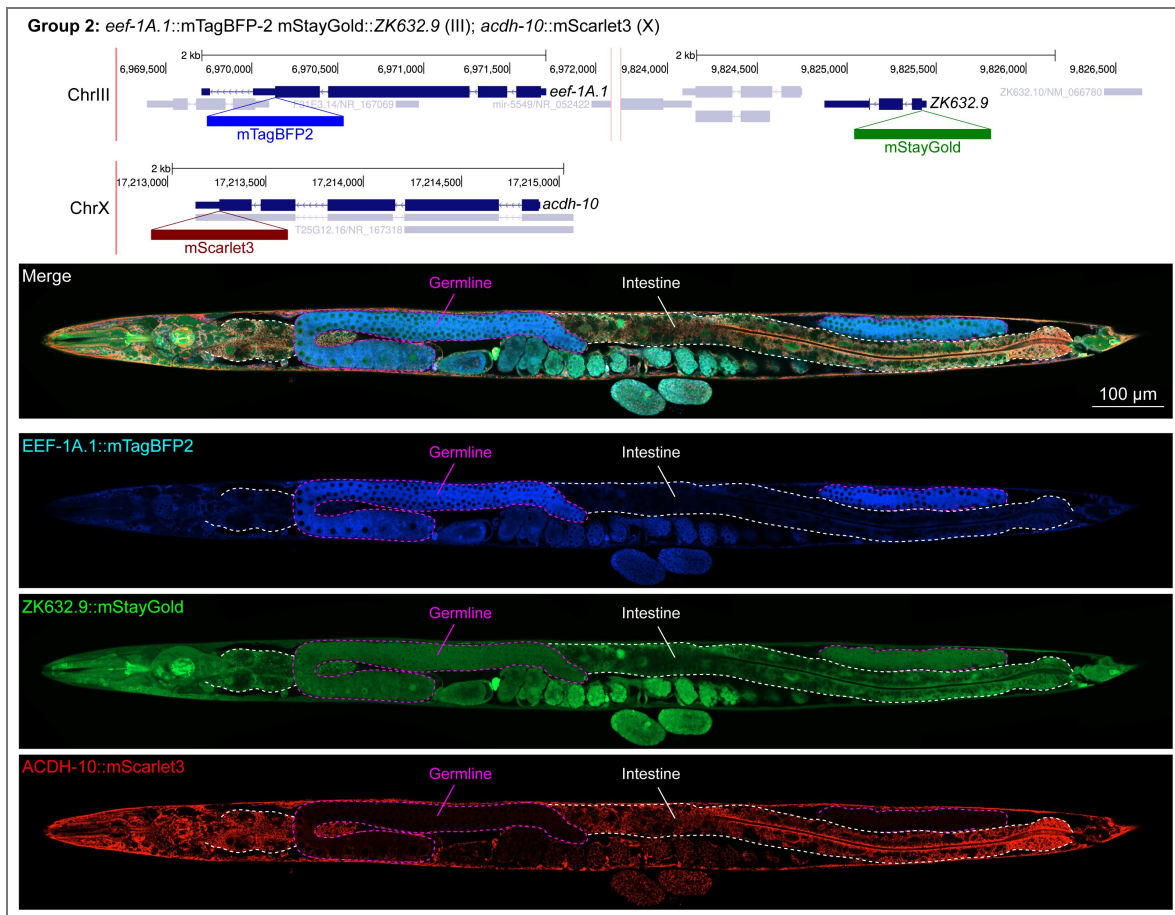


Figure 4. Expression and localization of genes in pool 2.

Expressions and localizations of translation elongation factor 1 alpha 1 *EEF-1A.1*::mTagBFP2, nuclear protein 1 homolog *ZK632.9*::mStayGold and acyl-CoA dehydrogenase *ACDH-10*::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. All three tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Germline and intestine of the animal are delineated in magenta and white dashed lines, respectively. The worm was imaged in separate overlapping frames with high magnification then stitched to reconstruct the whole animal. A single plane near the mid-plane of the animal was imaged.

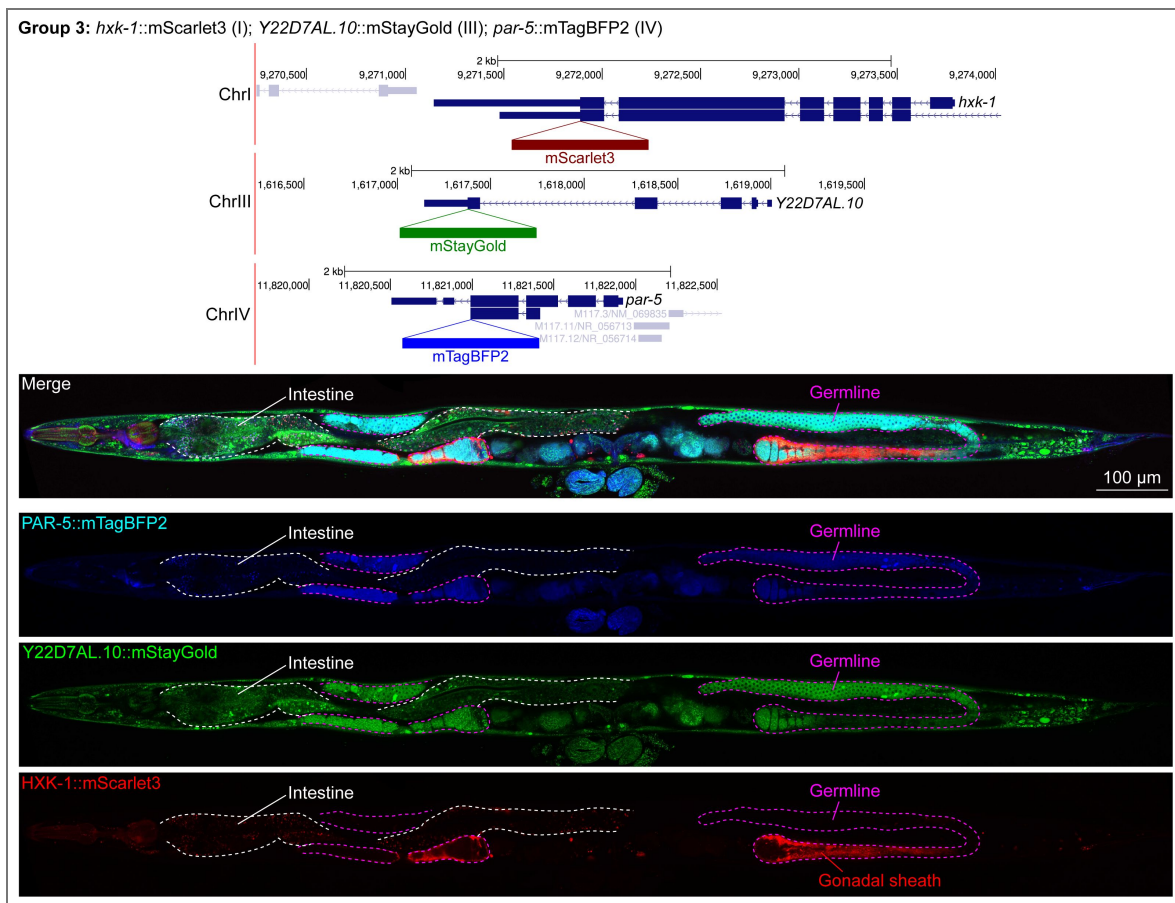


Figure 5. Expression and localization of genes in pool 3.

Expressions and localizations of 14-3-3 protein PAR-5::mTagBFP2, heat shock protein family E Y22D7AL.10::mStayGold and the glucokinase HXK-1::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. All three tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal.

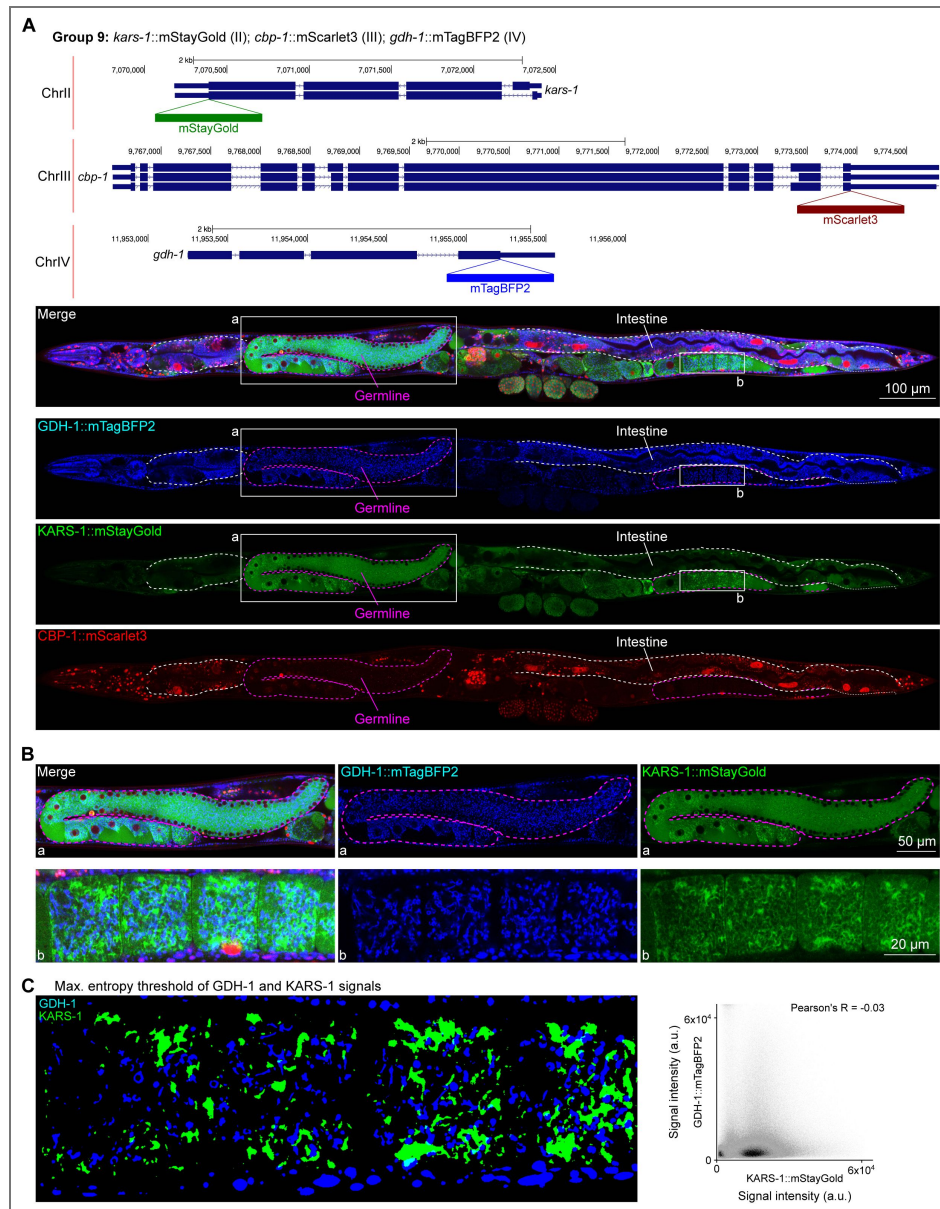


Figure 6. Expression and localization of genes in pool 9.

(A) Expressions and localizations of glutamate dehydrogenase GDH-1::mTagBFP2, lysyl-tRNA synthetase KARS-1::mStayGold and the CREB binding protein CBP-1::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. All three tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal. (B) Germline insets showing non-overlapping localizations of GDH-1 and KARS-1, predicted mitochondrial proteins. (C) Left, maximum entropy thresholds of GDH-1 and KARS-1 signals. Right, correlation plot of signal intensities of GDH-1 and KARS-1.

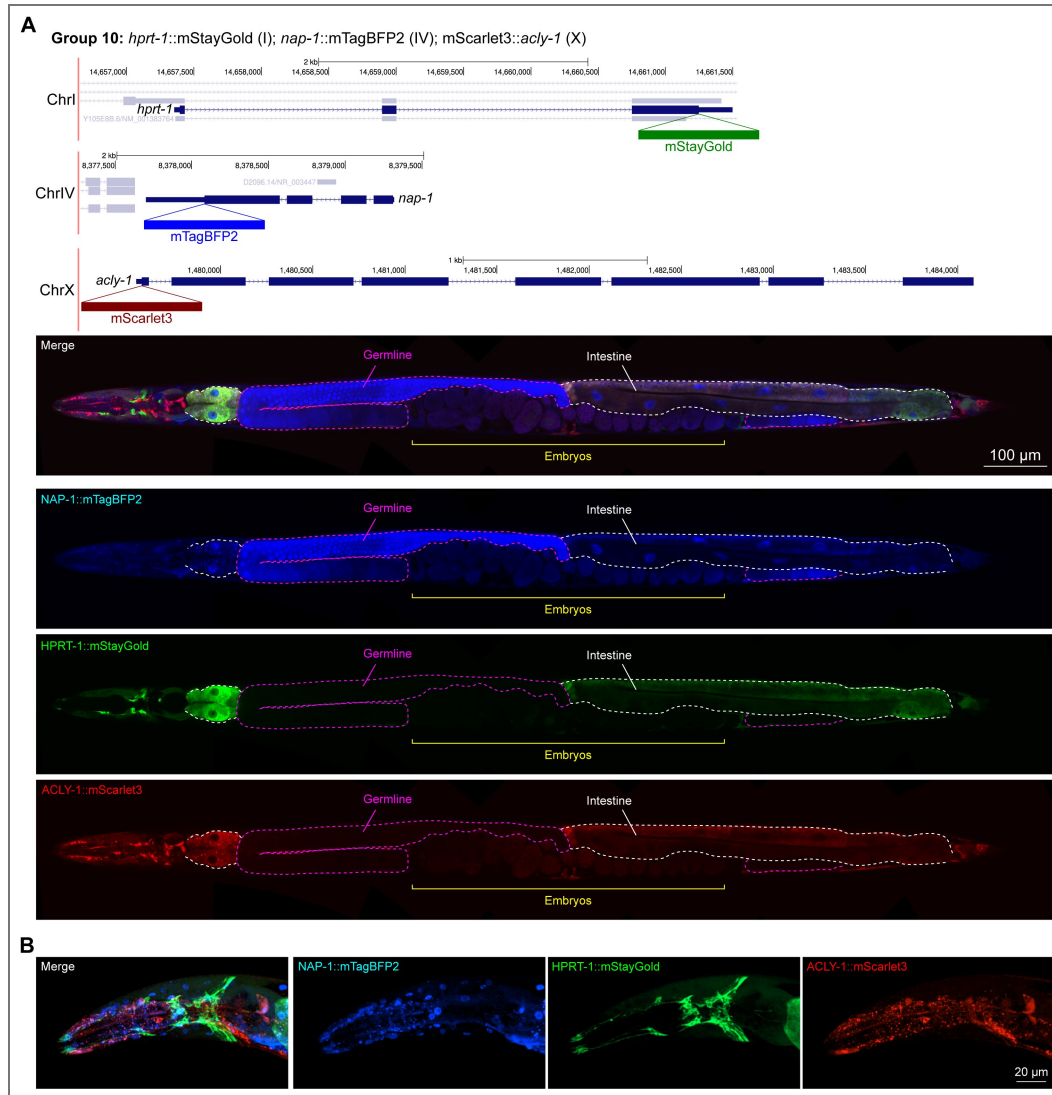


Figure 7. Expression and localization of genes in pool 10.

(A) Expressions and localizations of nucleosome assembly protein NAP-1::mTagBFP2, phosphoribosyl transferase HPRT-1::mStayGold, and ATP citrate lyase ACLY-1::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal. **(B)** Expressions and localizations of nucleosome assembly protein NAP-1::mTagBFP2, phosphoribosyl transferase HPRT-1::mStayGold, and ATP citrate lyase ACLY-1::mScarlet3 proteins in the head of a live adult hermaphrodite worm by confocal microscopy. Image shows a maximum intensity projection of a Z-stack optically sectioned at 1 μ m per stack.

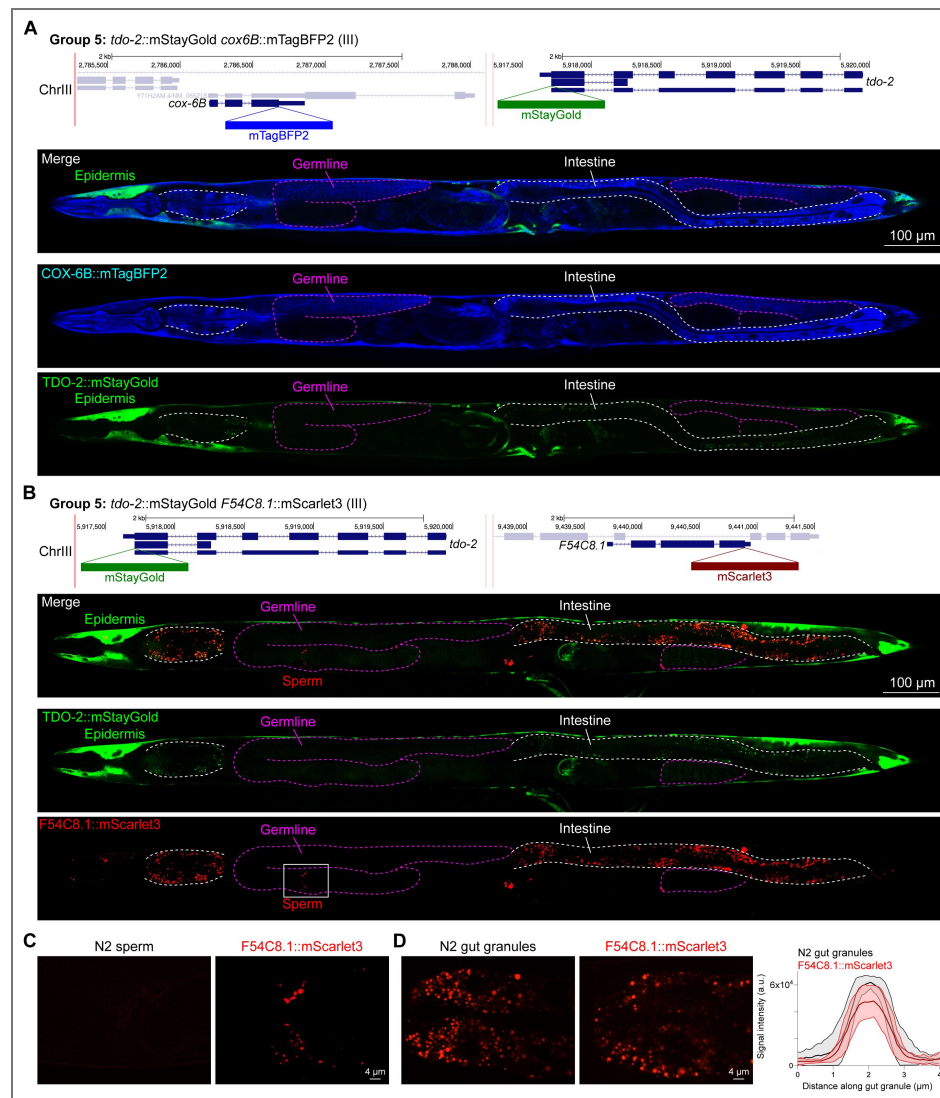


Figure 8. Expression and localization of genes in pool 5.

(A) Expressions and localizations of cytochrome C oxidase subunit 6B COX-6B::mTagBFP2 and TDO-2::mStayGold proteins in a live adult hermaphrodite worm by confocal microscopy. The two tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal. (B) Expressions and localizations of tryptophan 2,3-dioxygenase TDO-2::mStayGold and hydroxyacyl-CoA dehydrogenase F54C8.1::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. The two tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal. (C) F54C8.1::mScarlet3 signal in sperm compared to untagged N2 worms imaged under the same settings. (D) F54C8.1::mScarlet3 signal in gut granules compared to untagged N2 worms imaged under the same settings. N = 20 F54C8.1::mScarlet3 granules and 24 N2 granules quantified from 2 worms for each genotype. Lines, mean \pm SD.

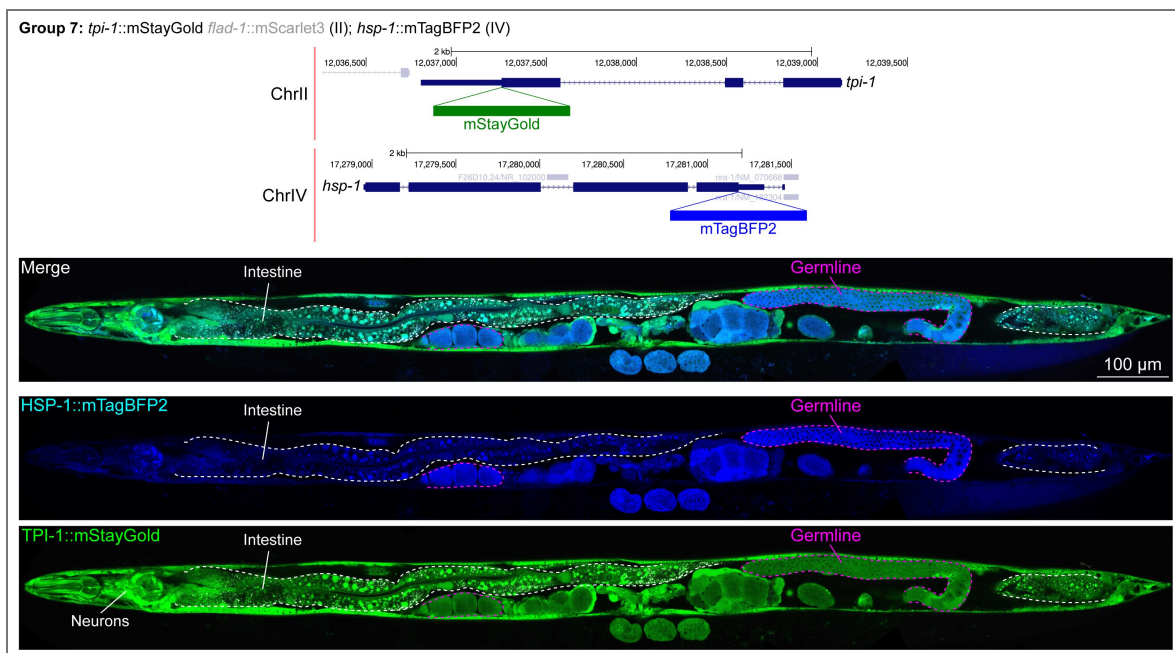


Figure 9. Expression and localization of genes in pool 7.

Expressions and localizations of heat shock protein family A HSP-1::mTagBFP2 and triosephosphate isomerase TPI-1::mStayGold proteins in a live adult hermaphrodite worm by confocal microscopy. The two tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal.

the gonad may have varying markers and metabolic specializations. Indeed, a study published around the time of our pilot experiment described metabolically distinct subpopulations of mitochondria depending on cellular energy demand (Ryu et al., 2024).

Furthermore, we found that the V-ATPase VHA-2, involved in proton transport and predicted to be localized to cell and vesicular membranes and detected in membrane fractions by mass spectrometry (Mawuenyega et al., 2003), displayed perinuclear localization in the gonad specifically (Fig. 10). This indicates that cell type-specific nuclear envelopes or endoplasmic reticula may have unique demands for pH control which are yet unexplored, to our knowledge.

In general, subcellular localizations were predicted more accurately than cell type-specific abundance, though we observed additional dynamic localization patterns unanticipated by molecular function. For instance, the DNA-binding nuclear phosphoprotein p8 homolog ZK632.9 was predicted to be nuclear but observed diffusely with modest nuclear enrichment (Fig. 4). We note that this may be due to its small size of 62 amino acids, which could be influenced by the larger 226 amino acid mStayGold tag. However, the tagged worm does not show any annotated ZK632.9 depletion phenotypes such as sterility (Simmer et al., 2003), indicating that it is to some extent functional. Additionally, the histone acetyltransferase 1 HAT-1 showed broadly nuclear localization as predicted but was cytoplasmic in early embryos (Fig. 10), indicating dynamic subcellular localization. Finally, the tryptophan 2,3-dioxygenase TDO-2 with previously unknown localization appeared cytoplasmic (Fig. 8A, B).

Discussion

We have provided proof of concept for the ease and efficiency of tagging at least three proteins at once during a single round of injection. This approach also provides the flexibility of tags being isolated together or individually, if desired. We also find that modern-day fluorophores allow the visualization of proteins at a wide range of expression levels, even at the lowest expression levels tested. Extrapolating our results to annotated transcript levels of *C. elegans* genes, we predict that at least 324 proteins could be directly visualized by mTagBFP2, 1578 with mStayGold, and 7249 with mScarlet3 for high throughput screening using low powered fluorescence stereomicroscopy. At minimum, this could enable annotation of one third of *C. elegans* proteins and at best provide sufficient markers to co-tag the entire genome in groups of 3 genes. In the latter scenario, brighter genes would be used as co-CRISPR markers grouped with dimmer genes which would be isolated among bright gene positive worms through alternative screening methods such as PCR or higher magnification, longer exposure fluorescence microscopy (Fig. 1A). Given many genes are not broadly expressed, their bulk transcript levels in whole worms are likely to underestimate expression within subsets of cells, which may aid in their detectability and isolation. Similarly, nuclear localized proteins will likely be easier to detect even at low expression levels, given the concentration of signal in small subcellular compartments. Even with the relatively small number of targets tested, we have observed numerous unanticipated patterns of tissue-specific enrichment and subcellular localization.

By tagging essential genes and in most cases observing no overt effect on viability of the animal, we have also shown that many N- or C-terminal tags do not disrupt protein function in an obvious way. Other groups targeting more specific protein classes (e.g., collagens) also observed that most tagged proteins retained function (Ragle et al., 2025). Given most genes we targeted were essential, we could infer the effect of tagging through the viability of tagged worms. We interpret failed tagging attempts as disruptions to protein function, which would necessarily prevent isolation of viable worms, which occurred in 20% of attempted tags. In our experience, inclusion of an 18 amino acid linker reduces the chances of disrupting protein function, and for all our tags we have included a universal 18 amino acid linker. However, we do fully recognize that N- or C-terminal tagging will not always produce properly localized and/or functional protein. We have shown a few examples of functional disruptions here, and it is well appreciated in the field that many types of proteins are difficult to tag without either disruption of function or localization, or both. We also recognize that in some cases where worms appear superficially wild type, tags may have more subtle effects on protein function which may impair fitness. One strategy to predict

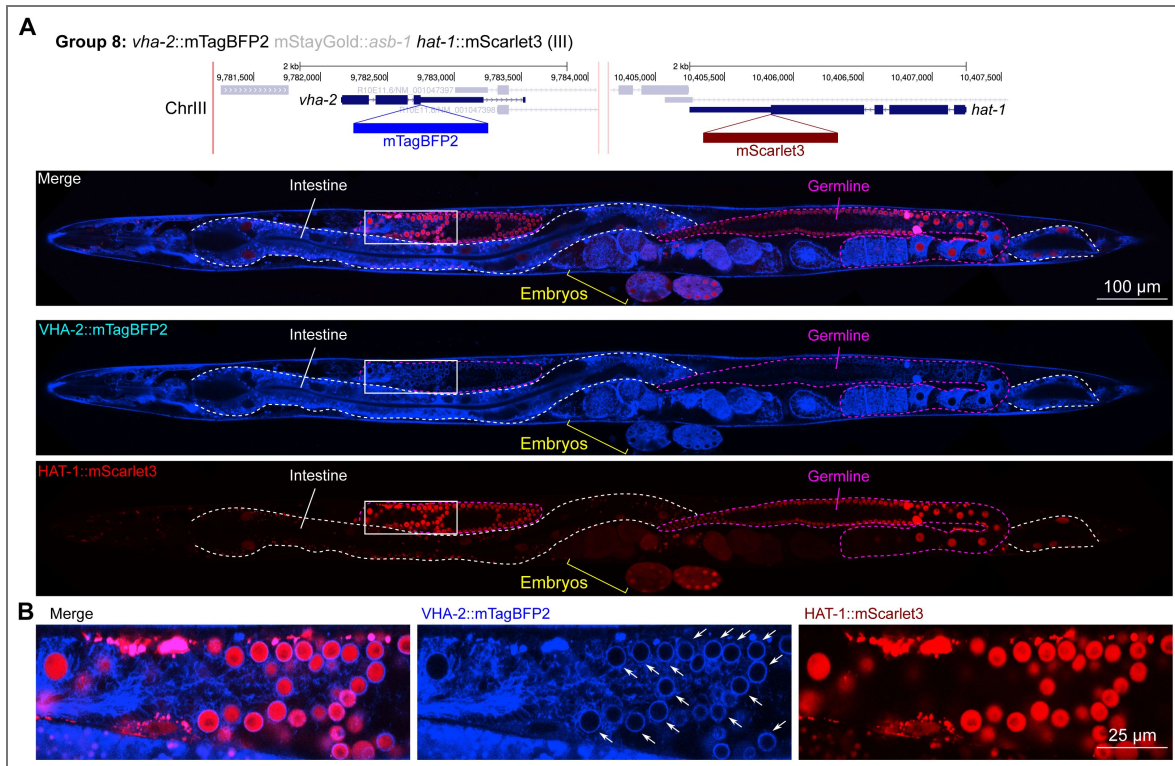


Figure 10. Expression and localization of genes in pool 8.

(A) Expressions and localizations of ATPase H⁺ transporting V0 subunit C VHA-2::mTagBFP2 and histone acetyltransferase HAT-1::mScarlet3 proteins in a live adult hermaphrodite worm by confocal microscopy. The two tags were isolated within the same individual then homozygosed at the F2 generation before imaging. Magenta and white dashed lines, germline and intestine respectively. Image reconstructed from multiple overlapping high magnification images. Single plane near the mid-plane of the animal. **(B)** Inset showing germ cell nuclei. The proton pump VHA-2 localizes to nuclear membranes in germ cells specifically, indicated by white arrows.

whether a fused protein may lose function is by comparing the predicted structure of the fusion to the original protein of interest with AlphaFold (Jumper et al., 2021 [↗](#)). However, even if the fused protein is predicted to fold normally, steric interactions, multimeric complex formation and other intrinsic properties of the protein such as solubility may still be affected and are difficult to predict based solely on structure. Smaller tags, including epitope tags like 3xFLAG, V5, and SPOT, can be expected to minimize the risk of such disruptions. However, these present their own problems. Classic epitope tags require fixation of samples, thereby preventing the ability for live imaging and introducing caveats of tissue deformation during fixation and permeabilization. The Alfa tag system that operates through stabilizing fluorescent protein-fused nanobodies (Götzke et al., 2019 [↗](#)) may, for steric reasons, carry the same risk of disruption of endogenous protein function or subcellular localization as covalent tagging. For genes with no obvious mutant phenotypes, assessing the functionality of tagged alleles will be difficult. For such genes, if an interaction-based phenotype such as synthetic lethality is annotated, the tagged gene can be crossed to the interactor to assess synthetic phenotypes. We envision that any novel or unexpected subcellular localizations observed will require validation by independent methods, as in every high-throughput experiment. Overall, the balance between throughput and functionality is, in our view, best struck with translational fusions of fluorophores to proteins of interest, which allows direct live visualization while carrying well understood limitations and artefacts.

Proteins expressed exclusively in a particular cell type and restricted to very specific subcellular sites, such as a handful of synapses in the nerve ring, pose significant challenges for accurate cell identification. Likewise, the cellular origin of secreted proteins may be difficult to trace based solely on translational reporter fusions. In such cases, the labeling of the entire cytoplasm of cells (or, alternatively, its membrane) may be useful. In theory this could relatively easily be achieved by using a double fluorophore tag, where the second fluorophore is separated by a splice leader or T2A sequence from the protein-tagging, first fluorophore. However, in practical terms, such a cassette is much larger than 1 kilobases, and in our experience larger inserts lead to a significant drop in the efficiency of successful insertion, therefore negatively affecting throughput. Where such cases are observed, a small, self-cleaving T2A peptide sequence, which frees up the tag from the fused protein, can be easily inserted in a second round of CRISPR/Cas9 to validate the cell of origin.

Fluorescent landmarks can facilitate delineation of cell boundaries, nuclei, cytoskeleton, or other organelles like mitochondria. Use of visible fluorophores such as mTagBFP2, mStayGold and mScarlet3 leaves the far-red spectrum open for the inclusion of markers such as membrane-targeted far-red fluorescent proteins like mRFP680 (Zhang et al., 2023 [↗](#)) or cell-permeant near-infrared dyes. In an ideal case, pools of target-specific mTagBFP2, mStayGold and mScarlet3 would be directly injected into integrated marker strains expressing mRFP680 targeted to landmarks of interest. Another option is to group genes strategically to include one ubiquitously expressed predicted cytoplasmic gene. With confocal or super-resolution microscopy, such impromptu landmarks may provide some indication of subcellular localization for particularly difficult proteins as a starting point.

Illustrating the discovery aspects of protein tagging, we found that two predicted mitochondrial proteins, KARS-1 and GDH-1, localized to distinct subsets of mitochondria. This is consistent with parallel reports of functionally differentiated mitochondria in cells with varying metabolic demands (Ryu et al., 2024 [↗](#)). To increase the chance of such discoveries, it may be informative to group multiple proteins predicted to localize to the same subcellular compartments together. With sufficient coverage, this can yield information on the differentiation or uniformity of organelles within the same cell. We also noted unexpected cell type dependent distributions of proteins involved in broadly important metabolic processes such as ACDH-10, which was depleted from the germline compared to other tissues, and HXK-1, which was highly enriched in the gonadal sheath. Notably, for these as well as other cases, scRNA-seq datasets were not sufficient to deduce *a priori* the observed cell type specific differences at the protein level. Importantly, many genes encoding metabolic enzymes including *acdH-10* and *hXK-1* have paralogs that likely perform similar catalytic functions. Yet, duplicate genes with identical functions are generally not evolutionarily stable

(Adler et al., 2014 [↗](#); Lynch and Conery, 2000 [↗](#)); thus such genes are likely to differ in some meaningful parameter (e.g., regulation or activity) that might align with tissue-specific functional needs. Fully annotating the expression patterns of paralogs at the protein level could indicate which tissues require unique metabolic needs and indicate which paralogous genes have undergone sub-versus neo-functionalization. For those proteins that are less functionally understood, unexpected distributions might indicate which merit further study.

We routinely inject 30 animals per day, with 5 worms per injection mix. In theory, this can enable injection of 6 pools targeting 18 genes per day, or 90 per week. A single lab can thus expect 1x coverage of the genome in approximately 290 weeks or 5.6 years, working at a modest pace, assuming that 30% of injections will need to be repeated. With simple quality control steps and careful template preparation, such as ensuring that purity of repair templates is sufficient, we project that fewer injections will need to be repeated. Notably, we find that previously untested guide RNAs and homology arms perform exceptionally well at novel loci, as we only tested one set of reagents for each locus which yielded satisfactory tagging rates. Recent advances in automated injection systems (Pan et al., 2024 [↗](#)) combined with fluorescent worm sorters may further accelerate coverage. The market cost of ordering crRNAs and annealing with tracrRNA is currently \$76 USD; the cost of each target-specific primer pair for this study was \$23 USD. Approximately \$8 USD of commercially purified Cas9 is used per injection mix; \$1.53 USD for spin columns. Without the one-time small cost of universal fluorophore template synthesis, the material cost of tagging each gene is \$109 USD and would be \$2.2 million USD for the entire genome. While the time, material and personnel costs would be daunting for an individual lab, it would be feasible for a group of committed labs within a few years. A centrally coordinated, but distributed effort would avoid wasteful duplications in gene tagging efforts that are already documented in the literature, such as PGL-1, DAF-16 and several Argonaute genes which each have been tagged at least 4 independent times with GFP (Leyhr et al., 2025 [↗](#)). The available *Caenorhabditis* Genetics Center (CGC) strain repository funded by NIH Office of Research Infrastructure Programs (P40 OD010440) as well as emerging database infrastructure dedicated to annotating endogenously tagged strains (Leyhr et al., 2025 [↗](#)) will facilitate strain as well as reagent (e.g., guide RNA sequence) cataloging and dissemination.

Materials and methods

Nematode maintenance

All *C. elegans* strains were maintained at 20 °C on nematode growth medium (NGM) agar plates with OP50 *Escherichia coli* as a food source (Brenner, 1974 [↗](#)). All CRISPR/Cas9 insertion strains were generated in the Bristol N2 background.

CRISPR/Cas9-mediated genomic insertions

Sequences of oligonucleotides, fluorophore templates and CRISPR RNAs (crRNAs) are provided in Supplementary Table 1. Cas9 protein was ordered from Integrated DNA Technologies (Park Coralville, IA, USA) (Alt-R® S.p. Cas9 Nuclease V3, 500 µg, Cat: 1081059), aliquoted at 5 mg/mL concentration, and stored at -80°C until use. tracrRNA was ordered from IDT (Alt-R® CRISPR-Cas9 tracrRNA, 20 nmol, Cat: 1072532), resuspended in the provided nuclease-free duplex buffer to 0.4 µg/µl (18 µM), aliquoted, and stored at -80°C until use (Ghanta and Mello, 2020 [↗](#)).

All guide RNAs were designed using IDT's guide RNA design tool. The highest 'on-target' scoring guide within 10 bp of the desired insert site was selected and ordered at 2 nmol scale as a custom Alt-R™ crRNA from IDT, resuspended in 20 µL nuclease-free duplex buffer (provided by IDT with tracrRNA order), and stored at 100 µM at -20°C.

Single-stranded DNA repair templates were generated as described (Eroglu et al., 2023 [↗](#)) with minor modifications. Double-stranded linear DNA templates encoding *C. elegans* codon-optimized mTagBFP2, mStayGold(J), and mScarlet3 were ordered as 'FragmentGENEs' from GENEWIZ (South Plainfield, NJ, USA) and used as universal templates. Primers encoding 45 bp gene-specific homology followed by universal priming sequences (Eurofins Genomics, 10 nmol scale) were used

to introduce homology arms to universal fluorophore templates. One primer was phosphorylated at the 5' end to enable strand-specific digestion by lambda exonuclease. Templates were amplified using Q5 polymerase from New England Biolabs (NEB; M0491) with 0.5 ng of template per reaction at an annealing temperature of 70°C and extension time of 30 seconds with 40 cycles in total. For each gene, a single 50 µL reaction was run. After amplification, individual 50 µL reactions were pooled in sets of three as indicated in Table 1 before proceeding to column purification (PureLink™ PCR Purification Kit, Invitrogen, Cat: K310001). Purified pools of double stranded DNA were digested with lambda exonuclease (NEB, cat: M0262) for 20-30 minutes at 37°C followed by micro-column purification (Monarch® Spin PCR & DNA Cleanup Kit (5 µg)), NEB, Cat: T1130) using a modified protocol described previously (Eroglu et al., 2023). For re-runs of pools 4, 6 and 10 which failed initially, we regenerated the repair templates and ensured that after each column purification, the A260/230 ratio of the purified DNA was ≥ 2.2 and A260/280 was 1.8 ± 0.05 when measured with a Nanodrop spectrophotometer.

Pooled injection mixes were set up by first annealing the crRNAs with tracrRNA as follows: 3.8 µL of tracrRNA and 0.4 µL each of 3 target-specific crRNAs were mixed, then incubated at 95 °C for 5 min followed by 10 °C for 5 min. The Cas9 RNP complex (Paix et al., 2015) was formed by incubating 2.55 µL of tracr-crRNA annealed guide pools with 0.5 µL of Cas9 for 5 min at room temperature. Following Cas9 RNP formation, 4 µL of pooled single stranded DNA template (≥ 400 ng/µL) was added. Constituted injection mixes were injected into *C. elegans* gonads using standard practice. Progeny of injected worms were screened for fluorescence under a Leica M205 FCA fluorescence stereomicroscope with a Planapo 1.6x M-series objective, pE 300 white MB light source and the following filters: Filter set ET DAPI BP - M205FA/M165FC for mTagBFP2; Filter set ET YFP - M205FA/M165FC for mStayGold; and Filter set ET mCherry - M205FA/M165FC for mScarlet3. Worms positive for fluorescence were isolated on individual plates to establish strains.

Confocal microscopy

Worms at the adult stage were immobilized with 50mM sodium azide on 5% agarose pads and imaged on a Zeiss LSM980 laser scanning confocal microscope (C-Apochromat 40x/1.20 water objective) with the 405, 488 and 561nm lasers at variable power settings. Detection was on spectral channels of the 32-channel GaAsP-PMT detector with gain at 800 V for all strains except for group 2 (Fig. 4) which was at 725V. For mTagBFP2, detection wavelengths were set to 411-481 nm; for mStayGold, detection was at 490-552 nm; for mScarlet3, detection was at 570-694 nm.

Multitracking with the LSM980 acousto-optic tunable filter (AOTF) was used to excite and detect mTagBFP2 and mScarlet3 simultaneously on the first light track (405 nm and 561 nm) while mStayGold was excited and detected on a second track (488 nm) with switching between tracks line by line. Scanning was bidirectional with 8x line averaging; pinhole was set to 31 µM (1.00 AU for mScarlet3). Worms were imaged at 1x sampling (213.13 µm x 213.13 µm at 2576 x 2576 pixels) then stitched together on ImageJ (FIJI) (Preibisch et al., 2009; Schindelin et al., 2012) to reconstitute the whole worm. Worms were linearized using the straighten function on ImageJ.

Software and statistical analysis

Genome plots showing gene loci were adapted from the University of California Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu>) (Casper et al., 2026) and represent ce11 version of the *Caenorhabditis elegans* genome. Plots were generated and statistical analyses performed on GraphPad Prism version 10.6.1.

Data availability

All strains will be made available at the CGC. Source data for Figs. 3 and 8D are provided in Supplementary Table 2.

Acknowledgements

We thank Chi Chen for support with microinjections, members of the Hobert lab, as well as Stephen Nurrish, Emily Bayer, Luisa Cochella, Geraldine Seydoux, Jean-Louis Bessereau, Hari Shroff, David Sherwood, Jake Leyhr, Brent Derry and Piali Sengupta for comments on the manuscript. Matthew Eroglu is supported by the Leon Levy Foundation and the New York Academy of Sciences through a Leon Levy Neuroscience Scholarship. This work was supported by the Howard Hughes Medical Institute.

Additional files

[Supplementary Tables.](#) 

Additional information

Funding

Funder	Grant reference number	Author
Leon Levy Foundation (LLF)		Matthew Eroglu
Howard Hughes Medical Institute (HHMI)		Oliver Hobert

Author ORCID iDs

Matthew Eroglu:  <https://orcid.org/0000-0003-2679-1376>

Oliver Hobert:  <https://orcid.org/0000-0002-7634-2854>

References

- Adler M**, Anjum M, Berg OG, Andersson DI, Sandegren L (2014) High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Molecular Biology and Evolution* **31**:1526-1535 <https://doi.org/10.1093/molbev/msu111> | [PubMed](#)
- Altun ZF**, Herndon LA, Wolkow CA, Crocker C, Lints R, Hall DH (2002) WormAtlas.
- Ando R**, Shimozono S, Ago H, Takagi M, Sugiyama M, Kurokawa H, Hirano M, Niino Y, Ueno G, Ishidate F, et al. (2024) StayGold variants for molecular fusion and membrane-targeting applications. *Nature Methods* **21**:648-656 <https://doi.org/10.1038/s41592-023-02085-6> | [PubMed](#)
- Arribere JA**, Bell RT, Fu BXH, Artiles KL, Hartman PS, Fire AZ (2014) Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**:837-846 <https://doi.org/10.1534/genetics.114.169730> | [PubMed](#)
- Bayer EA**, Mango SE, Hobert O, Schier AF (2025) The DEG/ENaC channel DEGT-1 is a proprioceptor of *C. elegans* foregut movement. *Current biology: CB* **35**:5332-5343.e5. <https://doi.org/10.1016/j.cub.2025.09.065> | [PubMed](#)
- Brenner S** (1974) The genetics of *Caenorhabditis elegans*. *Genetics* **77**:71-94 <https://doi.org/10.1093/genetics/77.1.71> | [PubMed](#)
- Casper J**, Speir ML, Raney BJ, Perez G, Nassar LR, Lee CM, Hinrichs AS, Gonzalez JN, Fischer C, Diekhans M, et al. (2026) The UCSC Genome Browser database: 2026 update. *Nucleic Acids Research* **54**:D1331-D1335 <https://doi.org/10.1093/nar/gkaf1250> | [PubMed](#)
- Cook SJ**, Jarrell TA, Brittin CA, Wang Y, Bloniarz AE, Yakovlev MA, Nguyen KCQ, Tang LT-H, Bayer EA, Duerr JS, et al. (2019) Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**:63-71 <https://doi.org/10.1038/s41586-019-1352-7> | [PubMed](#)
- Davidi D**, Weill U, Krieger G, Avihou Z, Milo R, Schuldiner M (2019) Assessment of GFP tag position on protein localization and growth fitness in yeast. *Journal of Molecular Biology* **431**:636-641 <https://doi.org/10.1016/j.jmb.2018.12.004> | [PubMed](#)

- Eroglu M, Yu B, Derry WB** (2023) Efficient CRISPR/Cas9 mediated large insertions using long single-stranded oligonucleotide donors in *C. elegans*. *The FEBS journal* <https://doi.org/10.1111/febs.16876> | PubMed
- Eroglu M, Zocher T, McAuley J, Webster R, Xiao MZX, Yu B, Mok C, Derry WB** (2024) Noncanonical inheritance of phenotypic information by protein amyloids. *Nature Cell Biology* 1-13 <https://doi.org/10.1038/s41556-024-01494-9> | PubMed
- Gadella TWJ, van Weeren L, Stouthamer J, Hink MA, Wolters AHG, Giepmans BNG, Aumonier S, Dupuy J, Royant A** (2023) mScarlet3: a brilliant and fast-maturing red fluorescent protein. *Nature Methods* **20**:541-545 <https://doi.org/10.1038/s41592-023-01809-y> | PubMed
- Gao SM, Qi Y, Zhang Q, Guan Y, Lee Y-T, Ding L, Wang L, Mohammed AS, Li H, Fu Y, et al.** (2024) Aging atlas reveals cell-type-specific effects of pro-longevity strategies. *Nature Aging* **4**:998-1013 <https://doi.org/10.1038/s43587-024-00631-1> | PubMed
- Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al.** (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**:1775-1787 <https://doi.org/10.1126/science.1196914> | PubMed
- Ghaddar A, Armingol E, Huynh C, Gevirtzman L, Lewis NE, Waterston R, O'Rourke EJ** (2023) Whole-body gene expression atlas of an adult metazoan. *Science Advances* **9**:eadg0506 <https://doi.org/10.1126/sciadv.adg0506> | PubMed
- Ghanta KS, Mello CC** (2020) Melting dsDNA Donor Molecules Greatly Improves Precision Genome Editing in *Caenorhabditis elegans*. *Genetics* **216**:643-650 <https://doi.org/10.1534/genetics.120.303564> | PubMed
- Götzke H, Kilisch M, Martínez-Carranza M, Sograte-Idrissi S, Rajavel A, Schlichthaerle T, Engels N, Jungmann R, Stenmark P, Opazo F, et al.** (2019) The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nature Communications* **10**:4403 <https://doi.org/10.1038/s41467-019-12301-7> | PubMed
- Grün D, Kirchner M, Thierfelder N, Stoeckius M, Selbach M, Rajewsky N** (2014) Conservation of mRNA and protein expression during development of *C. elegans*. *Cell Reports* **6**:565-577 <https://doi.org/10.1016/j.celrep.2014.01.001> | PubMed
- Guo X, Lu J, Miao L, Shen E** (2025) Mitochondrial Proteome Reveals Metabolic Tuning by Restricted Insulin Signaling to Promote Longevity in *Caenorhabditis elegans*. *Biology* **14**:279 <https://doi.org/10.3390/biology14030279> | PubMed
- Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R, et al.** (2020) WormBase: a modern Model Organism Information Resource. *Nucleic Acids Research* **48**:D762-D767 <https://doi.org/10.1093/nar/gkz920> | PubMed
- Huff J** (2015) The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. *Nature Methods* **12**:i-ii <https://doi.org/10.1038/nmeth.f.388>
- Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK** (2003) Global analysis of protein localization in budding yeast. *Nature* **425**:686-691 <https://doi.org/10.1038/nature02026> | PubMed
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al.** (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583-589 <https://doi.org/10.1038/s41586-021-03819-2> | PubMed
- Kumar A, Wu Y, Christensen R, Chandris P, Gandler W, McCreedy E, Bokinsky A, Colón-Ramos DA, Bao Z, McAuliffe M, et al.** (2014) Dual-view plane illumination microscopy for rapid and spatially isotropic imaging. *Nature Protocols* **9**:2555-2573 <https://doi.org/10.1038/nprot.2014.172> | PubMed
- Lambert TJ** (2019) FPbase: a community-editable fluorescent protein database. *Nature Methods* **16**:277-278 <https://doi.org/10.1038/s41592-019-0352-8> | PubMed

- Leyhr J, Chi Q, Zeng L, Shen E-Z, Zou W, Sherwood D (2025) WormTagDB: A Systematic Survey of Endogenously Tagged Proteins in *C. elegans* and Roadmap Towards the Tagged Proteome. *bioRxiv* <https://doi.org/10.64898/2025.12.02.691955> | PubMed
- Leyva-Díaz E, Cesar M, Pe K, Jordá-Llorens JI, Valdivia J, Hobert O (2025) Alternative splicing controls pan-neuronal homeobox gene expression. *Genes & Development* **39**:209-220 <https://doi.org/10.1101/gad.352184.124> | PubMed
- Li J, Cai T, Wu P, Cui Z, Chen X, Hou J, Xie Z, Xue P, Shi L, Liu P, *et al.* (2009) Proteomic analysis of mitochondria from *Caenorhabditis elegans*. *Proteomics* **9**:4539-4553 <https://doi.org/10.1002/pmic.200900101> | PubMed
- Liu Y, Beyer A, Aebersold R (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**:535-550 <https://doi.org/10.1016/j.cell.2016.03.014> | PubMed
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151-1155 <https://doi.org/10.1126/science.290.5494.1151> | PubMed
- Mawuenyega KG, Kaji H, Yamuchi Y, Shinkawa T, Saito H, Taoka M, Takahashi N, Isobe T (2003) Large-scale identification of *Caenorhabditis elegans* proteins by multidimensional liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research* **2**:23-35 <https://doi.org/10.1021/pr025551y> | PubMed
- Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, Stefanik D, Tan K, Trapnell C, Kim J, *et al.* (2019) A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* **365**:eaax1971 <https://doi.org/10.1126/science.aax1971> | PubMed
- Paix A, Folkmann A, Rasoloson D, Seydoux G (2015) . High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201**:47-54 <https://doi.org/10.1534/genetics.115.179382> | PubMed
- Palmer E, Freeman T (2004) Investigation Into the use of C- and N-terminal GFP Fusion Proteins for Subcellular Localization Studies Using Reverse Transfection Microarrays. *Comparative and Functional Genomics* **5**:342-353 <https://doi.org/10.1002/cfg.405> | PubMed
- Pan P, Zoberman M, Zhang P, Premachandran S, Bhatnagar S, Pilaka-Akella PP, Sun W, Li C, Martin C, Xu P, *et al.* (2024) Robotic microinjection enables large-scale transgenic studies of *Caenorhabditis elegans*. *Nature Communications* **15**:8848 <https://doi.org/10.1038/s41467-024-53108-5> | PubMed
- Preibisch S, Saalfeld S, Tomancak P (2009) Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**:1463-1465 <https://doi.org/10.1093/bioinformatics/btp184> | PubMed
- Ragle JM, Pooranachithra M, Ashley GE, Cadena E, Blank B, Kang K, Chen C, Bhowmick AR, Mercado SH, Wells TE, *et al.* (2025) Multiscale patterning of a model apical extracellular matrix revealed by systematic endogenous protein tagging. *bioRxiv* <https://doi.org/10.1101/2025.05.14.653803> | PubMed
- Rankin BR, Moneron G, Wurm CA, Nelson JC, Walter A, Schwarzer D, Schroeder J, Colón-Ramos DA, Hell SW (2011) Nanoscopy in a Living Multicellular Organism Expressing GFP. *Biophysical Journal* **100**:L63-L65 <https://doi.org/10.1016/j.bpj.2011.05.020> | PubMed
- Reilly MB, Cros C, Varol E, Yemini E, Hobert O (2020) Unique homeobox codes delineate all the neuron classes of *C. elegans*. *Nature* **584**:595-601 <https://doi.org/10.1038/s41586-020-2618-9> | PubMed
- Reilly MB, Tekieli T, Cros C, Aguilar GR, Lao J, Toker IA, Vidal B, Leyva-Díaz E, Bhattacharya A, Cook SJ, *et al.* (2022) Widespread employment of conserved *C. elegans* homeobox genes in neuronal identity specification. *PLoS genetics* **18**:e1010372 <https://doi.org/10.1371/journal.pgen.1010372> | PubMed
- Roux AE, Yuan H, Podshivalova K, Hendrickson D, Kerr R, Kenyon C, Kelley D (2023) Individual cell types in *C. elegans* age differently and activate distinct cell-protective responses. *Cell Reports* **42**:112902 <https://doi.org/10.1016/j.celrep.2023.112902> | PubMed
- Ryu KW, Fung TS, Baker DC, Saoi M, Park J, Febres-Aldana CA, Aly RG, Cui R, Sharma A, Fu Y, *et al.* (2024) Cellular ATP demand creates metabolically distinct subpopulations of mitochondria. *Nature* **635**:746-754 <https://doi.org/10.1038/s41586-024-08146-w> | PubMed

- Santella A, Kolotuev I, Kizilyaprak C, Bao Z (2022) Cross-modality synthesis of EM time series and live fluorescence imaging. *eLife* **11**:e77918 <https://doi.org/10.7554/eLife.77918> | PubMed
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**:676-682 <https://doi.org/10.1038/nmeth.2019> | PubMed
- Schwahnhäuser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473**:337-342 <https://doi.org/10.1038/nature10098> | PubMed
- Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PVE, Kamath RS, Fraser AG, Ahringer J, Plasterk RHA (2003) Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS biology* **1**:E12 <https://doi.org/10.1371/journal.pbio.0000012> | PubMed
- Spieth J, Lawson D, Davis P, Williams G, Howe K (2014) Overview of gene structure in *C. elegans*. *WormBook: The Online Review of C. Elegans Biology* 1-18 <https://doi.org/10.1895/wormbook.1.65.2> | PubMed
- Subach OM, Cranfill PJ, Davidson MW, Verkhusha VV (2011) An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLoS One* **6**:e28674 <https://doi.org/10.1371/journal.pone.0028674> | PubMed
- Taylor SR, Olson C, Ripoll-Sanchez L, Valperga G, McWhirter R, Barney ST, Atkinson A, Goel S, Weinreb A, Hardin A, *et al.* (2025) A gene expression atlas of a juvenile nervous system. *bioRxiv* <https://doi.org/10.1101/2025.11.21.689793> | PubMed
- Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, *et al.* (2021) Molecular topography of an entire nervous system. *Cell* **184**:4329-4347.e23. <https://doi.org/10.1016/j.cell.2021.06.023> | PubMed
- Vergara HM, Pape C, Meechan KI, Zinchenko V, Genoud C, Wanner AA, Mutemi KN, Titze B, Templin RM, Bertucci PY, *et al.* (2021) Whole-body integration of gene expression and single-cell morphology. *Cell* **184**:4819-4837.e22. <https://doi.org/10.1016/j.cell.2021.07.017> | PubMed
- Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**:227-232 <https://doi.org/10.1038/nrg3185> | PubMed
- Witvliet D, Mulcahy B, Mitchell JK, Meirovitch Y, Berger DR, Wu Y, Liu Y, Koh WX, Parvathala R, Holmyard D, *et al.* (2021) Connectomes across development reveal principles of brain maturation. *Nature* **596**:257-261 <https://doi.org/10.1038/s41586-021-03778-8> | PubMed
- Wu Y, Kumar A, Smith C, Ardiel E, Chandris P, Christensen R, Rey-Suarez I, Guo M, Vishwasrao HD, Chen J, *et al.* (2017) Reflective imaging improves spatiotemporal resolution and collection efficiency in light sheet microscopy. *Nature Communications* **8**:1452 <https://doi.org/10.1038/s41467-017-01250-8> | PubMed
- Yofe I, Weill U, Meurer M, Chuartzman S, Zalckvar E, Goldman O, Ben-Dor S, Schütze C, Wiedemann N, Knop M, *et al.* (2016) One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. *Nature Methods* **13**:371-378 <https://doi.org/10.1038/nmeth.3795> | PubMed
- Yu C-C (Jay), Orozco Cosio DM, Boyden ES (2022) ExCel: Super-resolution imaging of *C. elegans* with expansion microscopy. *Methods in Molecular Biology* **2468**:141-203 https://doi.org/10.1007/978-1-0716-2181-3_9 | PubMed
- Zhang H, Papadaki S, Sun X, Wang X, Drobizhev M, Yao L, Rehbock M, Köster RW, Wu L, Namikawa K, *et al.* (2023) Quantitative assessment of near-infrared fluorescent proteins. *Nature methods* **20**:1605-1616 <https://doi.org/10.1038/s41592-023-01975-z> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

Eroglu and Hobert demonstrate that injecting CRISPR guides and repair constructs to target three genes at a time, tagging each with a different fluorescent protein, and selecting which gene to tag with which fluorophore based on genes' expression levels, can improve efficiency of gene tagging.

Strengths:

This manuscript demonstrates that three genes can be targeted efficiently with three different fluorophores. It also presents some practical considerations, like using the fluorophore least complicated by agar/worm autofluorescence for genes with low expression levels, and cost calculations if the same methods were used on all genes.

Weaknesses:

Eroglu has demonstrated in a previous publication that single-stranded DNA injection can increase efficiency of CRISPR in *C. elegans*, while inserting two fluorescent proteins and a co-CRISPR marker into three loci, and Paix et al 2015 demonstrated simultaneous insertion of two fluorescent tags. The current work is valuable and incremental advance. In general, I applaud the authors' willingness to strategize about how whole proteome tagging might be accomplished. I predict that the advance here will be one of many small advances that will get the field to that goal. The title oversells the advance presented, in my view, since seems like one among many key advances, and the first sentence of the Discussion seems a more apt summary of the key advance here.

Some injections targeted genes on the same chromosome together, which will create unnecessary issues when doing crossing that will be useful for some future experiments. This made me wonder if injecting 3 together really is helpful vs targeting each gene separately, since only 5 worms need to be injected. It cuts time down by 2/3, but perhaps avoiding targeting the same chromosome with two tags would be useful.

The limited utility of current blue fluorescent proteins makes me wonder if it's worth using at this stage, before there are better blue fluorescent proteins, or better yet, far red, to avoid issues with live imaging under phototoxic UV or near-UV illumination.

<https://doi.org/10.7554/eLife.110717.2.sa4>

Reviewer #2 (Public review):

Original Review:

The manuscript by Eroglu and Hobert presents a set of strains each harboring up to three fluorescently tagged endogenous proteins. While there is technically nothing wrong with the method and the images are beautiful, we struggled to appreciate the advance of this work - who is this paper for?

As a technical method, the advance is minimal since the first author had already demonstrated that three mutations (fluorophore insertion and co-CRISPR marker) could be introduced simultaneously.

As a pilot for creating genome-scale resources, it is not clear whether three different fluorophores in one animal, while elegantly designed and implemented, will be desired by

the broader community.

Finally, the interpretation of the patterns observed in the created lines leaves much to be desired. A Table with all the observations must be included and can replace the tedious (and often wrong) descriptions of the observations with the different lines. It would be too much to point out every mistaken expectation of protein expression. Two examples include:

The expectation that ACDH-10 is enriched in the intestine and epidermal tissues (hypodermis) is naïve - there are multiple paralogs of this protein (look at WormPaths or WormFlux) that may share functions in different tissues. There is also no reason to assume that fatty acid metabolism does not occur in other tissues (including the germline). Finally, there are no published studies about this enzyme, so we really don't know for sure what it's doing.

The expectation that HXK-1 is ubiquitously expressed is similarly naïve. There are three paralogous enzymes that are all associated with the same reaction, and we have shown that these three function redundantly *in vivo*, perhaps in different tissues (PMID: 40011787). Moreover, single cell RNA-seq data (PMID: 38816550) also shows enrichment of *hvk-1* in gonadal sheath cells.

The table should have at least the following information: gene/protein name - Wormbase ID - TPM levels of single cell data assigned to tissues for L2, L4 and adult (all published) - tissues in which expression is observed in the lines presented by the authors.

Other points:

(1) We would encourage the authors to provide systematic validation of the reported insertions. The manuscript reports that 24 of 30 tags were isolated and visible but does not clearly state whether each isolated line was confirmed by sequence-level validation to be correctly in-frame and free of unintended mutations at the target locus.

(2) The manuscript presents aggregated success counts (e.g., 8/10 mTagBFP2 tags, 9/10 mStayGold, 7/10 mScarlet3) and useful narrative descriptions of injection outcomes. We suggest also to include per-locus success rates.

(3) For pools that required re-injection after initial failures, we would like to see a description of the specific changes that were made to the injection mixes or procedures (e.g., new repair template prep, different Cas9 reagent lot, guide redesign). This will be useful troubleshooting information for others.

(4) The authors states that the fluorophore sequences are codon-optimized for *C. elegans*. We suggest they provide the exact donor/tag sequences used specifically state whether the fluorophore sequences contain any synthetic/artificial introns or other sequence modifications (e.g., silent PAM-disrupting mutations) were included in the donor templates.

(5) Page 3: Include a reference for "The *C. elegans* genome encodes around 20,000 genes"

We hope these comments are useful.

Comments on Revised Version:

Overall, we found the responses to be quite recalcitrant.

We have one remaining composite concern about the comparison between observed expression patterns with the new strains versus published data.

First, the authors only report patterns for one stage while it should be not too much effort to image the different life stages. However, since this is a revision, we are not formally requesting they do this.

Second, in the now provided Table (thank you) 'observed expression' (last column) is lacking for 9 of the 30 proteins, and for 6 of these the procedure was not successful. Why not report patterns for the other three? It is confusing also because on page 5, the authors say that "overall, 24 of 30 tags ...all of which were visible with fluorescence stereomicroscopy" - are we missing something? Also, they then said that they "obtained 6/9 of the originally failed tags"; why are the corresponding patterns not included in table 1, and are 9 proteins still labeled as "no" in the "success?" Column?

Third, we strongly feel that the response to our comments about expression patterns is not adequate. On page 5 the authors say that "all proteins were expected to be ubiquitously expressed" and that "scRNA-seq indicated that transcript abundance was ubiquitous and without strong tissue-specific enrichment with few exceptions". However, in their rebuttal, the authors now argue for tissue-specific expression for proteins with paralogs, turning around their own argument! Moreover, their Table indicates that many genes show tissue-enriched expression by RNA-seq while many of their tagged proteins exhibit ubiquitous expression.

Overall, this indicates that both the overall accomplishment of generating tagged protein strains and analyzing their expression is oversold.

<https://doi.org/10.7554/eLife.110717.2.sa3>

Reviewer #3 (Public review):

Summary:

The authors argue that establishing the expression pattern and sub-cellular localisation of an animal's proteome will highlight hypotheses for further study. This claim is probably accepted by many in the community. This manuscript seeks to confirm the feasibility of establishing such a resource, by using current transgenic methods to knock in DNA encoding different colored fluorescent tags into *C. elegans* genes.

Strengths:

The authors make the points above. For example, they provide evidence that the *C. elegans* germline harbors two populations of mitochondria that differ qualitatively in the proteins they express. They also confirm that labelling the whole proteome is an achievable goal with relatively limited resources and time.

Weaknesses:

The work is somewhat incremental in that it uses existing transgenic technology. Cell biology in *C. elegans* is challenging because of the small size of many of its cells, notably neurons. This can make establishing the sub-cellular localisation of a fluorescently tagged protein, or co-localizing it with another protein, tricky. The authors point out in their introduction that advances in light microscopy such as diSPIM, STED and ISM (a close relative of SIM), have increased the resolution of light microscopy. They also point out that recent advances in expansion microscopy can similarly help overcome the resolution limit. However, they do not use these technologies to characterize their transgenic strains.

<https://doi.org/10.7554/eLife.110717.2.sa2>

Reviewer #4 (Public review):

Summary:

Tagging the entire proteome of a metazoan would be a landmark achievement, providing a powerful complement and extension to existing "omic" catalogs in model systems. Here, Eroglu and Hobert argue that efficiently tagging multiple loci in a single "batch" would make the community-based achievement of this goal realistic. They provide rigorous evidence that such an approach is indeed feasible, exploring issues related to efficiency, design and screening strategies, disruption of gene function, and the potential for endogenously tagged alleles to reveal unexpected aspects of protein expression and localization. While the work has some minor gaps that are important to rigorously assess the feasibility of the proposed effort, the detailed and valuable insights that emerge should provide impetus to the community to coordinate efforts to make this ambitious goal a reality.

Strengths:

The work has numerous strengths. The authors provide compelling evidence that:

- three distinct loci can be efficiently targeted with three distinct fluorescent tags in a single injection.
- thoughtful targeting design can reduce the likelihood of disruption of function by the tag.
- systematic design principles based on expression level and predicted localization/function can be used to optimize tagging strategies.
- the resulting tags can provide unexpected insight into patterns of protein production and subcellular localization.

Not all of these advances are novel in themselves, but taken together, they represent an important technical and conceptual advance. The most important strength comes from the exceptionally high value of the goal itself, in that the work is that it has the potential to spur a community-wide effort toward achieving the ambitious goal of proteome-wide tagging.

Weaknesses:

The work's shortcomings are minor.

- One concern has to do with the feasibility of the proposed screening strategies. The experimental design cleverly coinjects tags for three loci in different gene expression 'zones'; this expression level determines which tag will be used. As the authors allude to, there is an important distinction between genes with the same overall FKPM value between those that are expressed broadly and those focally expressed in a specific tissue. The proposed strategy claims that there are a sufficient number of highly expressed genes "to be used as visible markers" for recovering successfully edited animals. It would be useful for the authors to discuss the issue of broad vs focused expression among this set of genes a bit more thoroughly, with an eye toward the issue of how likely it is that these genes could indeed consistently be used as visible markers, particularly for those at the low end of this limit.
- What fraction of the proteome (on a per-gene basis) is secreted proteins? How difficult will it be to screen these for successful tags? Are there specific tags that would be more optimal for secreted proteins? (The authors mention the use of an SL2 or T2A cassette to label the cells in which these proteins are expressed but note that there are technical challenges associated with doing this at scale.)
- For secreted and/or weakly expressed genes, it would be useful for the authors to estimate for what fraction of these would successful insertions need to be screened by PCR, and what resources (time and money) this would likely entail.
- For how many genes would a single tag not capture all predicted isoforms?

- Finally, some readers might object to the authors' assertion in the abstract that this work is "a first step in this direction" (presumably referring to designing a strategy for whole-proteome tagging). There is no concern that the authors are disregarding the extensive work of other groups, as they explicitly mention the contributions of other groups to the foundation that enables the present work. However, the spirit of the abstract could be misinterpreted by a well-intentioned reader.

<https://doi.org/10.7554/eLife.110717.2.sa1>

Author response:

The following is the authors' response to the original reviews.

eLife Assessment

The nematode C. elegans is an ideal model in which to achieve the ambitious goal of a genome-wide atlas of protein expression and localization. In this paper, the authors explore the utility of a new and efficient method for labeling proteins with fluorescent tags, evaluating its potential to be the basis for a larger, genome-wide effort that is likely to be very useful for the community. While the evidence for the method itself is solid, carrying out this project at a large scale will require significant additional feasibility studies.

We appreciate the editor's recognition that the evidence for our method is solid and that a genome-wide protein atlas in *C. elegans* would be highly valuable to the community. However, we respectfully disagree that "significant additional feasibility studies" are required. Take the yeast proteome-wide GFP tagging project (Huh et al., Nature 2003). It achieved ~75% coverage of ~6,000 proteins directly from an established protocol without any prior significant feasibility studies, at least to our knowledge. While the *C. elegans* genome is 3 times in size, we would argue that our tagging protocol may even be less labor intensive as it does not involve any cloning and the screening is visual, requiring no molecular biology skills. Reviewer 3 notes: "They also provide convincing evidence that labelling the whole proteome is an achievable goal with relatively limited resources and time."

Our pilot study validates all key parameters for genome-wide scaling: editing efficiency at novel loci with untested reagents, viability of tagged worms, and detectability of multiple spectrally separated fluorophores across expression ranges. These address the core technical, biological, and practical challenges of large-scale endogenous tagging in a multicellular organism, leaving no fundamental barriers in our view.

The proposed cost and timeline align quite favorably with established large-scale consortium projects: e.g., ENCODE pilot analyzed 1% of the human genome at ~\$55 million over 4 years; Mouse Knockout Consortium scaled to ~20,000 genes over 20 years (ongoing) with ~\$100 million; Human Protein Atlas mapped ~87% of proteins with antibodies in fixed cells (through much more labor intensive methods) over 20+ years at >\$100 million. With ~8% of *C. elegans* genes already tagged (WormTagDB) and labs already tagging entire gene classes (PMID: 40463100), scaling our protocol to the proteome is feasible, potentially covering the genome in 5-6 years by a single lab or faster with distributed effort at a reagent cost of merely \$2.2 million. The main barriers now are funding commitment and assembling collaborators, not further feasibility testing.

Public Reviews:

Reviewer #1 (Public review):

Summary:

Eroglu and Hobert demonstrate that injecting CRISPR guides and repair constructs to target three genes at a time, tagging each with a different fluorescent protein, and selecting which gene to tag with which fluorophore based on genes' expression levels, can improve the efficiency of gene tagging.

Strengths:

This manuscript demonstrates that three genes can be targeted efficiently with three different fluorophores. It also presents some practical considerations, like using the fluorophore least complicated by agar/worm autofluorescence for genes with low expression levels, and cost calculations if the same methods were used on all genes.

Weaknesses:

*Eroglu has demonstrated in a previous publication that single-stranded DNA injection can increase the efficiency of CRISPR in *C. elegans* while inserting two fluorescent proteins and a co-CRISPR marker into three loci. The current work is, therefore, an incremental advance. In general, I applaud the authors' willingness to think ahead to how whole proteome tagging might be accomplished, but I predict that the advance here will be one of many small advances that will get the field to that goal.*

Our manuscript indeed builds on prior multiplex editing (including our own co-CRISPR work), but the manuscript's primary contribution is not a novel technical breakthrough *per se*. Instead, our main goal was to pilot and strategize a feasible path to whole-proteome tagging in *C. elegans* and, most critically, test the following key parameters: (1) success rate of triple pools with prior untested reagents at novel targets; (2) utility of fluorophores across expression levels; (3) major effects on tagged protein function. In prior multiplexing, we used two targets which we already knew could be edited quite efficiently, with the 3rd target a point mutation with nearly 100% efficiency. Thus, it was not at all clear that picking 3 random genes and replacing the 3rd highly efficient locus with another less efficient large insertion would work or be sufficiently scalable for thousands of novel genes with unvalidated reagents at first pass.

The title vastly oversells the advance in my view, and the first sentence of the Discussion seems a more apt summary of the key advance here.

Some injections target genes on the same chromosome together, which will create unnecessary issues when doing necessary backcrossing, especially if the mutation rate is increased by CRISPR.

We disagree with the reviewer's assessment of the need for backcrossing, for two reasons: (1) Prior studies have shown that off-target mutations are not a serious concern in *C. elegans* (reviewed in PMID: 26336798). For instance, WGS of strains after CRISPR/Cas9 found negligible off-target effects (PMID: 25249454, PMID: 30420468 – using similar RNP/ssDNA method and multiple guides; PMID: 23979577, PMID: 27650892 using other methods). Targeted sequencing studies have reported similar findings, using various CRISPR/Cas9 methods, with essentially no mutations at sites other than the intended target (PMID: 23995389; PMID: 23817069). (2) If the goal is to tag the entire genome, the introduction of backcrossing should not reasonably be a routine part of the initial tagging.

Lastly, if one really does want to backcross, the existence of tags on the same chromosome is actually an advantage because it permits selection for recombinants with wild-type chromosomes.

Also, the need for backcrossing and perhaps sequencing made me wonder if injecting 3 together really is helpful vs targeting each gene separately, since only 5 worms need to

| *be injected.*

Apart from our disagreement regarding backcrossing, we are puzzled by the reviewer's comment. Why would one do single tagging at a time, rather than triple tagging if the whole point is to scale up tagging? It is important to keep in mind that the rate limiting step for tagging the whole genome is the number of injections that can be done per day. Since there is no cloning to generate the repair templates/guides and all other reagents are commercially available and not sample specific, these can be prepared quite rapidly. Being able to isolate multiple lines (together or independently) from the same injection increases throughput 3-fold and in our view does not provide any disadvantages as individual tags can be isolated independently if desired.

Beyond the numerous technical advantages pooling provides (also lower cost and throughput for making injection mixes as well as imaging), our results show that it yields epistemic benefits as well: we would never have noted the subcellular pattern in Fig. 6B, C with different sets of mitochondria being marked by different mitochondrial proteins had we imaged them separately or even aligned to a pan-mitochondrial landmark. As we mentioned in the discussion, grouping proteins predicted to localize to the same compartment together can simultaneously test how uniform or differentiated such compartments are during the screen.

| *The limited utility of current blue fluorescent proteins makes me wonder if it's worth using at all at this stage, before there are better blue (or far red) fluorescent proteins.*

We do not think that the utility of current BFPs is that limiting. At least the theoretical brightness of mTagBFP2 is comparable to that of EGFP (PMID: 30886412), which was useful for the bulk of currently tagged proteins. Due to modestly higher autofluorescence in the blue spectrum, the practical brightness is somewhat less ideal, but we have shown that many proteins are expressed high enough to be detected quite well with mTagBFP2 by eye at low magnification. We also note that many tags that are not visible by eye under a dissection scope become visible with long exposure cameras of widefield microscopes or modern confocal (GaAsP) detectors, so the list of genes detectable with mTagBFP2 is likely to be much higher. We routinely use mTagBFP2 to super-resolve subnuclear structures with endogenous tags (e.g., in the nucleolus), with some tags having lower annotated FPKMs than the genes tested here.

| *Some literature reviews, particularly in the Introduction and Abstract, rely too much on recent examples from the authors' laboratory instead of presenting the state of the field. I'd like to have known what exactly has been done with simultaneous injection targeting multiple loci more thoroughly, comparing what has been accomplished to date by various laboratories' advances to date.*

We are not sure what the reviewer is referring to. In the Abstract, we do not refer to any literature. In the Introduction, we cite 28 papers, 6 of those from our lab (4 of which providing examples of protein tags). We do not believe that this can be fairly called an unbalanced presentation of the state of the field.

This being said, we have gladly expanded our Introduction to provide more background on co-CRISPRing. Labs have routinely used co-conversion ("coCRISPR") markers for picking out their intended edits (e.g., point mutations or insertions), as it has been shown by multiple groups that a CRISPR/Cas9 edit at one locus correlates with efficiency at other simultaneous targets (PMID: 25161212). Generally, making point mutations with the Cas9/RNP protocol is highly efficient, especially at specific loci such as *dpy-10*. However, multiple FP-sized insertions have not been routinely attempted. We and only one other group have successfully attempted it using previously working targets and reagents (e.g., 28% in PMID: 26187122). Importantly, the efficiency of such multiple insertions has never been assessed at scale and

using entirely untested reagents at novel sites – critical parameters to determine for a whole genome approach. So, we test here (1) the efficiency of triple insertions and (2) the chance of getting them with new and untested guides and reagents.

In our view, since we have to use some injection/coCRISPR marker anyway for those genes which are not expressed at dissecting-scope visible levels (likely most genes), using highly expressed intended targets as improvised markers in a pooled approach makes our approach much more efficient. It allows us to find the worms with the highest chance of yielding CRISPR insertions, which we can screen with higher power methods for the dimmer targets, while enabling us to co-isolate other intended targets. Insertions, being often heterozygous in F1, can be segregated independently if desired, or homozygosed together to facilitate maintenance then outcrossed individually by those interested in studying specific genes in more detail.

In the revised version of this manuscript, we now discuss some of these points in the introduction section:

“Currently, around 1554 proteins representing 8% of the proteome are estimated to have been endogenously tagged (Leyhr et al., 2025). However, at current rates, tagging the proteome is projected to take around 100 years and likely involve numerous duplicate attempts on a small number of commonly studied proteins (Leyhr et al., 2025). It will thus be crucial for the field to coordinate tagging efforts and scale up tagging protocols to enable coverage of the entire genome at a reasonable timescale and cost. Given the number of injections is a major time-limiting factor, pooling multiple injections into one would at minimum cut tagging time by a factor of 3. In *C. elegans*, screening for novel CRISPR/Cas9-induced genomic edits is already facilitated either by use of co-injection markers (i.e., plasmids that form extrachromosomal arrays) that yield phenotypes or fluorescence in progeny of successfully injected worms, or co-editing well characterized loci using established and highly efficient reagents which likewise yield visible phenotypes. In the latter approach, termed “co-CRISPR”, worms edited at the marker locus are most likely to also carry the intended edit (Arribere et al., 2014). Recent methods for CRISPR/Cas9 mediated genomic insertions have pushed efficiencies to sufficient levels to simultaneously insert multiple fluorophores (e.g., mNeonGreen and mScarlet) as well as a co-CRISPR marker (*dpy-10*) at three independent loci in a single injection (Eroglu et al., 2023; Paix et al., 2015). These attempts pooled reagents previously established to work efficiently and targeted genes that were known to yield functional fusion proteins when tagged. Thus, while in principle current methods could allow tagging of at least 3 independent loci in one injection if a co-CRISPR marker is omitted, it is not known to what extent such an approach could be generalized across the genome with previously unvalidated reagents (i.e., guides and repair template homology arms) at novel loci to yield functional tags”

Reviewer #2 (Public review):

The manuscript by Eroglu and Hobert presents a set of strains each harboring up to three fluorescently tagged endogenous proteins. While there is technically nothing wrong with the method and the images are beautiful, we struggled to appreciate the advance of this work - who is this paper for?

We consider this paper to have two purposes: (1) motivate the community to come together to consider such genome-wide tagging approach; (2) provide a reference point for funding agencies that such an aim is not unreasonable and will provide novel interesting insights.

As a technical method, the advance is minimal since the first author had already demonstrated that three mutations (fluorophore insertion and co-CRISPR marker) could be introduced simultaneously.

We agree that the basic principle is similar. However, it was not clear that triple pooling three novel large edits would work, given the numbers in our original paper or that it would be scalable.

The *dpy-10* coCRISPR marker previously used is a highly efficient single site, with close to 100% hit rate. We also knew in the earlier study that the two pooled insertions already worked quite efficiently and did not disrupt the function of targeted proteins. Exchanging these plus *dpy-10* for three novel tags was not guaranteed to succeed for many potential reasons, including both biological and technical. For instance, such a “marker free” approach necessitates that a significant number of targets in the genome should be expressed highly enough to be visible by fluorescence stereomicroscopy when tagged with current best fluorophores. The chance of disrupting gene function by tagging was also not explored in detail in *C. elegans*, nor whether one untested guide is generally sufficient. We think that establishing these parameters was meaningful and necessary for the goal of whole genome tagging. We have clarified some of these points in the text.

As a pilot for creating genome-scale resources, it is not clear whether three different fluorophores in one animal, while elegantly designed and implemented, will be desired by the broader community.

The usage of three different fluorophores is largely driven by the ability to co-inject and therefore cut injection effort by a factor of three. Moreover, having all three fluorophores together facilitates imaging and maintenance. Lastly, co-labeling has the potential to reveal unexpected patterns of co-localization or lack thereof (example: two mitochondrial proteins that we found to not have overlapping distribution). We clarified this point in the revised text in both the results and discussion.

Finally, the interpretation of the patterns observed in the created lines is somewhat lacking. A Table with all the observations must be included. This can replace the descriptions of the observations with the different lines, which could be somewhat laborious for the reader, and are often wrong. There are numerous mistaken expectations of protein expression here, but two examples include:

We are not convinced that our expectations are mistaken. Below we respond to the reviewer’s specific examples, and we are open to hear from the reviewer about additional cases.

(1) The expectation that ACDH-10 is enriched in the intestine and epidermal tissues (hypodermis).

There are multiple paralogs of this protein (see WormPaths or WormFlux) that may share functions in different tissues. There is also no reason to assume that fatty acid metabolism does not occur in other tissues (including the germline). Finally, there are no published studies about this enzyme, so we really don't know for sure what it's doing.

The expression of *acdH-10* is annotated in multiple scRNA datasets as intestine and epidermal enriched (CeNGEN/Taylor et al. 2021, highest in epidermis; Ghaddar et al 2023 highest in intestine). We did not mean to imply that fatty acid metabolism does not occur in the gonad, nor that a paralog of *acdH-10* could not be performing the same function in tissues where *acdH-10* is not expressed.

However, this raises an important question: why have different paralogs doing the same thing? Duplicate genes with the same function are generally not evolutionarily stable (PMID: 11073452, PMID: 24659815). That there are such striking tissue specific expression patterns of an essential or widely expressed protein class suggests that paralogs of the gene likely differ in some meaningful parameter that might align with tissue-specific functional needs or

regulation. The reviewer's statement that 'there are no published studies about this enzyme, so we really don't know for sure what it's doing' is in fact an excellent demonstration of our point; finding out where the duplicates are expressed can provide a starting point to uncover potential differences between the paralogs. At the very least it can delineate to what degree paralogs diverge in their expression across the proteome and identify which such cases merit further study. In a more ideal scenario, prior information of protein function could indicate that the involved pathway requires tissue specific regulation.

(2) *The expectation that HXK-1 is ubiquitously expressed.*

Three paralogous enzymes are all associated with the same reaction, and we have shown that these three function redundantly in vivo, perhaps in different tissues (PMID: 40011787).

The cited paper (PMID: 40011787) does not show where they are expressed. We discussed redundancy/paralogs above in point 1, and in our view the same applies here. They may perform the same reaction but are likely to differ in some meaningful way, be it regulation or rate of activity, for them to be stably maintained as functional genes over evolution.

Moreover, single-cell RNA-seq data (PMID: 38816550) also show enrichment of hxx-1 in gonadal sheath cells.

The Ghaddar et al. and CeNGEN/Taylor et al. datasets do not show this. The scRNA paper cited (PMID: 38816550) also shows enrichment in neurons, pharynx, coelomocyte and germ cells which we did not note. In our view, these in fact further support our goals: often, transcript datasets alone (frequently used to infer tissue function) do not sufficiently predict protein expression. One can post hoc find an scRNA-seq dataset that aligns somewhat with our protein observations, but how does one know which to trust *a priori*? Disagreements between transcript datasets will ultimately require resolution at the protein level, in our view.

To clarify these points, we added the following to the discussion section:

“We also noted unexpected cell type dependent distributions of proteins involved in broadly important metabolic processes such as ACDH-10, which was depleted from the germline compared to other tissues, and HXK-1, which was highly enriched in the gonadal sheath. Notably, for these as well as other cases, scRNA-seq datasets were not sufficient to deduce *a priori* the observed cell type specific differences at the protein level. Importantly, many genes encoding metabolic enzymes including *acd-10* and *hxx-1* have paralogs that likely perform similar catalytic functions. Yet, duplicate genes with identical functions are generally not evolutionarily stable (Adler et al., 2014; Lynch and Conery, 2000); thus such genes are likely to differ in some meaningful parameter (e.g., regulation or activity) that might align with tissue-specific functional needs. Fully annotating the expression patterns of paralogs at the protein level could indicate which tissues require unique metabolic needs and indicate which paralogous genes have undergone sub- versus neo-functionalization. For those proteins that are less functionally understood, unexpected distributions might indicate which merit further study.”

The table should have at least the following information: gene/protein name - Wormbase ID - TPM levels of single cell data assigned to tissues for L2, L4, and adult (all published) - tissues in which expression is observed in the lines presented by the authors.

We added some of this information such as annotated expression levels in young adults from various scRNA datasets (but not larval datasets as we did not image these). We note that each of these studies use different pipelines and report different metrics (scaled TPM/Z-score versus Seurat average expression versus TPM), so comparisons between them are not informative unless they are integrated and analyzed together.

Reviewer #3 (Public review):
Summary:

*The authors argue that establishing the expression pattern and subcellular localisation of an animal's proteome will highlight many hypotheses for further study. To make this point and show feasibility, they developed a pipeline to knock in DNA encoding fluorescent tags into *C. elegans* genes.*

Strengths:

*The authors effectively make the points above. For example, they provide evidence of two populations of mitochondria in the *C. elegans* germline that differ qualitatively in the proteins they express. They also provide convincing evidence that labelling the whole proteome is an achievable goal with relatively limited resources and time.*

We appreciate the referee's recognition that whole proteome tagging is feasible.

Weaknesses:

*Cell biology in *C. elegans* is challenging because of the small size of many of its cells, notably neurons. This can make establishing the sub-cellular localisation of a fluorescently tagged protein, or co-localizing it with another protein, tricky. The authors point out in their introduction that advances in light microscopy, such as diSPIM, STED, and ISM (a close relative of SIM), have increased the resolution of light microscopy. They also point out that recent advances in expansion microscopy can similarly help overcome the resolution limit.*

*(1) Have the authors investigated if the three fluorescent tags they use are appropriate for super-resolution microscopy of *C. elegans*, e.g., STED or SIM? Would Elektra be better than mTAGBFP2? How does mScarlet3-S2 compare to mScarlet 3?*

All three tags work for ISM (i.e., Airyscan). We previously tried Electra (not for the genes tested here) but could not isolate positive tags. Given Electra is not that much brighter on paper than mTagBFP2 we did not pursue it further, though we recognize that these may simply have been unlucky injections. mScarlet3-S2 is quite a bit dimmer than mScarlet3 on paper – the advantage is that it has higher photostability. In our view, the limiting factor will be having FPs that are bright enough to screen, image and scale to the whole genome, so brightness will likely provide an advantage over photostability at this stage.

(2) Have the authors investigated what tags could be used in expansion microscopy - that is, which retain antigenicity or even fluorescence after the protocol is applied? It may be useful to add different epitope tags to the knock-in cassettes for this purpose.

mSG and mSc3 retain fluorescence after fixing with formaldehyde. We have not tested mTagBFP2 fluorescence in fixed worms. We agree that adding different epitope tags would be useful.

The paper is fine as it stands. The experiments above could add value to it and future-proof it, but are not essential. If the experiments are not attempted, the authors could refer to the points above in the discussion.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

(1) Merged figures appear saturated, and use colors that won't work for red-green colorblind viewers.

For all figures, we also show individual channels separately, which is common practice for making fluorescence images accessible to colorblind readers (PMID: 33788834). Figures highlighting non-overlap like 6B and C are already in accessible colors when merged (blue/green) and include a numerical quantification. 3-color RGB images preserve the greatest information for the highest number of individuals.

(2) Targeting ubiquitously expressed genes as a proof of concept gives me some concern that this might underestimate the challenges that may be experienced with less widely expressed genes.

While the genes were predicted to be ubiquitously expressed, many were not in practice, like HXK-1 and F54C8.1, which were also among the lower expressed genes on our list and highly cell type restricted. As discussed, the more tissue restricted a gene, the likelier that bulk RNA levels underestimate expression. Such genes are therefore more likely to be detected in a specific tissue. We routinely isolate tissue restricted endogenous tags, including those expressed in only a few neurons, with bulk FPKMs lower than the ranges tested in this manuscript.

(3) Some results are not shown or referenced (autofluorescence, for example, is shown using a schematic in Figure 1C).

We now provide representative images alongside what would be expected to be observed by eye during screening.

(4) It would be useful to describe how to recover worms from what is shown in Figure 1A.

In the revised version, we added the following in the caption for Fig. 1A:

“Selected worms expressing the brighter tag can be screened for dimmer tags by higher magnification and long exposure imaging. Worms can be recovered directly from slides if immobilized by levamisole as described (Ghanta et al., 2021). Alternatively, single hermaphrodite worms can be isolated, allowed to lay eggs, then screened.”

(5) A blue bar of data must be missing from Figure 3B injection pool 5.

As stated in the text, “All but one tag (cox-6B::mTagBFP2) was visible in the F1 generation of injected P0 animals, and these were subsequently isolated among F2 worms positive for the other tags in the pool.”

To clarify that data points are not unintentionally omitted, we added the following text to the caption of Fig. 3B:

“For group 5 including cox-6B::mTagBFP2, worms with detectable levels of mTagBFP2 fluorescence were not recovered in the F1 generation but were isolated among progeny of F1s positive for mStayGold and mScarlet3; we were thus unable to quantify efficiency for this locus at F1.”

(6) Some expression or localization patterns were unexpected, but complications like germline silencing and protein mislocalization, with a small fraction localizing normally and rescuing function, were not presented as possibilities. Viability is used to confirm function, but without presenting whether this means 100% viability, less, or just the ability to maintain a strain.

We already do discuss mislocalization and functionality issues in the Discussion, as well as tradeoffs of alternate methods. Any existing method to observe biological molecules, be it protein, RNA or DNA, has multiple drawbacks and sources of artifacts, which are unlikely to be fully eliminated in the foreseeable future.

In regard to germline silencing of endogenously tagged genes in *C. elegans*, there is actually very little evidence for this. Collectively, various labs have now generated over 200 reporter alleles of germline-expressed genes (WormTagDB), with robust expression throughout the germline and retention of function. Likewise, numerous of our tags across fluorophores showed robust germline expressions including EEF-1A.1::mTagBFP2, Y22D7AL.10::mStayGold, and HAT-1::mScarlet3. In fact, overall transcript levels generally tended to underestimate germline enrichment at the protein level. We note that single-copy transgenes driven by *eef-1A.1/eft-3* promoter by itself are frequently not expressed in the germline (PMID: 31064766); that we could detect EEF-1A.1 robustly in the germline when tagged endogenously is evidence that silencing is unlikely to be a widespread concern, and at the least less of a concern than single copy transgenes. We appreciate that for a transgene, presence/absence of specific sequence elements and genomic loci play a role in expression, but an endogenous tag captures all such information at a given locus.

Indeed, we found only two reports of endogenous tags being silenced in the germline, the first being a novel tag (not fluorophore) which initially prevented expression at the tagged locus (PMID: 30109984), but after making changes to the sequence to avoid silencing signals the authors could rescue expression and thereafter saw robust expression in various novel contexts with this tag. The second example (PMID: 34547227) leaves open the possibility that germline repression of that particular gene might be a part of its endogenous regulation.

Nevertheless, given it is probably rare if occurs at all, it will likely take a large scale tagging effort to uncover such cases at sufficient numbers to study. In our view, this further justifies tagging at large, ideally genomic, scales. If we do discover that there are numerous annotated germline proteins which we don't observe by tagging, that would be interesting to study on its own.

(7) *Halotag* is presented in the Discussion as a small tag, but it is bigger than GFP.

Thank you for catching this. We have removed the discussion of Halotag. Given the comparable size to FPs, it would be unlikely to alleviate issues of tag functionality.

(8) It would be useful to include FPKMs and viability percentages in Table 1.

FPKM is included in column 6, but the title for this column is cut off. In the revised table FPKM values are now shown more clearly across stages.

We did not quantify viability percentage. In our view it does not yield an informative metric when there is little information about the protein's required dosage for function, which was the case for most proteins here. A haplosufficient gene might yield a full brood size even if 50% of protein function is lost; conversely, a highly dose sensitive protein could yield penetrant and severe inviability with mild perturbation of function. It also is not actionable information at this stage if there is no alternate tagging strategy as a baseline of comparison. The worms we picked to image all have viable embryos as adults, so in those individuals the genes were likely to be sufficiently expressed and functional.

(9) *Because establishing that a guide works well is a limiting step for many CRISPR experiments (once a guide works well, it's easy to inject 5 worms and get lines), I wondered if testing that for many genes is what is really needed in the field at this stage.*

Guide quality is rarely an issue in *C. elegans*, as for all the genes here we tried only one guide, all of which were previously untested. We now clarified this in the discussion section:

“Notably, we find that previously untested guide RNAs and homology arms perform exceptionally well at novel loci, as we only tested one set of reagents for each locus which yielded satisfactory tagging rates.”

(10) *For a manuscript where the injection is so central to what was done, I was surprised to read in the Acknowledgments that all of the injections were done by someone who is not included as an author.*

We are likewise surprised by such a comment but gladly clarify: Chi Chen has been with us as an expert microinjection specialist for more than 25 years and her very important technical contributions have been acknowledged in many dozen papers. Multiple authorship guidelines, including COPE's and ICMJE's, state that technical contributions alone do not qualify for authorship.

Reviewer #2 (Recommendations for the authors):

(1) *We would encourage the authors to provide systematic validation of the reported insertions. The manuscript reports that 24 of 30 tags were isolated and visible, but does not clearly state whether each isolated line was confirmed by sequence-level validation to be correctly in-frame and free of unintended mutations at the target locus.*

We appreciate the reviewer's concerns on fidelity. These parameters have been assessed in prior published work (e.g., PMID: 30504364, PMID: 34748534) and in our hands are in the range of 80% whenever we sequence non-fluorescent tags of similar sizes. The efficiencies we observed are high enough that one can expect to recover numerous worms with the exact intended sequence for each target, though we would argue mutations within the FP reporter are less likely to matter if it retains high fluorescence.

(2) *The manuscript presents aggregated success counts (e.g., 8/10 mTagBFP2 tags, 9/10 mStayGold, 7/10 mScarlet3) and useful narrative descriptions of injection outcomes. We also suggest including per-locus success rates.*

Figure 3B shows per locus success rate and source data is provided for this figure. Each dot is an individual injection and the Y axis is per locus rate. We now worded this more clearly in the figure's caption.

“Total insertion efficiencies per locus for the indicated targets across injection pools.”

(3) *For pools that required re-injection after initial failures, we would like to see a description of the specific changes that were made to the injection mixes or procedures (e.g., new repair template prep, different Cas9 reagent lot, guide redesign). This will be useful troubleshooting information for others.*

We re-made the exact same injection mix but with nanodrop to ensure the purity of the repair templates as assessed by absorbance ratios (A260/230 and A260/280) were sufficient after each purification step. No other changes were made. This is now specified in the methods section in the following way:

“For re-runs of pools 4, 6 and 10 which failed initially, we regenerated the repair templates and ensured that after each column purification, the A260/230 ratio of the purified DNA was

≥2.2 and A260/280 was 1.8 ± 0.05 when measured with a Nanodrop spectrophotometer.”

(4) The authors state that the fluorophore sequences are codon-optimized for C. elegans. We suggest they provide the exact donor/tag sequences, specifically state whether the fluorophore sequences contain any synthetic/artificial introns, or whether other sequence modifications (e.g., silent PAM-disrupting mutations) were included in the donor templates.

This information is provided in Supplementary Table 1.

(5) Page 3: Include a reference for "The C. elegans genome encodes around 20,000 genes"

We added a reference to the most recent release of the genome (WS237, May 2013). Spieth et al., 2014.

<https://doi.org/10.7554/eLife.110717.2.sa0>