High-Throughput Library Transgenesis in *Caenorhabditis elegans* via Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS)

Zachary C. Stevenson[1], Megan J. Moerdyk-Schauwecker[1], Stephen A. Banse[1], Dhaval S. Patel[2,3,4], Hang Lu[2,3], and Patrick C. Phillips[1,&]

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, United States of America.

[2]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States of America.

[3]Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, United States, Atlanta, GA 30332, United States of America.

[4]Present address: NemaLife Inc., Lubbock, TX, 79409. [&]Corresponding Author pphil@uoregon.edu

[&] Corresponding Author pphil@uoregon.edu

Running head: High-Throughput Library Transgenesis in *C. elegans*

Abstract

High-throughput transgenesis using synthetic DNA libraries is a powerful method for systematically exploring genetic function. Diverse synthesized libraries have been used for protein engineering, identification of protein-protein interactions, characterization of promoter libraries, developmental and evolutionary lineage tracking, and various other exploratory assays. However, the need for library transgenesis has effectively restricted these approaches to single-cell models. Here we present Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS), a simple yet powerful approach to large-scale transgenesis that overcomes typical limitations encountered in multicellular systems. TARDIS splits the transgenesis process into a two-step process: creation of individuals carrying experimentally introduced sequence libraries, followed by inducible extraction and integration of individual sequences/library components from the larger library cassette into engineered genomic sites. Thus, transformation of a single individual, followed by lineage expansion and functional transgenesis, gives rise to thousands of genetically unique transgenic individuals. We demonstrate the power of this system using engineered, split selectable TARDIS sites in *Caenorhabditis elegans* to generate (1) a large set of individually barcoded lineages and (2) transcriptional reporter lines from pre-defined promoter libraries. We find that this approach increases transformation yields up to approximately 1000-fold over current single-step methods. While we demonstrate the utility of TARDIS using *C. elegans*, in principle the process is adaptable to any system where experimentally generated genomic loci landing pads and diverse, heritable DNA elements can be generated.

Introduction

Transgenesis, which is the specific and heritable introduction of foreign DNA into genomes, has been a central tool for functional analysis and genetic engineering for nearly 40 years. The power of transgenesis is due in part to the wide variety of assays and techniques that are built upon controlled introduction of novel DNA sequences into a native genome. While there are many uses for transgenesis, in practice most can be grouped into those inserting a small number of known sequences (specific transgenesis) and those introducing many sequence variants from experimental libraries (exploratory transgenesis). While the ability to perform specific transgenesis has become a de facto requirement for all model organisms, exploratory transgenesis remains effectively limited to single-cell models (both prokaryotic and eukaryotic) caused by biological limitations generated by inheritance in multicellular organisms. In single-cell models, high-throughput transgenesis has been used for exploratory sampling of sequence space using protein interaction libraries (Joung et al., 2000), barcode-lineage tracking libraries (Levy et al., 2015; Nguyen Ba et al., 2019), directed evolution (Packer and Liu, 2015), synthetic promoter library screens (Wu et al., 2019), and mutagenesis screens (Bock et al., 2022; Erwood et al., 2022; Kim et al., 2022; Sánchez-Rivera et al., 2022). Despite the usefulness of such experiments in single-celled systems, either in microorganisms or in cell culture, increasing transgenic throughput in multicellular models holds the potential to expand the impact of exploratory transgenesis in

functional domains, such as inter-tissue signaling, neuronal health, and animal behavior, that are dependent on multicellular interactions and therefore difficult to replicate in single-cell models.

Exploratory transgenesis in single-cell models has been facilitated by the availability of *in-vitro*-generated DNA libraries, selectable markers, plasmids, *in vivo* homologous recombination, and most importantly, the ability to massively parallelize transgenesis using microbial transformation or eukaryotic cell transfection/transduction. Currently, there is no practical means to make populations of uniquely transgenic individuals from sequence libraries at a similar scale in animal systems due to the Weismann Barrier (Weismann, 1893): the split between soma and germline. The requirement that the germline be accessible and editable has forced animal systems into a transgenic bottleneck compared to single-cell systems because it is very difficult to introduce exogenous DNA directly into the germline in a high-throughput manner, relying instead on injection, bombardment, or some other physical intervention. This low-throughput limitation in animals dramatically reduces the sequence diversity that can be sampled, effectively preventing large-scale exploratory experiments from being performed. Attempts have been made to parallelize transgenic creation in multicellular model organisms, for example, the development of Brainbow (Livet et al., 2007; Weissman and Pan, 2015), ifgMosaic analysis (Pontes-Quero et al., 2017), P[acman] libraries in *Drosophila* (Venken et al., 2009), and multiple types of transformation in plants (Ismagul et al., 2018; Xu et al., 2022). In *Caenorhabditis elegans*, CRISPR technology combined with custom engineered sites within the genome ("landing pads") has facilitated the generation of single-copy integrations (Malaiwong et al., 2023; Nonet, 2020, 2021; Silva-García et al., 2019; Stevenson et al., 2020; Vicencio et al., 2019), and attempts have been made to multiplex transgenesis using traditional integration methods in conjunction with specialized landing pad systems (Gilleland et al., 2015; Kaymak et al., 2016; Mouridi et al., 2022; Radman et al., 2013). While these efforts have increased throughput over standard single copy integration methods, throughput still remains too low for effective exploratory transgenesis, and in some cases requires significant additional labor, cost, equipment and/or expertise.

Here we present "Transgenic Arrays Resulting in Diversity of Integrated Sequences" (TARDIS) (Stevenson et al., 2021), a simple yet powerful alternative to traditional single-copy transgenesis. TARDIS greatly expands throughput by explicitly separating and reordering of the conceptual steps of transgenesis (Figure 1). To increase throughput, TARDIS begins with an *in vitro* generated DNA sequence library that is introduced into germ cells via traditional low throughput methods (i.e., germline transformation, Fig 1). While traditional transgenesis typically couples the physical introduction of DNA into cells with the integration of a selected sequence from the original library, the DNA sequences in TARDIS are designed to be incorporated in large numbers into diverse, heritable sub-libraries (TARDIS libraries), rather than be directly integrated into the desired genomic locus. In addition to the sequence library, a functioning selectable marker is also included to stabilize the inheritance of the TARDIS library over generations. These TARDIS libraries function to create "metaploidy" – expanding the total number of alleles available for inheritance, essentially making the worm genetically "bigger on the inside." TARDIS library-bearing animals are then allowed to propagate under selection to generate a large population of TARDIS library carriers. After population expansion, genome integration of a single sequence unit is performed by inducing a double strand break at a genetically engineered landing pad. This landing pad is designed to both integrate a sequence unit and act as a second selectable marker. We chose *C. elegans* to validate the TARDIS approach because *C. elegans* naturally form extrachromosomal arrays that can be several megabases in size (Carlton et al., 2022; Lin et al., 2021; Mello et al., 1991; Stinchcomb et al., 1985) from injected DNA, which simplifies the generation of heritable "TARDIS library arrays" (TLA) that encompass sequence significant diversity.

We demonstrate the functionality of TARDIS for two use cases: unique animal barcoding and promoter library transgenesis. Barcoding has been widely adopted in microbial systems for evolutionary lineage tracking (Jahn et al., 2018; Levy et al., 2015; Nguyen Ba et al., 2019) and for developmental lineage tracking in animals (Kebschull and Zador, 2018; McKenna et al., 2016). In microbial systems, barcode libraries have relied on highly diverse randomized oligo libraries, as compared to animal systems, which have relied on CRE recombinases or randomized Cas9-induced mutations. Here we present a novel TARDIS barcoding system for an animal model which mimics the scope and diversity previously only possible using microbial systems. Our results show that large, heritable libraries containing thousands of barcodes can be created and maintained as extrachromosomal arrays. Individual sequences are selected and removed from the library upon experimental induction of Cas9 in a proportion consistent with the composition of the TLA with rare overrepresented sequences. We found that TARDIS is also compatible with the integration of large promoters and can be used to simultaneously integrate promoters into multiple genomic locations, providing a tool for multiple insertions at defined locations across the genome. While we demonstrate the system's advantages in *C. elegans*, in principle, the system is adaptable for any situation where the sequences for integration can be introduced with high diversity and heritability, and where a genomic site for integration can be made or is available.

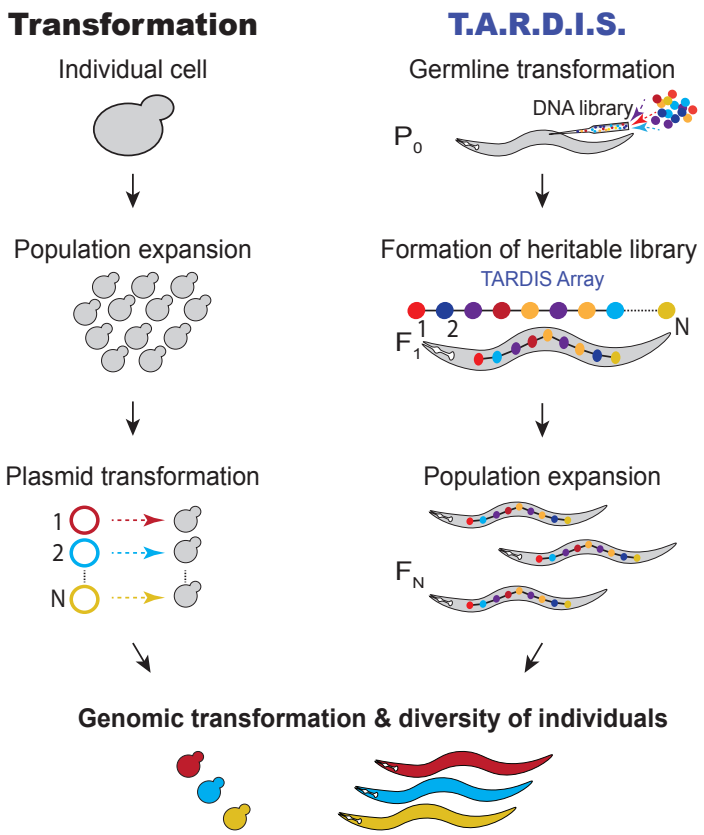


Figure 1: **Transformation compared to TARDIS**. For transformation, a large population of cells are individually transformed with a DNA library, resulting in a diverse population of individuals. TARDIS achieves a diversity of individuals by splitting transgenesis into two separate processes: (1) The introduction of a diverse library, which is formed into a TARDIS library array, passed down to future generations and thus replicated; and (2) an event that triggers the integration a sequence from the library at random, resulting in a diversity of integrated sequences.

Results

Generation of barcode landing pad

We designed a specific landing pad for the introduction and selection of small barcode fragments from high-diversity, multiplexed barcode libraries (Figure 2). This landing pad was designed to be targeted by Cas9 and requires perfect integration on both the 5' and 3' ends of a synthetic intron for functional hygromycin B resistance. Current split selection landing pads only provide selection on one side of the double-strand break, which can result in a small percentage of incomplete integrations (Stevenson et al., 2020). To fully test a large library approach, the requirement of genotyping to identify correct integrations must be overcome. A split-selection, hygromycin resistance (HygR) system was chosen for its simplicity and integration-specific selection. A unique synthetic CRISPR guide RNA target sequence was created by removing coding sequence on both sides of an artificial intron, resulting in a non-functional HygR gene. By removing critical coding sequence on both sides of the gene, only 'perfect' integration events will result in hygromycin resistance (Figure 2A). The synthetic landing pad was integrated at Chromosome II: 8,420,157, which has previously been shown to be permissive for germline expression (Dickinson et al., 2015; Frøkjær-Jensen et al., 2012, 2008).

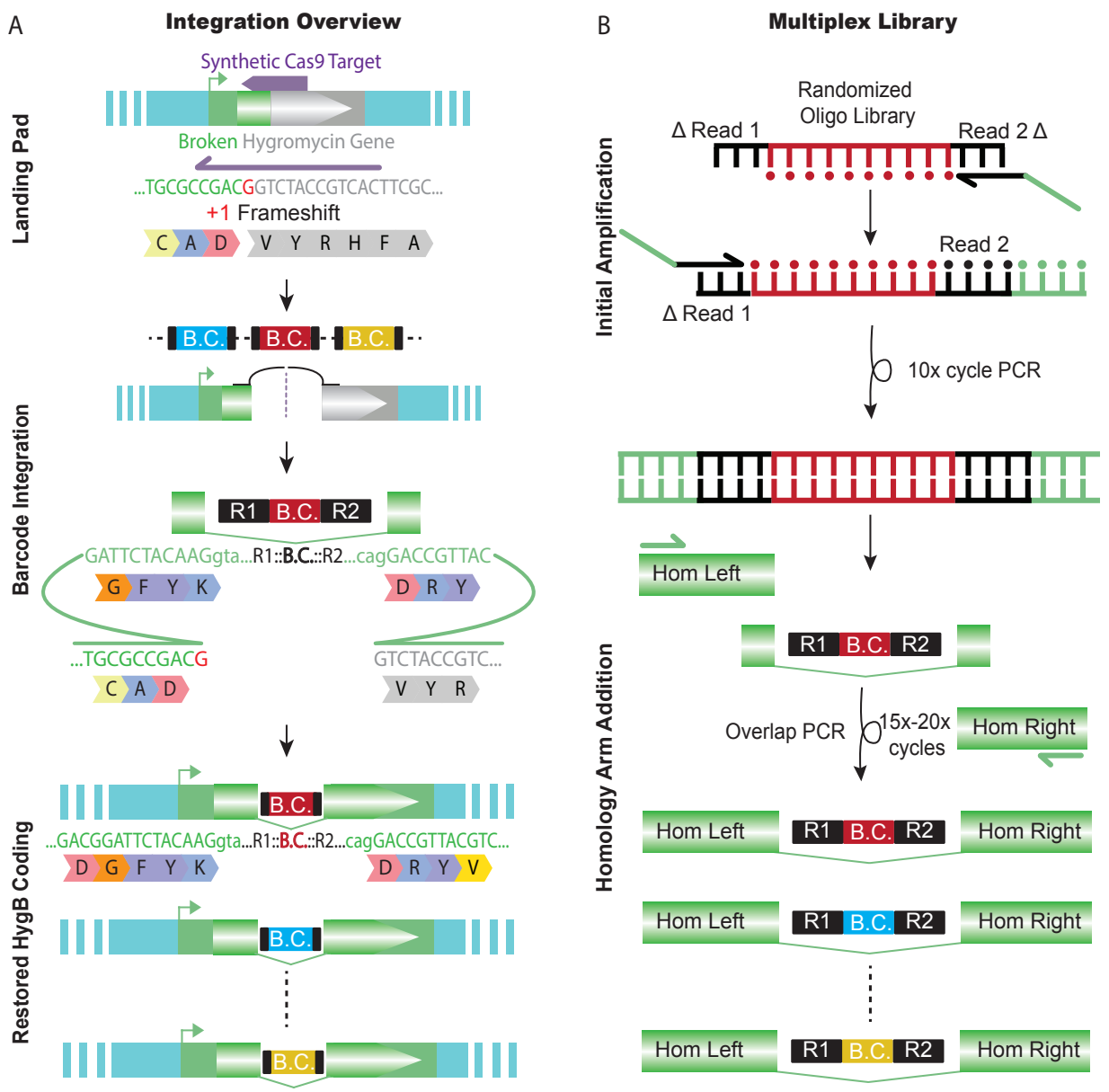


Figure 2: Barcode landing pad and diverse donor library. A) Schematic design for the barcode landing pad and integration. A broken hygromycin resistance gene is targeted by Cas9, which repairs off the TARDIS array, integrating a barcode and restoring the functionality of the gene. B) The TARDIS multiplex library was created from a randomized oligo library, which underwent 10 cycles of PCR to make a dsDNA template. The barcode fragment was then added into a three fragment overlap PCR to add homology arms and make the final library for injection.

Generation of high diversity donor library and TARDIS arrays

Transgenes or DNA sequences can be cloned into plasmid vectors for injections in *C. elegans*. However, the cloning process is laborious, and the plasmid vector is unnecessary for integration into an array or the genome. We sought to provide a protocol for library generation that maximized diversity and eliminated the requirement of cloning (Figure 2B). Oligo libraries have been used for barcoding (Levy et al., 2015) and for identification of promoter elements (de Boer et al., 2020) in yeast, but practical implementation of large synthetic libraries for transgenesis has never been performed in an animal system. We used randomized synthesized oligos to build a highly diverse library of barcodes, similar to the one described by Levy et. al. (2015), via complexing PCR. Given randomized bases present at the 11 nucleotide positions centrally located within the barcode, our base library can yield a theoretical maximum of

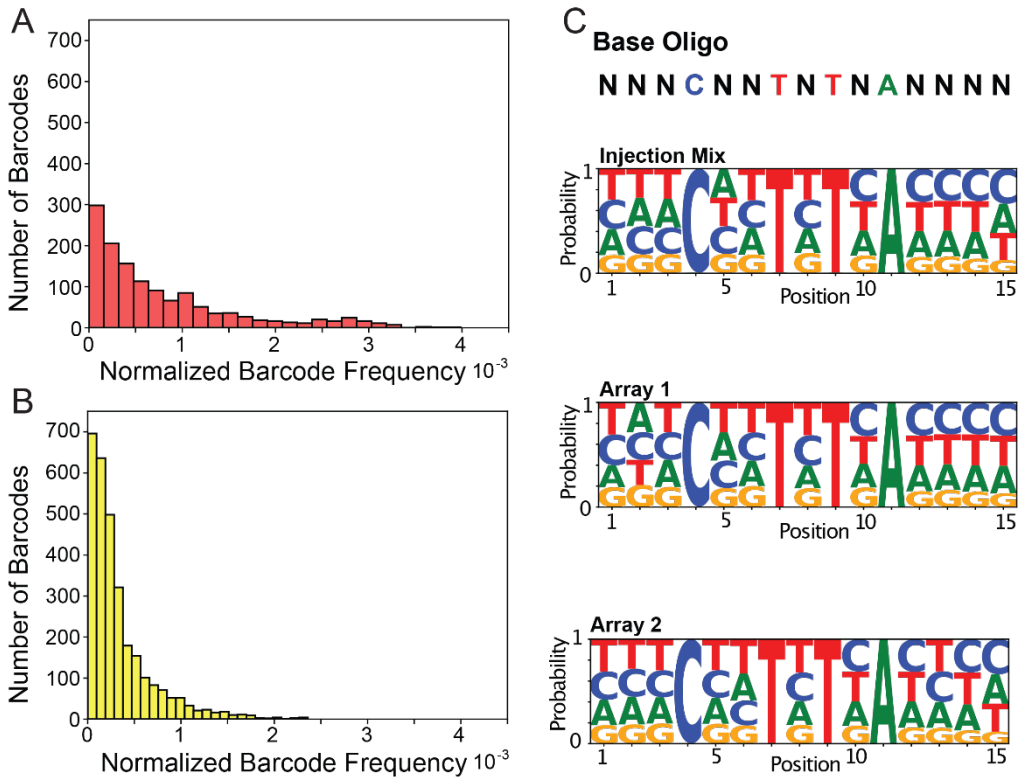


Figure 3. TARDIS Library Arrays can contain large barcode diversity. A) Frequency distribution of 1,319 unique barcodes in array 1 (PX816). B) Frequency distribution of the 3,001 unique barcode sequences in array 2 (PX817). C) Sequence logo probabilities of the 15 base pair positions of the barcodes in the injection mix, array 1 and array 2.

approximately 4.2 million sequences. Our overlap PCR approach achieves high levels of diversity with minimal 'jackpotting' — sequences with higher representation than expected (Figure 3—figure supplement 1). With low coverage sequencing, we found almost 800,000 unique barcode sequences, providing a large pool of potential sequences that can be incorporated into TARDIS arrays. Only 472 sequences were overrepresented (counts greater than 50), accounting for approximately 6.7% of the total reads and only approximately 0.06% of the unique barcodes detected.

We injected our complexed barcodes and isolated individual TARDIS array lines, each containing a subset of the barcode library (Figure 3). Individual injected worms were singled, and we identified four arrays from three plates. Arrays 1 and 2 were identified on separate plates, and were therefore derived from independent array formation events, while array 3, profile 1 and array 3, profile 2 were both identified on the same plate. Analysis of array diversity within these lines shows, somewhat unexpectedly, that during array formation a subset of barcode sequences tended to increase in frequency (Figure 3A and B). Higher frequency barcodes in arrays tend to be independent of the jackpotted sequences of the injection mix as very few are represented in the set of high frequency barcodes from the injection mix. The high frequency barcodes also varied between arrays.

We found that array formation does not seem to favor any particular barcode sequence motif (Figure 3C) and that arrays can range considerably in diversity. Array 1 had 1,319 unique barcode sequences, array 2 had 3,001 unique barcode sequences, array 3 profile 1 had 91 unique barcode sequences, and array 3 profile 2 had 204 unique barcode sequences (Figure 3—figure supplement 2). Across the four arrays, we found a total of 4,395 unique barcode sequences. When we compared the individual sequences incorporated during the three independent injections, we found little overlap. 96.5% (4395/4553) of the identified sequences were unique to one injection, 3.0% (136) were incorporated twice, and 0.5% (22) were recovered from all three injections. In contrast to the diversity between injection events, a similar comparison of the two profiles derived from a single injection for array 3 showed considerable overlap, with 68% (62/91) of the profile 1 sequences also being present in profile 2. Overall, our results

suggest our complexing PCR oligo library can produce a highly diverse library and that arrays can store a large diversity of unique sequences.

The distribution of element frequency within a given array follows a clear Poisson distribution. Arrays 1 and 2 show more diversity, with barcode frequencies more similar to one another than the two profiles isolated from array 3 (Figure 3—figure supplement 2). The null assumption is that the array is formed from a simple sample of the injected barcodes in equal proportions. However, arrays have been already reported to jackpot certain sequences. For example, when Lin et. al. (2021) injected fragmented DNA, they found that larger fragments were favored in the assembly. In our case, we find some barcode sequences become jackpotted, despite being identical in size. A possible explanation is that early in formation, arrays are replicating sequences, possibly to reach a size threshold. Consistent with this hypothesis, arrays with higher barcode diversity had frequencies closer to one another, while arrays with lower diversity had wider frequency ranges.

Integration from TARDIS array to F1

Our primary motivation in developing the TARDIS method was to utilize individual sequences from the TARDIS array as integrated barcodes. To assay the integration efficiency, we performed TARDIS integration on two biological replicates from a TARDIS array line (PX786) synchronized in the presence of G-418. Out of the 100 L1's per plate initially plated on antibiotic free plates, an average of 41 worms (N=255 plates) for replicate one, and 62 worms for replicate two (N=125 plates), survived to the next day. These surviving individuals contained the array, allowing them to survive early life G-418 exposure, and generally showed fluorescent co-marker expression as well. Following heat shock to induce Cas9, replicate one produced 104 plates with hygromycin resistant individuals, indicating barcode integration, and replicate two produced 71. These results suggest that approximately 200-300 worms need to be heat shocked to obtain an integrated line when using 150bp homology arms and relatively small inserts such as the barcodes. To assay the integration frequency from the array to the F1, we performed TARDIS integration on four biological replicates derived from PX786. We found that frequency of integration for barcodes in F1 individuals was strongly correlated with the barcodes' frequency in the TLA (Figure 4A), ($R \approx 0.96$, $p \approx 5.7\text{x}10^{-154}$). Notably, there are two replicated outliers across the four biological replicates. One barcode (TTAAATTATCACATG), tended to integrate more often than would be predicted by its frequency in the array, while barcode (GCTCATTCTGACGTA) integrated less frequently than expected (Figure 4—figure supplement 1). In general, however, we did not observe any noticeable bias in sequence motif selection following integration (Figure 4B). Several individual lineages were isolated from the population with hygromycin selection, validating functional restoration of the *HygR* gene, and three were randomly chosen for Sanger sequencing to confirm perfect barcode integration. As expected, these sequenced barcodes were also found amongst the barcode sequences of the array.

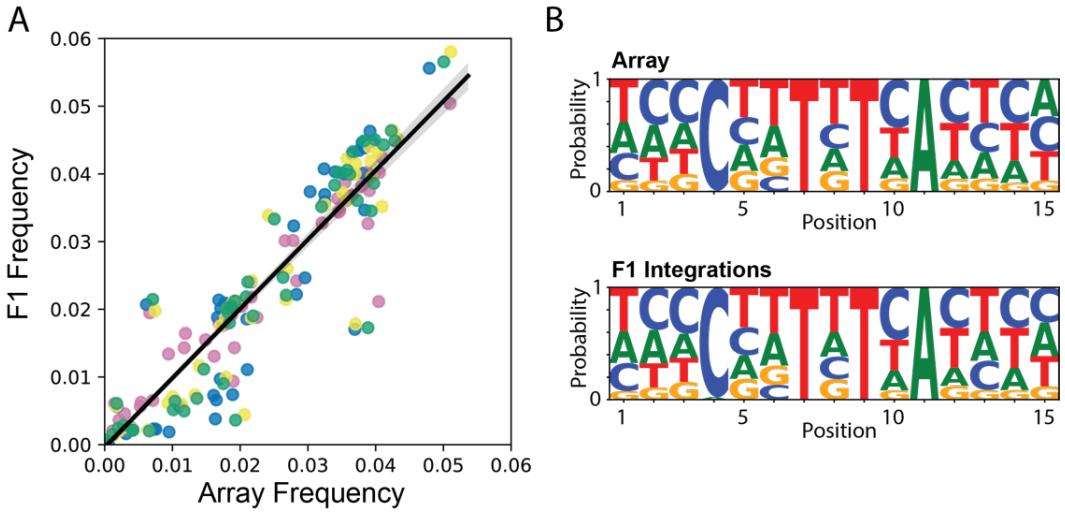


Figure 4: Integration frequency from TARDIS Library Array to F1. A) Frequency of integration from TARDIS Library Array to the F1, R ≈ 0.96, $p \approx 5.7\text{x}10^{-154}$. Different colors represent four biological replicates. Line shading represents 95% confidence interval. B) Sequence

probabilities of PX786 compared to the F1 integrations (91 unique barcodes were identified in the array and 118 in the F1s, with a five read threshold).

181

182 **Generation and integration of TARDIS promoter library**

183 For testing insertion of promoter libraries via TARDIS, two separate landing pad sites utilizing split selection were
184 engineered in chromosome II (Figure 5A). The first contained the 3' portions of both the *mScarlet-I* and the *HygR*
185 genes in opposite orientation to each other and separated by a previously validated synthetic Cas9 target (Stevenson
186 et al., 2020). Similarly, the second landing pad site contained 3' portions of *mNeonGreen* and *Cbr-unc-119(+)*
187 separated by the same synthetic Cas9 target, allowing both sites to be targeted by the same guide. These landing pads
188 were engineered into an *unc-119(ed3)* background to allow for selection via rescue of the Uncoordinated (Unc)
189 phenotype. A strain containing only the split *mScarlet-I*/split *HygR* landing pad was also constructed, in which case a
190 copy of *Cbr-unc-119(+)* was retained at the landing pad site. Repair templates contained the 5' portion of the
191 respective selective gene, a lox site allowing for optional removal of the selective gene after integration (by expression
192 of Cre) and the chosen promotors in front of the 5' portion of the respective fluorophore. The selective gene and
193 fluorophore fragments contained >500bp overlaps with the landing pad to facilitate homology directed repair. Correct
194 homology directed repair at both junctions resulted in worms that were fluorescent, hygromycin resistant and had wild
195 type movement.

196 The initial promoter library tested was composed of 13 promoters targeted to a single landing pad site with split
197 *mScarlet-I* and split *HygR* (Table 1). These promoters ranged in size from 330-5545bp (total repair template length of
198 2238-7453bp). Seven different array lines were generated which exhibited distinct profiles when probed by PCR as a
199 crude measure of array composition and diversity (Figure 5—figure supplement 1A). Promoter specific PCR showed
200 these arrays to contain 2-13 of the 13 injected promoters, with a mean of 10.7 and a median of 12 (Figure 5—figure
201 supplement 1B). For the selected line (PX819), 12 promoters were incorporated into the TARDIS array. From this
202 line, approximately 200 G-418 resistant L1s (ie. those containing the array) were plated onto each of 60 plates and
203 then heat shocked as L2/L3s to initiate integration. Hygromycin resistant individuals were recovered from 59 of the
204 60 plates, indicating one or more integration events on each of those plates. Four individuals were singled from each
205 of these plates, with the intent of maximizing the diversity of fluorescent profiles and analyzed by PCR to identify the
206 integrated promoters (Figure 5—figure supplement 1B). Based on the banding patterns, 83 of these PCR products
207 were sequenced with nine different promoters confirmed as integrated (Table 1 and Figure 5B). This included both
208 the smallest (*aha-1p*) and the largest promoter (*nhr-67p*) in the set. Notably, two of the three promoters that were in
209 the array but not recovered as integrants were found to be integrated in a subsequent experiment (see below),
210 suggesting the failure to be recovered in this case was likely due to the array composition rather than any properties
211 of these particular promoters. For approximately half of the plates, two or more promoters were identified from the
212 four worms chosen. Of the 83 PCR products sequenced, five had incorrect sequences and/or product sizes inconsistent
213 with the promoter identified and three failed to prime. Additionally, several samples failed to amplify or gave a non-
214 specific banding pattern and likely also represent incorrect integrations.

215 To test if TARDIS could be used to target multiple sites simultaneously, a second promoter library containing seven
216 promoters targeted to each site *(ahr-1p*, *ceh-10p*, *ceh-20p*, *ceh-40p*, *ceh-43p*, *hlh-16p*, *mdl-1p*) was injected into worms
217 containing both landing pad sites. Five plates of mixed stage worms were heat shocked, and worms that were both
218 hygromycin resistant and had wild-type movement were found on three of those plates. Worms that were hygromycin
219 resistant but retained the Unc phenotype were also observed on some plates, representing individuals with integrations
220 at a single site. For two of the plates a single pair of integrations was observed, in both cases being *ahr-1p::mScarlet*
221 plus *hlh-16p::mNeonGreen*. For the third plate, two different combinations were recovered: *ahr-1p::mScarlet* plus
222 *mdl-1p::mNeonGreen* and *ceh-40p::mScarlet* plus *ceh-10p::mNeonGreen* (Figure 5C). While multi-site CRISPR is
223 known to be possible (Arribere et al., 2014), these results suggest that TARDIS provides a unique way to engineer
224 multiple locations using a single injection.

225 When transcriptional reporter lines were examined by fluorescent microscopy, expression of the fluorophores was
226 concentrated in but not exclusive to the nucleus, consistent with the presence of nuclear localization signals (NLS) on
227 the fluorophores. For all promoters, expression was seen in at least one previously reported tissue (Table 1) but was

absent in one or more tissues for several of the promoters. Expression of single copy reporters is frequently more spatially restricted than that from integrated or extrachromosomal arrays (Aljohani et al., 2020). The differences in expression pattern may also reflect differences in the region used as the promotor or the fact that only a single developmental stage (late L4/early adult) was examined. Overall, we find that TARDIS can be used to screen functional libraries, either individually or in combination.

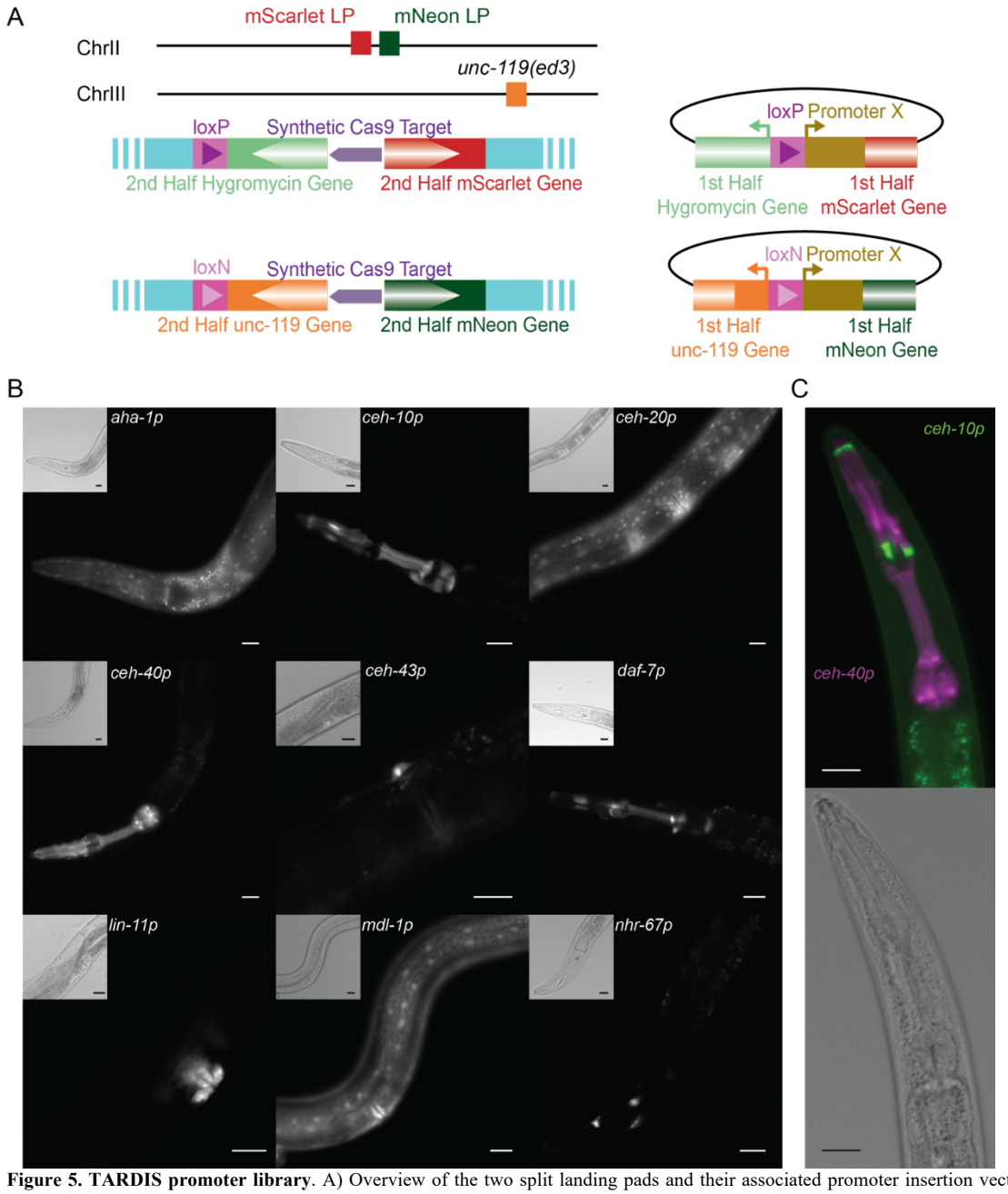


Figure 5. TARDIS promoter library. A) Overview of the two split landing pads and their associated promoter insertion vectors. Both the selective marker and the fluorophore expression are restored upon correct integration. B) Transcriptional reporters for nine genes were recovered from a single heatshock of a single TARDIS array line (PX819). Integration was into the single *mScarlet-I*/HygR landing pad. Main images show mScarlet-I expression for the indicated reporter while insets show polarized image of the same region. C) Example simultaneous, dual integration from a single TARDIS array into the double landing pad strain with PEST. *ceh-10p*::*mNeonGreen*::*PEST* is false colored green and *ceh-40p*::*mScarlet-I*::*PEST* is false colored magenta. All scale bars represent 20µm

Discussion

Here we present the first implementation of a practical approach to large-scale library transgenesis in an animal system (Figure 1). Building on over a half century of advancements in *C. elegans* genetics, we can now make thousands of specific, independent genomic integrations from single microinjection events that traditionally yield at most a small handful of transgenic individuals. Increasing transgenesis throughput has long been desired, and in *C. elegans* several attempts have been made to multiplex transgenic protocols. Library mosSCI and RMCE, which both introduce a multiplexed injection mixture and do indeed achieve multiple integrations (Kaymak et al., 2016; Nonet, 2020). However, just as in the case of standard mosSCI or single donor injections for RMCE, anti-array screening, genotyping, and the direct integration of the process limit the multiplex potential substantially. One group has adopted arrays with small pools of guides coupled with heatshock inducible Cas9 to produce randomized mutations at targeted locations (Froehlich et al., 2021). This protocol shares similarities with TARDIS, in that diverse arrays are coupled with inducible Cas9. However, the focus of that technology was to produce randomized genomic edits, and it does not produce precise, library integrations into the genome. Recently, another group (Mouridi et al., 2022) built on the utility of heatshock Cas9 and integrated three individual sequences from an array. While these prior multiplexed methods made substantial contributions in improving the efficiency of specific transgenesis, none have yet demonstrated multiplexing beyond tens of unique sequences—orders of magnitude below what would be needed for exploratory transgenesis. TARDIS therefore provides the first true library-based approach for multiplexing transgenesis in *C. elegans*.

TARDIS as a method for creating barcoded individuals

Genetic barcode libraries have been applied to many high-throughput investigations to reduce sequencing costs and achieve a higher resolution within complex pools of individuals. By focusing the sequencing reads on a small section of the genome, a larger number of individual variants can be identified or experimentally followed. This critical advancement has led to the widespread use of barcoding for evolutionary lineage tracking in microbial systems (Blundell and Levy, 2014; Kasimatis et al., 2021; Levy et al., 2015; Levy, 2016; Nguyen Ba et al., 2019; Venkataram et al., 2016)–uncovering the fitness effects of thousands of individual lineages without requiring large coverage depth of the whole genome. In addition to this application, using barcoded individuals can be used to facilitate any application that involves screening a large pool of diverse individuals within a shared environment. For example, barcodes have been used in microbial studies investigating pharmaceutical efficacy (Smith et al., 2011) and barcoded variant screening (Emanuel et al., 2017). The TARDIS-based system presented here provides an approximately 1,000X fold increase in barcoding throughput in *C. elegans*, making it a unique resource among multicellular models that allows the large diversity pool and design logic of microbial systems to be adapted to animal models.

While we designed our barcode sequence units for the purpose of barcoding individuals, this approach could also prove useful in future optimization and functional understanding of array-based processes. In particular, the high sequence diversity but identical physical design of the synthetic barcode library may provide a unique window into extrachromosomal array biology that would be helpful in optimizing sequence units for incorporation into heritable TLAs. For example, an unexpected result of the barcoding experiment was the discovery that a small minority of sequences were overrepresented, or 'jackpotted,' in the TLA relative to their frequency in the injection mix (Figure 3 and Figure 3—figure supplement 1). Our expectation was that arrays would form in an equal molar fashion proportional to the injection mix based on the model that arrays are formed by physical ligation of the injected DNA fragments (Mello et al., 1991). Deviations from random array incorporation have been observed before, and a bias for incorporating larger fragments has been proposed as an explanatory mechanism (Lin et al., 2021). Our results suggest that the ultimate array composition is not directly proportional to the molarity of the injected fragments, or strictly weighted towards the size of the fragment as has been suggested. In contrast, we propose that array size affects the maintenance of extrachromosomal arrays. As such, selection can act to increase the rate of recovery for arrays that have increased their size through random amplification of some sequences by an unknown process early in the formation of the array or by expansion of similar sequences by DNA polymerase slippage during replication, as has been well documented for native chromosomes (Levinson and Gutman A., 1987). These hypotheses would be consistent with observations of Lin et. al. (2021), if the underlying mechanism for their observation is that inclusion of larger fragments tends to be positively correlated with ultimate array size, and therefore likelihood of maintenance.

TARDIS as a method for the introduction of promoters and other large constructs

While the barcode approach demonstrates the potential for using TARDIS to integrate large numbers of 433bp PCR products, previous work using CRISPR/Cas9 initiated homology directed repair has suggested that integration efficiencies decrease with the size of the insert (Dickinson and Goldstein, 2016). We therefore implemented TARDIS for integrating promoters cloned into a vector backbone and ranging in size from 330bp to 5.5kb, to determine TARDIS functionality under a physically different use case directed specifically at functional analysis. We found that promoter libraries could be integrated into either single sites or two sites simultaneously. Unsurprisingly, the frequency at which various promoters were recovered varied from array to array (for example *ahr-1p* was never recovered in the single site integration experiment, despite being present in the array, while it was the most common promoter recovered in the two-site integration experiment) and likely reflects the same relationship between integration frequency and prevalence in the array, as was seen with the distribution of insert abundance for the barcodes. While we showed that plasmid donors can be used in the TARDIS pipeline, neither of our two arrays contained all 13 plasmids. Given that the estimated 1-13MB size of arrays (Carlton et al., 2022) would be adequate to hold copies of each of the plasmids, as well as the extreme diversity obtained when using smaller DNA fragments, differential presence of a given promotor fragment was somewhat unexpected. This may reflect a preferential use of linear fragments in the *in-situ* assembly of arrays. Future use of linear fragments where feasible may increase incorporation and overall diversity (Priyadarshini et al., 2022).

For both the one and two site promoter library integrations, transgenic individuals were readily detected, suggesting the TARDIS method for integration was highly efficient. It has long been understood that successful CRISPR editing at one site significantly increases the chances of successful editing at a second site. This is the premise behind commonly used co-conversion screening strategies (also referred to as co-CRISPR), such as the *dpy-10* screen commonly used in *C. elegans* (Arribere et al., 2014; Ward, 2015). Here we show that same type of co-conversion also occurs when using only "large" (>1kb), plasmid-based repair templates containing gene-sized repair constructs. Additionally, we have simultaneously targeted the same two landing pads presented here using standard CRISPR techniques and find that approximately half of hygromycin resistant individuals also have rescue of the Unc phenotype (i.e. editing has occurred at both sites; data not shown). Given the high rate of co-conversion, this work demonstrates multiplex integrations are possible not only by targeting multiple repair templates to a single site but also by simultaneously utilizing multiple insertion sites.

In order to recover individual edits most efficiently, given the high frequency of integration using TARDIS, we recommend to either heat shock small cohorts of array bearing individuals, such that most cohorts only yield one edited individual or to screen multiple individuals per cohort. Additionally, while split-selection methods allow for direct verification of integration, depending on the downstream use case, integrations should be confirmed by sequencing as errors can still occur, including internal deletions within the insert.

Expansion of TARDIS to other multicellular systems

Unlocking the investigative potential of transgenesis in animal systems would enable exploratory experiments normally restricted to single-cell models. For example, alanine scanning libraries and protein-protein interactions (Cunningham and Wells, 1989; Matthews, 1996; Wells, 1991), CRISPR library screening (Bock et al., 2022), and promoter library generation (Delvigne et al., 2015; Zaslaver et al., 2006). While we demonstrate the use of TARDIS in *C. elegans* here, the intellectual underpinnings of the approach are agnostic to the research model used. Conceptually, TARDIS facilitates high-throughput transgenesis by using two engineered components: a heritable TARDIS Library containing multiplexed transgene-units and a genomic split selection landing pad that facilitates integration of single sequence units from the library. To generate the first TARDIS libraries, we capitalized on the endogenous capacity of *C. elegans* to assemble experimentally provided DNA into heritable extrachromosomal arrays. Extrachromosomal arrays are formed from exogenous DNA, are megabases in size (Lin et al., 2021; Woglar et al., 2020), do not require specific sequences to form and replicate, and can be maintained in a heritable manner via selection (Mello et al., 1991). These qualities make them suitable for use as a heritable library upon which TARDIS can be based. To adopt TLAs in systems beyond *C. elegans*, methods must be adopted to introduce large heritable libraries into the germline, as most systems do not maintain extrachromosomal arrays. In mice, the locus *H11* has been used for large transgenic insertions (Liu et al., 2022), while in *Drosophila*, the use of PhiC31-mediated transgenesis coupled with bacterial artificial chromosomes (BACs) have allowed for many approximately 10kb+ sized fragments

to be integrated into their respective genomes (Venken et al., 2006). Each of these large integration strategies can provide a vehicle for stable inheritance of a TLA.

The second component of the TARDIS integration system is a pre-integrated landing pad sequence. We have generally favored split selectable landing pads (SSLPs) that use HygR for its effectiveness (Mouridi et al., 2022; Stevenson et al., 2020, 2021). The SSLPs are engineered to accept experiment-specific units from the array. For example, here we used SSLPs designed to accept barcodes for experimental lineage tracking and promoters for generation of transcriptional reporters. To translate TARDIS to other systems, a genomic site needs to be engineered to act as a landing pad that can utilize sequence units from the TLA and can be customized to the specific system and use. Because TLAs allow the experimenter to design the library of interest and the landing pad to recapitulate the strengths of single-cell systems, adoption of TARDIS in multicellular animal experiments can leverage the high-resolution, high-diversity exploratory space of DNA synthesis. In addition to adapting assays currently restricted to single-cell models, TARDIS also opens the door to animal-specific uses, such as developmental biology, neurobiology, endocrinology, and cancer research.

In developmental genetics, the lack of large-library transgenesis has resulted in 'barcode' libraries in a different form, utilizing randomized CRISPR-induced mutations to form a unique indel. For example, GESTALT (McKenna et al., 2016) creates a diversity of barcodes *in-vivo* via random indel formation at a synthetic target location. LINNAEUS (Spanjaard et al., 2018) similarly utilizes randomized targeting of multiple RFP transgenes to create indels, allowing for cells to be barcoded for single cell sequencing. TARDIS barcodes do not rely on randomized indel generation and thus can be much simpler to implement with sequencing approaches outlined above.

In-vivo cancer models have also adopted the high-resolution, high-variant detection of barcodes for the study of tumor growth and evolution. Rogers at al., developed Tuba-seq (Rogers et al., 2017; Winslow, 2022), a pipeline that takes advantage of small barcodes allowing for *in-vivo* quantification of tumor size. In Tuba-seq, barcodes are introduced via lentiviral infection, leading to the barcoding of individual tumors. TARDIS brings the multiplexed library into the animal context without requiring viral vectors or intermediates, thereby allowing large *in-vivo* library utilization and maintenance. Capitalizing on the large sequence diversity possible within synthesized DNA libraries with a novel application in multicellular systems generates new opportunities for experimental investigation in animal systems heretofore only possible within microbial models.

Summary

In summary, here we have presented Transgenic Arrays Resulting in Diversity of Integrated Sequences (TARDIS), a simple yet powerful approach to transgenesis that overcomes the limitations of multicellular systems. TARDIS uses synthesized sequence libraries and inducible extraction and integration of individual sequences from these heritable libraries into engineered genomic sites to increase transgenesis throughput up to 1000-fold. While we demonstrate the utility of TARDIS using *C. elegans*, the process is adaptable to any system where experimentally generated genomic loci landing pads and diverse, heritable DNA elements can be generated.

Materials and Methods

Table 2: Key Reagents.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Genetic reagent (*C.elegans*)	*aha-1p*	wormbase.org	WBGene00000095	
Genetic reagent (*C.elegans*)	*hlh-16p*	wormbase.org	WBGene00001960	

Genetic reagent (*C. elegans*)	*ceh-40p*	wormbase.org	WBGene00000461	
Genetic reagent (*C. elegans*)	*ceh-10p*	wormbase.org	WBGene00000435	
Genetic reagent (*C. elegans*)	*ahr-1p*	wormbase.org	WBGene00000096	
Genetic reagent (*C. elegans*)	*mdl-1p*	wormbase.org	WBGene00003163	
Genetic reagent (*C. elegans*)	*egl-43p*	wormbase.org	WBGene00001207	
Genetic reagent (*C. elegans*)	*ceh-20p*	wormbase.org	WBGene00000443	
Genetic reagent (*C. elegans*)	*ceh-43p*	wormbase.org	WBGene00000463	
Genetic reagent (*C. elegans*)	*daf-7p*	wormbase.org	WBGene00000903	
Genetic reagent (*C. elegans*)	*lin-11p*	wormbase.org	WBGene00003000	
Genetic reagent (*C. elegans*)	*egl-46p*	wormbase.org	WBGene00001210	
Genetic reagent (*C. elegans*)	*nhr-67p*	wormbase.org	WBGene00003657	
Strain, strain background (*Caenorhabditis elegans*)	*N2*	Caenorhabditis Genetics Center		
Strain, strain background (*Caenorhabditis elegans*)	*N2-PD1073*	doi: 10.17912/micropub.biology.000518		Available from the Caenorhabditis Intervention Testing Program- upon request (https://citp.squarespace.com/)-
Strain, strain background (*Caenorhabditis elegans*)	PX740	This paper		N2-PD1073 fxIs47 [*rsp-0p:: 5' ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3' ΔHygR::unc-54 3'::loxP*]
Strain, strain background (*Caenorhabditis elegans*)	GT331	This paper		aSi9[*lox2272 Cbr-unc-119(+) lox2272 + loxP 3'3' ΔHygR + 3' ΔmScarlet-I::PEST*]*; unc-119(ed3)*
Strain, strain background	GT332	This paper		aSi10[*lox2272 Cbr-unc-119(+) lox2272 + loxP 3' ΔHygR + 3' ΔmScarlet-I*]*; unc-119(ed3)*

(*Caenorhabditis elegans*)				
Strain, strain background (*Caenorhabditis elegans*)	GT336	This paper		aSi12*[lox2272 rps-0p::HygR + hsp-16.41p::Cre::tbb-2 3'UTR + sqt-1(e1350) lox2272 + loxN 3' ΔCbr-unc-119(+)::tjp2a_guide:: 3' ΔmNeonGreen::PEST::egl-13nls::tbb-2 3'UTR]* aSi9*[lox2272 Cbr-unc-119(+) lox2272 + loxP 3'ΔHygR::tjp2a guide::3'ΔmScarlet-I::PEST::egl-13nls::tbb-2 3'UTR] II; unc-119(ed3) III*
Strain, strain background (*Caenorhabditis elegans*)	GT337	This paper		aSi13[*lox2272 + loxN 3' ΔCbr-unc-119(+) + 3' ΔmNeonGreen::PEST]* aSi14*[lox2272 + loxP 3' ΔHygR + 3' ΔmScarlet-I::PEST*]*; unc-119(ed3),*
Strain, strain background (*Caenorhabditis elegans*)	QL74	Gift from QueeLim Ch'ng		oxEx1578 [*eft-3p::GFP + Cbr-unc-119(+)*] 6x outcross EG4322
Strain, strain background (*Caenorhabditis elegans*)	PX786	This paper		fxEx23 [TARDIS #5 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Read 2::Intron3′:: 3′ ΔHygR* (89 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3'::loxP*
Strain, strain background (*Caenorhabditis elegans*)	PX816	This paper		fxEx25 [TARDIS #1 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Rea d2::Intron3′:: 3′ ΔHygR* (1,319 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3'::loxP*
Strain, strain background (*Caenorhabditis elegans*)	PX817	This paper		fxEx26 [TARDIS #2 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Rea d2::Intron3′:: 3′ ΔHygR* (3,001 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3'::loxP*
Strain, strain background (*Caenorhabditis elegans*)	PX818 profile 1	This paper		fxEx27 [TARDIS #3 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Rea d2::Intron3′:: 3′ ΔHygR* (91 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR::unc-54 3'::loxP*
Strain, strain background (*Caenorhabditis elegans*)	PX818 profile 2	This paper		fxEx28 [TARDIS #4 *5′ ΔHygR::Intron5′::Read1::NNNCNNTNTNANNNN::Rea d2::Intron3′:: 3′ ΔHygR* (204 Unique Sequences) *hsp-16.41p::piOptCas9::tbb-2 34′ UTR + rsp-27p::NeoR::unc-54 3′ UTR + U6p:: GCGAAGTGACGGTAGACCGT*]; fxSi47[*rsp-0p:: 5′ ΔHygR:: GCGAAGTGACGGTAGACCGT :: 3′ ΔHygR'::unc-54 3'::loxP*

Strain, strain background (*Caenorhabditis elegans*)	PX819	This paper		N2 fxEx24 [(*rps-0p:: 5' ΔHygR +loxP + aha-1p::SV40 NLS:: 5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + ahr-1p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + ceh-10-1p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + ceh-20p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + ceh-40p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: ΔHygR +loxP + ceh-43p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + daf-7p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: ΔHygR +loxP + egl-43p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + hlh-16p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + lin-11p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + mdl-1p::SV40 NLS::5' ΔmScarlet-I) + (rps-0p:: 5' ΔHygR +loxP + nhr-67p::SV40 NLS::5' ΔmScarlet-I) + hsp-16.41p::piOptCas9::tbb-2 34' UTR + prsp-27::NeoR::unc-54 3' UTR]; aSi10[lox2272 + Cbr-unc-119(+) + lox2272 + loxP + 5' ΔHygR::unc-54 3' UTR + 5' ΔmScarlet-I::egl-13 NLS::tbb-2 3' UTR, II:*8420157]*; unc-119(ed3) III*
Strain, strain background (*Caenorhabditis elegans*)	EG4322	doi.org/10.1038/ng.248; Caenorhabditis Genetics Center		
Strain, strain background (*Escherichia coli*)	PXKR1	This paper		NA22 transformed with pUC19
Recombinant DNA reagent	Plasmid pDSP15	This paper	193853 (Addgene)	*5' ΔHygR::loxP::MCS::5' Δ mScarlet-I*
Recombinant DNA reagent	Plasmid pDSP16	This paper	193854 (Addgene)	*5' ΔCbr-unc-119(+)::loxN::MCS::5' Δ 5'mNeonGreen*
Recombinant DNA reagent	Plasmid pMS84	This paper	193852 (Addgene)	*U6p*::GGACAGTCCTGCCGAGGTGG
Recombinant DNA reagent	Plasmid pZCS36	This paper	193048 (Addgene)	*hsp16.41p::Cas9(dpiRNA)::tbb-2 '3UTR*
Recombinant DNA reagent	Plasmid pZCS38	This paper	193049 (Addgene)	*rsp-27p::NeoR::unc-54 3' UTR*
Recombinant DNA reagent	Plasmid pZCS41	This paper	193050 (Addgene)	*U6p*::GCGAAGTGACGGTAGACCGT
Sequence-based reagent	ZCS422	This paper		Design and construction of barcode donor library
Commercial assay kit	DNA Clean and Concentrator	Zymo Research	Cat No.: D4004	
Commercial assay kit	Genomic DNA Clean and Concentrator	Zymo Research	Cat. No.: D4011	

Commercial assay kit	Zymoclean Gel DNA Recovery Kit	Zymo Research	Cat. No.: D4008	
Commercial assay kit	Zyppy Plasmid Miniprep Kit	Zymo Research	Cat. No.: D4019	
Software, algorithm	Cutadept	doi.org/10.14806/ej.17.1.200	Version 4.1	
Software, algorithm	AmpUMI	doi.org/10.1093/bioinformatics/bty264	Version 1.2	
Software, algorithm	Starcode	doi.org/10.1093/bioinformatics/btv053	Version 1.4	
Software, algorithm	Google colab	colab.research.google.com		
Software, algorithm	python (version)	Guido van Rossum, 1991	Version 3.7.13	
Software, algorithm	Juypter notebook (IPython)	doi:10.3233/978-1-61499-649-1-87	Version 7.9.0	
Software, algorithm	matplotlib	doi.org/10.5281/zenodo.3898017	Version 3.7.13	
Software, algorithm	Fiji	imagej.net/software/fiji/	Version 2.9.011.53t	
Chemical compound, drug	G-418	GoldBio (CAS Number 108321-42-2)	Cat No.: G-418-5	
Chemical compound, drug	Hygromycin B	GoldBio (CAS Number 31282-04-9)	Cat No.: H-270-10-1	

General TARDIS reagents

Strains generated for this publication along with key plasmids and reagents are listed in the Key Resources Table. A full list of all plasmids is given in Supplementary file 1. All plasmids were cloned by Gibson Assembly following the standard NEB Builder HiFi DNA Assembly master mix protocol [New England Bio Labs (NEB), Massachusetts, USA], unless otherwise indicated. All plasmids have been confirmed by restriction digest, Sanger sequencing, and/or full plasmid sequencing. All primers used in the construction and validation of plasmids are listed in Supplementary file 2.

To generate our heatshock inducible Cas9, *hsp16.41p::Cas9dpiRNA::tbb-2 '3UTR,* the *hsp16.41* promoter was amplified from pMA122 (Addgene ID34873) (Frøkjær-Jensen et al., 2012). The germline licensed Cas9 and tbb-2 3' UTR were amplified from pCFJ150-Cas9 (dpiRNA) (Addgene ID107940) (Zhang et al., 2018). All fragments were assembled into PCR linearized pUC19 vector (NEB) to give the final plasmid pZCS36.

To generate a standard empty guide vector, *U6p*::(empty)gRNA, the U6p and gRNA scaffold from pDD162 (Addgene ID47549) (Dickinson et al., 2015) was amplified and assembled into PCR linearized pUC19 to generate pZCS11.

382 To generate *rsp-27p::NeoR::unc-54 3' UTR,* the full resistance cassette was amplified from pCFJ910 (Addgene
383 ID44481) and assembled into PCR linearized pUC19 vector to give pZCS38.

384 **Genomic DNA isolation for array and integrant characterization**

385 For processing large populations of worms, a widely used bulk lysis protocol was adapted (Fire Lab 1997 Vector
386 Supplement, February 1997). In brief, 450µl of worm lysis buffer (0.1M Tris-Cl pH8.0, 0.1M NaCl, 50mM EDTA
387 pH8.0, and 1% SDS) and 20µl 20mg/ml proteinase K were added to approximately 50µl of concentrated worm pellet.
388 Samples were inverted several times to mix and incubated at 62C° for 2 hours. After incubation, samples were checked
389 under the microscope to ensure no visible worm bodies were left in the solution. Chip DNA binding buffer (Zymo,
390 California, USA) was added in a 2:1 ratio and gently inverted to mix. Samples were then purified with Zymo -Spin
391 IIC-XLR columns following manufacture protocol. Samples were eluted in 50µl of water. Each sample was then
392 digested with 10mg/ml RNase A (ThermoFisher Scientific Massachussets, USA, Cat. No. EN0531) at 42C° for 2
393 hours. Genomic DNA was then reisolated by adding a 2:1 ratio of Chip DNA binding buffer and purifying with Zymo-
394 Spin IIC-XLR columns. Final genomic samples were quantified by Nanodrop.

395 For individual worm lysis, individual array bearing worms were isolated and lysed in 4µl of EB (Zymo, Cat. No.:
396 D3004-4-16) buffer with 1mg/ml proteinase K (NEB). Each sample was rapidly frozen in liquid nitrogen and then
397 thawed to disrupt the cuticle and then incubated at 58C° for 1 hour, with a subsequent incubation at 95C° for 20
398 minutes to inactivate the proteinase K.

399 **TARDIS integration- general protocol**

400 On Day 0, TARDIS array bearing *C. elegans* grown to a high density of gravid adults were hypochlorite synchronized
401 in NGM buffer (Leung et al., 2011) and grown overnight in 15ml NGM with G-418 (1.56mg/ml) at 15C° with nutation.
402 On Day 1, L1s were washed three times with NGM buffer to remove G-418, plated onto media without selective agent
403 and continued to be grown at 15°C. On Day 2, L2/L3s were heat-shocked at 35.5°C for one hour. After heat shock,
404 worms were grown at 25°C until gravid adults when Hygromycin B was top spread on plates at a final concentration
405 of 250µg/ml.

406 **Construction of landing pad for barcodes**

407 To create the barcode landing pad, an intermediate Chr. II insertion vector, pZCS30 was built from pMS4 by using
408 PCR to remove the *let-858* terminator. pZCS30 served as the vector backbone for pZCS32. To assist in cloning, the
409 backbone was split into two PCR fragments. The broken *HygR* gene was amplified in two parts, *rsp-0p::5'ΔHygR* and
410 *3'ΔHygR::unc-54* 3' UTR, from pCFJ1663 (Addgene ID51484). Overlapping PCR was used to fuse both *HygR*
411 fragments. The resulting broken *HygR* cassette removed the intron found in pCFJ1663 as well as four codons from
412 exon one and three codons from exon two, while also creating +1 frameshift and a reverse orientation guide RNA
413 target for pZCS41. A second overlapping PCR was performed to fuse the broken *HygR* cassette to backbone fragment
414 two. The resulting two-part clone was then assembled to give pZCS32.

415 The barcode landing pad TARDIS strain, PX740 was created by injecting a mixture of 10ng/µl pZCS32, 50ng/µl
416 pMS8, and 3ng/µl pZCS16 (Addgene ID154824) (Stevenson et al., 2020) into the gonad of young adult N2-PD1073
417 (Teterina et al., 2022) hermaphrodites. Screening and removal of the SEC were performed following (Dickinson et
418 al., 2015). Presence of the correct insertion was confirmed by Sanger sequencing using the primers listed in
419 Supplementary file 3.

420 To create the barcode landing pad targeting guide RNA, *U6p*:: GCGAAGTGACGGTAGACCGT, the guide sequence
421 GCGAAGTGACGGTAGACCGT was added by overlapping primers to the vector pZCS11 to give the final construct
422 pZCS41.

423 **Design and construction of barcode donor library**

424 Oligo ZCS422 was ordered with 11 randomized N's (hand mixed bases) [Integrated DNA Technologies (IDT), Iowa,
425 USA] and has the following sequence: CTACACGACGCTCTTCCGATCTNNNCNNTNTNANNNNAGATCG
426 GAAGAGCACACGTCTG. Four 'hard-coded' base pairs were included within the randomized sequence. ZCS422
427 was used as the core for the generation of two separate complexing PCR barcode homologies referred to as "barcode-

428 15X" and "barcode-20X" to denote the number of complexing cycles (Figure 2). All PCRs were performed using the
429 high-fidelity Q5 polymerase (NEB) as per manufacture instructions. All primers used for barcode synthesis can be
430 found in Supplementary file 4. For both "barcode-15X" and "barcode-20X, the left and right homology arms were
431 generated separately by PCR and purified by gel extraction. An initial 10 cycle PCR was performed to convert the
432 oligo into a 201bp double stranded product which was gel extracted with Zymo clean Gel DNA Recovery Kit (Cat.
433 No.: D4008) following manufacture protocol. The low cycle number was done to retain diversity and to minimize
434 effects of PCR jackpotting.

435 For "barcode-15X", to generate the complete donor homology, the double stranded barcode template was combined
436 with both the left and right homology arms for a three-fragment overlap PCR. To maximize diversity, high
437 concentrations of the individual templates were used. The reaction contained 52 fmol/µl of barcode template and 22
438 fmol/µl of left right arms in a single 50µL Q5 reaction. A total of 15 cycles were performed. The lower cycle was
439 again done to reduce PCR jackpotting. The single product was gel extracted as a 433bp fragment. The final donor
440 fragment is referred to as 'barcode-15X.'

441 To generate "barcode-20X" a similar three-fragment overlap PCR was used. 4.3fmol/µl of barcode template,
442 15.33fmol/µl of left arm, and 3.3fmols/µl of right arm were combined across six Q5 50ul reactions and a total of 20
443 cycles were performed. The right arm concentration was lower caused by low concentration extraction. The single
444 product was gel extracted as a 433bp fragment. The final donor fragment is referred to as 'barcode-20X.'

445 **Generation of barcode TLA lines**

446 The TARDIS array bearing line PX786 was created by injecting 50ng/µl of barcode-15X, 10ng/µl pZCS38, 15ng/µl
447 pZCS41, 5ng/µl pZCS16, and 20ng/µl pZCS36 into young adult PX740 hermaphrodites. Individual injected worms
448 were grown at 15C° for four days and then treated with G-418 (1.56mg /ml). A single stable array line was isolated
449 and designated PX786.

450 The TARDIS array bearing lines PX816, PX817, PX818 profile 1 and PX818 profile 2 were created by injecting
451 100ng/µl of barcode-20X, 10ng/µl pZCS38, 15ng/µl pZCS41 and 20ng/µl pZCS36. Individual injections were grown
452 at 15°C for four days and then treated with G-418 (1.56mg/ml). Full genotypes are provided in Supplementary file 7
453 as the full genotypes cannot be contained within a table.

454 **Estimation of barcode integration frequency population sample preparation**

455 PX786 was grown to gravid adults in the presence of G-418 with concentrated NA22 transformed with pUC19 for
456 ampicillin resistance as a food source (designated PXKR1). Once gravid, the strain was hypochlorite synchronized
457 and grown overnight in 15ml NGM buffer with G-418 at 15°C with nutation. For each of the four replicates, a
458 synchronized L1 population was divided in half. The first half was pelleted by centrifugation (2,400xg for two
459 minutes) and frozen (-20°C) until processed. These samples represented the array-bearing samples. Another sample
460 of approximately 150,000 L1s was plated to large NGM and subjected to the standard TARDIS heat shock and grown
461 until the population was primarily gravid adults. Then, this population was hypochlorite synchronized and grown in
462 NGM buffer at 15°C with Hygromycin B (250 µg/ml). These entire samples were pelleted and frozen, representing
463 the F1 samples.

464 **PCR for barcode quantification**

465 Several different PCRs were performed depending on the intended downstream sequencing quantification. See Figure
466 2-figure supplement 1 for a schematic layout of the different PCR steps. The primers used for barcode quantification
467 are given in Supplementary file 5. To quantify the diversity of arrays from either a bulk population or individual
468 worms, two separate PCRs were performed to quantify the diversity of arrays.

469 The first PCR (Amplicon one-array) was performed for three cycles to add Unique Molecular Identifiers (UMI),
470 allowing for downstream de-duplication. For each sample either 100ng of genomic DNA (bulk samples) or the entirety
471 of the worm lysate (single worms) was used as the template. PCR samples were then purified using the Zymo DNA
472 Clean and Concentrator-5 Kit (Cat. No.: D4004) following manufacture protocol and eluted with 24µl water. Samples
473 were not quantified prior to the next step as most DNA will not be from the target PCR product. A second PCR

474 (Amplicon two) using the entire 24µl of the extract from the previous step was performed for 24 cycles to added
475 indices. In some cases, a smaller, non-specific product was also formed, so each sample was run on a 2% agarose gel
476 and extracted for the 169bp size product.

477 Two separate PCRs were performed to quantify the diversity of integrated barcode sequences. The first PCR
478 (Amplicon one-integrant) was performed for three cycles to add UMI sequences. For each sample, 100ng of genomic
479 DNA was used as the template. PCR products were then purified as described above and followed the Amplicon two
480 protocol. Each product was quantified on a Synergy H1 plate reader using software Gen 5 3.11. Samples were mixed
481 at an equal molar ratio for a 20nM final concentration for Illumina Sequencing.

482 **Illumina sequencing and data processing for barcode characterization**

483 To quantify the diversity of barcodes in each sample, PCR products were sequenced on either a single NextSeq 500
484 lane or NovaSeq SP, with single read protocols performed by the Genomics and Cell Characterization Facility (GC3F)
485 at the University of Oregon. Compressed fastq files were processed with cutadept 4.1 (Martin, 2011) to remove low
486 quality reads (quality score < 30, max expected error=1, presence of 'N' within the read) and trimmed to 87bp. For
487 the NextSeq lane the specific nextseq trim=30 command was used. The sequences were then demultiplexed using
488 cutadept. For duplicate removal, AmpUMI (Clement et al., 2018) in "processing mode" was used with umi regex
489 "CACIIIIIIIIIIGAC" for individual index files. Deduplicated reads were then trimmed to 15 base pairs with cutadapt
490 for each file. Starcode (Zorita et al., 2015) was then used for mutation correction and counting of each barcode
491 sequence. Each unique sequence was only kept if its final length was 15 base pairs. For the injection mix, each unique
492 barcode was kept regardless of total reads. For all TARDIS arrays and F1 integrations, we used the observed plateau
493 in the number of observed unique barcodes for various count cutoffs to establish a conservative threshold of five or
494 more reads for true barcode sequence (Figure 3—figure supplement 3). Visualizations were created with Python 3.7.13
495 (Guido van Rossum, 1991) and matplotlib 3.5.2 (Hunter, 2007). Sequence logos were created with Logomaker (Tareen
496 and Kinney, 2020). Correlation and p-values generated by scipy input stats.pearsonr (Virtanen et al., 2020). This
497 statistical test was chosen because the relationship from array to integration is approximately linear. All data were
498 processed in Jupyter Notebooks (Kluyver et al., 2016) utilizing Google Colaboratory (colab.research.google.com). All
499 python code is available on Figshare.

500 **Design of landing pads for transcriptional reporters**

501 The utilized fluorophores, mScarlet-I (Bindels et al., 2017) and mNeonGreen (Shaner et al., 2013) were synthesized
502 with the desired modifications as genes incorporated into the pUCIDT-KAN plasmid (IDT). First, a SV40 nuclear
503 localization sequence (NLS) was added after the 13th codon of the *mScarlet-I* gene. This same 66bp sequence was also
504 used in place of the first four codons of the *mNeonGreen* gene. Secondly, a PEST domain (Li et al., 1998) flanked by
505 MluI restriction endonuclease sites and an additional NLS from the *egl-13* gene (Lyssenko et al., 2007) were added to
506 the 3' end of the gene. The *C. elegans* Codon Adapter (https://worm.mpi-cbg.de/codons/cgi-bin/optimize.py)
507 (Redemann et al., 2011) was used to codon optimize both modified fluorophore sequences and to identify locations
508 for three synthetic introns. The first two introns contained 10-base pair periodic A_n/T_n-clusters (PATCs), which have
509 been shown to reduce rates of transgene silencing (Frøkjær-Jensen et al., 2016), while the third was a standard
510 synthetic intron. Finally, the 3' UTR of the *tbb-2* gene, which is permissive for germline expression (Merritt et al.,
511 2008), was added to the end of fluorophore genes. The modified *mScarlet-I* and *mNeonGreen* genes were PCR
512 amplified and assembled into NotI and SnaBI linearized pDSP1, a standard backbone vector derived from pUCIDT-
513 KAN. The resulting *mScarlet-I*-containing plasmid was designated pDSP6 and the *mNeonGreen* containing plasmid
514 was designated pDSP7. In addition, pDSP9, a version of mScarlet-I lacking the PEST destabilization sequence was
515 generated by PCR amplifying the shared *egl-13* NLS and *tbb-2* 3' UTR sequence from pDSP6 and then assembling
516 this fragment into MluI and SnaBI linearized pDSP6.

517 Landing pads were built using a modification of our previous split landing pad strategy (Stevenson et al., 2020). Each
518 landing pad contained the 3' portion of a selectable marker followed by a validated guide sequence and the 3' portion
519 of a fluorophore. The guide sequence (GGACAGTCCTGCCGAGGTGGAGG) has no homology in the *C. elegans*
520 genome and has been previously shown to allow for efficient editing (Stevenson et al., 2020). This sequence was
521 targeted by the plasmid pMS84 which was made from pZCS2, a plasmid made in the same manner as pZCS11 but
522 which is missing a segment of the plasmid backbone, using the Q5 site-directed mutagenesis kit (NEB). *mScarlet-I*

523 was paired with a *HygR* marker (Dickinson et al., 2013) while *mNeonGreen* gene was paired with the *Cbr-unc-119(+)*
524 rescue cassette (Frøkjær-Jensen et al., 2008).

525 **Construction of split HygR /mScarlet-I landing pads**

526 The split *HygR*/*mScarlet-I* landing pad was inserted into the well-characterized *ttTi5605* Mos1 site on Chromosome
527 II (Frokjaer-Jensen *et al.*, 2008). pQL222 (a gift from Dr. QueeLim Ch'ng), a modified version of the pCFJ350
528 (Frøkjær-Jensen et al., 2012) in which the original resistance marker was changed to a kanamycin and zeocin cassette,
529 was digested with BsrGI to provide a linear vector backbone. The *Cbr-unc-119* gene, with a lox2272 sequence added
530 to the 5' end, and a multiple cloning site (MCS) with a lox2272 site added to the 3' end were PCR amplified from
531 pQL222. These two fragments were assembled into the linearized backbone to yield pDSP2.

532 Next, the 3' 949 bases of the *HygR* marker were amplified along with the *unc-54* 3'UTR from pDD282 (Dickinson et
533 al., 2015). The primers used were designed to invert the loxP sequence at the 3' end of *unc-54* 3'UTR from its original
534 orientation in pDD282 and to add the guide sequence to the 5' end of the *HygR* fragment. The 3' 821 bases of the
535 *mScarlet-I* gene along with the tbb-2 3' UTR were amplified from pDSP6. These these two amplicons were assembled
536 into a SbfI/SnaBI digested pDSP2 vector to yield pDSP61. Similarly, the *mScarlet-I* gene was amplified from pDSP9
537 and assembled into pDSP2 along with the *HygR* fragment to give pDSP62, a PEST-less version of the landing pad
538 construct. Both the PEST-containing and PEST-less versions of the split *HygR*/*mScarlet-I* landing pads were
539 integrated into QL74, a 6x outcross of EG4322 (Frøkjær-Jensen et al., 2008), using the standard MosSCI technique
540 (Frøkjær-Jensen et al., 2012) to yield strains GT331 and GT332.

541 **Construction of Split Cbr-unc-119(+)/mNeonGreen landing pad**

542 To construct the *Cbr-unc-119(+)*/*mNeonGreen* landing pad, we wanted to find a genomic safe harbor site permissive
543 to germline expression of transgenes. The *oxTi179* universal MosSCI site on Chromosome II permits germline
544 expression but interrupts *arrd-5*, an endogenous *C. elegans* gene. Therefore, CRISPR-mediated genome editing was
545 used to place the landing pad between *ZK938.12* and *ZK938.3*, two genes adjacent to *arrd-5* whose 3' UTRs face each
546 other. The genomic sequence CATGGTATAAAGTGAATCA<u>AGG</u> was targeted by the plasmid pDSP45 which was
547 made from pDD162 (Dickinson et al., 2013) using the Q5 site-directed mutagenesis kit (NEB).

548 Chromosomal regions II:9830799-9831548 and II:9831573-9832322 were amplified from genomic DNA for use as
549 homology arms. The self-excising cassette (SEC) was PCR amplified from pDD282 such that the loxP sites were
550 replaced by lox2272 sites. A MCS was amplified from pDSP2 while a linear vector backbone fragment was amplified
551 from pDSP1. All five of these PCR fragments were assembled into a circular plasmid, which was immediately used
552 as a template for seven synonymous single-nucleotide substitutions into the terminal 21bp of the *ZK938.12* gene
553 fragment by Q5 site-directed mutagenesis kit (NEB). The resultant plasmid was named pDSP47.

554 The 3' 846 based of the *Cbr-unc-119(+)* rescue cassette plus the 3' UTR were amplified from pDSP2 such that the
555 lox2272 sequence after the 3' UTR was replaced with a loxN site and the guide site
556 GGACAGTCCTGCCGAGGTGGAGG was added upstream of the coding sequence. The 3' 818 bases of
557 *mNeonGreen* plus the tbb-2 3' UTR were amplified from pDSP7. These two amplicons were assembled into StuI/AvrII
558 digested pDSP47 to yield pDSP63.

559 Following the protocol from Dickinson et al. (2015), the landing pad from pDSP63 was integrated into the GT331
560 strain using pDSP45 as the guide plasmid. Upon integration, this yielded strain GT336. Activation of the Cre
561 recombinase within the SEC by heat shock caused both the removal of the SEC from the *mNeonGreen* landing pad
562 and the *Cbr-unc-119(+)* cassette from the *mScarlet-I* landing pad. The combined effect of this double excision event
563 was to yield strain GT337 which has an Unc phenotype and no longer has the hygromycin resistance and Rol
564 phenotypes.

565 **Design and construction of promoter library**

566 Targeting vectors were constructed to provide the 5' portions of each split gene pairing. Both targeting vectors had
567 the same multiple cloning site, allowing promoter amplicons to be assembled into either vector using the same set of
568 primers. In addition, each selectable marker gene is flanked by a lox site that matches the sequence and orientation of

569 the lox site flanking the 3' portion of the marker in the genomic landing pad, allowing for the optional post-integration
570 removal of the selectable marker gene using Cre recombinase.

571 To construct the split *HygR*/*mScarlet-I* targeting vector, the *rps-0* promoter plus the 5' 627 bases of the *HygR* gene
572 were amplified from pDD282 such that a loxP site was added in front of the promoter sequence The MCS was
573 amplified from pDSP2 and the 5' 803 bases of the *mScarlet-I* gene were amplified from pDSP6. All three of these
574 amplicons were assembled into NotI/SnaBI digested pDSP1 to yield pDSP15.

575 To construct the split *Cbr-unc-19(+)*/*mNeonGreen* targeting vector, the promoter and the 5' 515 bases of the *Cbr-*
576 *unc-19(+)* were amplified from pDSP2 such that a loxN site was added prior to the promoter. The MCS was amplified
577 from pDSP2 and the 5' 830 bases of the *mNeonGreen* gene were amplified from pDSP7. All three of these amplicons
578 were assembled into NotI/SnaBI digested pDSP1 to yield pDSP16.

579 The entire intergenic region was used for *aha-1p*, *ahr-1p*, *ceh-20p*, *ceh-40p*, *egl-46p*, *hlh-16p* and *nhr-67p*. For *ceh-*
580 *43p* the 2096bp upstream of the *ceh-43* start codon was used. For *mdl-1p*, *egl-43p* and *ceh-10p,* the promoters describe
581 in (Reece-Hoyes et al., 2013) were used. For *daf-7p* and *lin-11p* the promoters described in (Entchev et al., 2015) and
582 (Marri and Gupta, 2009) respectively were used. Promoters were amplified from N2 genomic DNA using primers
583 designed to add the appropriate homology to the targeting vector and assembled into PCR linearized pDSP15 or
584 pDSP16 for split *HygR*/*mScarlet-I* and split *Cbr-unc-19(+)*/*mNeonGreen* respectively.

585 **Insertion of promoter libraries by TARDIS**

586 For integration of a promoter library into a single landing pad site, a mixture consisting of 15ng/µl guide plasmid
587 (pMS84), 20ng/µl *hsp16.41::Cas9* plasmid (pZCS36), 10ng/µl neomycin resistance plasmid (pZCS38) and
588 0.45fmol/µl of each of the 13 repair template plasmids (Table 1) was microinjected into the gonad arms of young adult
589 GT332 hermaphrodites. Individuals were incubated at 20°C and after three days treated with 1.56mg/ml G-418 to
590 select for array bearing individuals. Once stable array lines were obtained, integration was done using the standard
591 TARDIS protocol, using a density of approximately 200 L1s per plate.

592 For integration of a promoter library into two landing pad site, a mixture consisting of 15ng/µl guide plasmid (pMS84),
593 20ng/µl *hsp16.41::Cas9* plasmid (pZCS36), 0.5ng/µl neomycin resistance plasmid (pZCS38) and 0.45fmol/µl of each
594 of the 14 repair template plasmids (seven targeted to each site) was microinjected into the gonad arms of young adult
595 GT337 hermaphrodites. Individuals were incubated at 20°C and after three days treated with 1.56mg/ml G-418 to
596 select for array bearing individuals. Once a stable array line was obtained, plates of mixed stage worms were
597 transferred to plates without drug, heat shocked at 35.5°C for 1.5hrs and returned to 20°C. Three days after heat shock,
598 hygromycin was added at a final concentration of 250µg/ml.

599 For both scenarios, candidate worms (those which had both hygromycin resistance and wild-type movement) were
600 singled and screened by PCR. The identity of the integrated promoters was determined by Sanger sequencing of the
601 PCR product. Primers used for genotyping are given in Supplementary file 6.

602 **Microscopy**

603 Individual late L4/young adults were mounted on 2% agarose pads and immobilized with 0.5M levamisole. Imaging
604 was performed on a DeltaVision Ultra microscope (Cytiva, Massachusetts, USA) using the 20x objective and Acquire
605 Ultra software version 1.2.1. Fluorescent images were acquired using the orange (542/32nm) and green (525/48nm)
606 filter sets for mScarlet-I and mNeonGreen respectively. Light images were captured at 5% transmission and a 0.01
607 second exposure. Fluorescent images were captured at 5% transmission and a 2sec (*aha-1p*), 1sec (*ceh-40p*, *ceh-43p*,
608 *nhr-67p*, *ceh-10p::mNeonGreen*), 0.5sec (*ceh-10p::mScarlet-I*, *ceh-20p*, *daf-7p*) or 0.2sec (*lin-11p*, *mdl-1p*) exposure.
609 Images were processed in Fiji (ImageJ) version 2.9.0/1.53t.

610 **Accessibility of reagents, data, code, and protocols**

611 The authors affirm that all data necessary for confirming the conclusions of the article are present within the article,
612 figures, and tables. Plasmids pDSP15 (Addgene ID 193853), pDSP16 (Addgene ID193854), pMS84 (Addgene ID
613 193852), pZCS36 (Addgene ID 193048), pZCS38 (Addgene ID193049), and pZCS41 (Addgene ID 193050), are
614 available through Addgene and can be freely viewed and edited in ApE (Davis and Jorgensen, 2022) and other

compatible programs. Strains PX740, GT332 and GT337 are available from the *Caenorhabditis* Genetics Center (cgc.umn.edu). Strains and plasmids not available at a public repository are available upon request. Illumina sequencing data are available at BioProject ID: PRJNA893002. All other data, code, plasmid and landing sequences and original microscopy images are available on Figshare (Stevenson et al., 2022). We plan to continue to develop TARDIS technology and provided descriptions of updated libraries and advancements at: https://github.com/phillips-lab/TARDIS.

Acknowledgments

We thank the Phillips lab members for helpful suggestions and technical assistance. In particular, we thank Erin Jahahn for her assistance in early TARDIS multiple integrations, Zach Muñoz for his assistance injecting several TARDIS libraries, Kristin Robinson for creating PXKR1, and Ellie Laufer for her assistance in genotyping individual promoter integrations. We also thank Sihoon Moon, Hyun Jee Lee, and Eric Andersen for technical assistance in plasmid construction.

Funding

This work was funded by National Institutes of Health grants R35GM131838 awarded to PCP and R01AG056436 awarded to PCP and HL.

Conflict of Interest

The presented technology underlies U.S. patent application 17/236,556 and associated U.S. Provisional Application No. 63/013,365 (Inventors ZCS, SAB, PCP, assignee University of Oregon). The patent applicant/assignee and grant funding institutions had no involvement in the described work, including but not limited to experimental design, data analysis, interpretation, or manuscript preparation.

References

Aljohani MD, el Mouridi S, Priyadarshini M, Vargas-Velazquez AM, Frøkjær-Jensen C. 2020. Engineering rules that minimize germline silencing of transgenes in simple extrachromosomal arrays in *C. elegans*. *Nat Commun* **11**:6300. doi:10.1038/s41467-020-19898-0

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. doi:10.1534/genetics.114.169730

Bertrand V, Bisso P, Poole RJ, Hobert O. 2011. Notch-dependent induction of left/right asymmetry in *C. elegans* interneurons and motoneurons. *Current Biology* **21**:1225–1231. doi:10.1016/j.cub.2011.06.016

Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, Gadella TWJ. 2017. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **14**:53–56. doi:10.1038/nmeth.4074

Blundell JR, Levy SF. 2014. Beyond genome sequencing: Lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics* **104**:1–14. doi:10.1016/j.ygeno.2014.09.005

Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, Stanley G, Chen S, Garnett M, Li W, Moffat J, Qi LS, Shapiro RS, Shendure J, Weissman JS, Zhuang X. 2022. High-content CRISPR screening. *Nature Reviews Methods Primers* **2**:8. doi:10.1038/s43586-021-00093-4

Carlton PM, Davis RE, Ahmed S. 2022. Nematode chromosomes. *Genetics* **221**. doi:10.1093/genetics/iyac014

Clement K, Farouni R, Bauer DE, Pinello L. 2018. AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**:i202–i210. doi:10.1093/bioinformatics/bty264

Cunningham BC, Wells JA. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science (1979)* **244**:1081–1085. doi:10.1126/science.2471267

Davis MW, Jorgensen EM. 2022. ApE, A Plasmid Editor: A freely available DNA manipulation and visualization program. *Frontiers in Bioinformatics* **2**. doi:10.3389/fbinf.2022.818619

de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. 2020. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat Biotechnol* **38**:56–65. doi:10.1038/s41587-019-0315-8

Dickinson DJ, Goldstein B. 2016. CRISPR-Based methods for *Caenorhabditis elegans* genome engineering. *Genetics* **202**:885–901. doi:10.1534/genetics.115.182162

Dickinson DJ, Pani AM, Heppert JK, Higgins CD, Goldstein B. 2015. Streamlined genome engineering with a self-excising drug selection cassette. *Genetics* **200**:1035–1049. doi:10.1534/genetics.115.178335

Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. 2013. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat Methods* **10**:1028–1034. doi:10.1038/nmeth.2641

Emanuel G, Moffitt JR, Zhuang X. 2017. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat Methods* **14**:1159–1162. doi:10.1038/nmeth.4495

Entchev E v, Patel DS, Zhan M, Steele AJ, Lu H, Ch'ng Q. 2015. A gene-expression-based neural code for food abundance that modulates lifespan. *eLife* **4**. doi:10.7554/eLife.06259

Erwood S, Bily TMI, Lequyer J, Yan J, Gulati N, Brewer RA, Zhou L, Pelletier L, Ivakine EA, Cohn RD. 2022. Saturation variant interpretation using CRISPR prime editing. *Nat Biotechnol* **40**:885–895. doi:10.1038/s41587-021-01201-1

Fernandes JS, Sternberg PW. 2007. The tailless ortholog nhr-67 regulates patterning of gene expression and morphogenesis in the *C. elegans* vulva. *PLoS Genet* **3**:e69. doi:10.1371/journal.pgen.0030069

Froehlich JJ, Uyar B, Herzog M, Theil K, Glažar P, Akalin A, Rajewsky N. 2021. Parallel genetics of regulatory sequences using scalable genome editing *in vivo*. *Cell Rep* **35**:108988. doi:10.1016/j.celrep.2021.108988

Frøkjær-Jensen C, Davis MW, Ailion M, Jorgensen EM. 2012. Improved Mos1-mediated transgenesis in *C. elegans*. *Nat Methods* **9**:117–118. doi:10.1038/nmeth.1865

Frøkjær-Jensen C, Jain N, Hansen L, Davis MW, Li Y, Zhao D, Rebora K, Millet JRM, Liu X, Kim SK, Dupuy D, Jorgensen EM, Fire AZ. 2016. An abundant class of non-coding DNA can prevent stochastic gene silencing in the *C. elegans* germline. *Cell* **166**:343–357. doi:10.1016/j.cell.2016.05.072

Frøkjær-Jensen C, Wayne Davis M, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, Grunnet M, Jorgensen EM. 2008. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* **40**:1375–1383. doi:10.1038/ng.248

Gilleland CL, Falls AT, Noraky J, Heiman MG, Yanik MF. 2015. Computer-assisted transgenesis of *Caenorhabditis elegans* for deep phenotyping. *Genetics* **201**:39–46. doi:10.1534/genetics.115.179648

Guido van Rossum. 1991. Interactively testing remote servers using the Python programming language. *CWI quarterly* 283–303.

Gupta BP, Wang M, Sternberg PW. 2003. The *C. elegans* LIM homeobox gene *lin-11* specifies multiple cell fates during vulval development. *Development* **130**:2589–2601. doi:10.1242/dev.00500

Huang X, Powell-Coffman JA, Jin Y. 2004. The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in *C. elegans*. *Development* **131**:819–828. doi:10.1242/dev.00959

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**:90–95. doi:10.1109/MCSE.2007.55

Hwang BJ, Meruelo AD, Sternberg PW. 2007. *C. elegans* EVI1 proto-oncogene, EGL-43, is necessary for Notch-mediated cell fate specification and regulates cell invasion. *Development* **134**:669–679. doi:10.1242/dev.02769

Ismagul A, Yang N, Maltseva E, Iskakova G, Mazonka I, Skiba Y, Bi H, Eliby S, Jatayev S, Shavrukov Y, Borisjuk N, Langridge P. 2018. A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. *BMC Plant Biol* **18**:135. doi:10.1186/s12870-018-1326-1

Jahn LJ, Porse A, Munck C, Simon D, Volkova S, Sommer MOA. 2018. Chromosomal barcoding as a tool for multiplexed phenotypic characterization of laboratory evolved lineages. *Sci Rep* **8**:6961. doi:10.1038/s41598-018-25201-5

Jiang H, Guo R, Powell-Coffman JA. 2001. The *Caenorhabditis elegans hif-1* gene encodes a bHLH-PAS protein that is required for adaptation to hypoxia. *Proceedings of the National Academy of Sciences* **98**:7916–7921. doi:10.1073/pnas.141234698

Joung JK, Ramm EI, Pabo CO. 2000. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **97**:7382–7387. doi:10.1073/pnas.110149297

Kasimatis, K. R., Sánchez-Ramírez, S., & Stevenson, Z. C. (2021). Sexual dimorphism through the lens of genome manipulation, forward genetics, and spatiotemporal sequencing. *Genome Biology and Evolution*, *13*(2). doi: 10.1093/gbe/evaa243

Kaymak E, Farley BM, Hay SA, Li C, Ho S, Hartman DJ, Ryder SP. 2016. Efficient generation of transgenic reporter strains and analysis of expression patterns in *Caenorhabditis elegans* using library MosSCI. *Developmental Dynamics* **245**:925–936. doi:10.1002/dvdy.24426

Kebschull JM, Zador AM. 2018. Cellular barcoding: lineage tracing, screening and beyond. *Nat Methods* **15**:871–879. doi:10.1038/s41592-018-0185-x

Kim Y, Lee S, Cho S, Park J, Chae D, Park T, Minna JD, Kim HH. 2022. High-throughput functional evaluation of human cancer-associated mutations using base editors. *Nat Biotechnol* **40**:874–884. doi:10.1038/s41587-022-01276-4

Klabonski L, Zha J, Senthilkumar L, Gidalevitz T. 2016. A bystander mechanism explains the specific phenotype of a broadly expressed misfolded protein. *PLoS Genet* **12**:e1006450. doi:10.1371/journal.pgen.1006450

Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C. 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows In: Loizides F, Schmidt B, editors. positioning and power in academic publishing: players, agents and agendas. pp. 87–90.

Leung CK, Deonarine A, Strange K, Choe KP. 2011. High-throughput screening and biosensing with fluorescent *C. elegans* strains. *Journal of Visualized Experiments*. doi:10.3791/2745

Levinson G, Gutman A. G. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. doi:10.1093/oxfordjournals.molbev.a040442

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**:181–186. doi:10.1038/nature14279

Levy SLX. 2016. Genomic combinatorial screening platform. WO2017075529A1.

Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang C-C, Kain SR. 1998. Generation of destabilized green fluorescent protein as a transcription reporter. *Journal of Biological Chemistry* **273**:34970–34975. doi:10.1074/jbc.273.52.34970

Lin Z, Xie Y, Nong W, Ren X, Li R, Zhao Z, Hui JHL, Yuen KWY. 2021. Formation of artificial chromosomes in *Caenorhabditis elegans* and analyses of their segregation in mitosis, DNA sequence composition and holocentromere organization. *Nucleic Acids Res* **49**:9174–9193. doi:10.1093/nar/gkab690

Liu B, Jing Z, Zhang X, Chen Y, Mao S, Kaundal R, Zou Y, Wei G, Zang Y, Wang X, Lin W, Di M, Sun Y, Chen Q, Li Y, Xia J, Sun J, Lin C-P, Huang X, Chi T. 2022. Large-scale multiplexed mosaic CRISPR perturbation in the whole organism. *Cell* **185**:3008-3024.e16. doi:10.1016/j.cell.2022.06.039

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW. 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**:56–62. doi:10.1038/nature06293

Lyssenko NN, Hanna-Rose W, Schlegel RA. 2007. Cognate putative nuclear localization signal effects strong nuclear localization of a GFP reporter and facilitates gene expression studies in *Caenorhabditis elegans*. *Biotechniques* **43**:596–600. doi:10.2144/000112615

Malaiwong, N., Porta-de-la-Riva, M., & Krieg, M. (2023). FLInt: Single shot safe harbor transgene integration via fluorescent landmark interference. *G3 Genes|Genomes|Genetics*. doi:10.1093/g3journal/jkad041

Marri S, Gupta BP. 2009. Dissection of lin-11 enhancer regions in *Caenorhabditis elegans* and other nematodes. *Dev Biol* **325**:402–411. doi:10.1016/j.ydbio.2008.09.026

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**:10. doi:10.14806/ej.17.1.200

Matthews BW. 1996. Structural and genetic analysis of the folding and function of T4 lysozyme. *The FASEB Journal* **10**:35–41. doi:10.1096/fasebj.10.1.8566545

McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science (1979)* **353**. doi:10.1126/science.aaf7907

Mello CC, Kramer JM, Stinchcomb D, Ambros V. 1991. Efficient gene transfer in *C.elegans:* extrachromosomal maintenance and integration of transforming sequences. *EMBO J* **10**:3959–3970. doi:10.1002/j.1460-2075.1991.tb04966.x

Merritt C, Rasoloson D, Ko D, Seydoux G. 2008. 3′ UTRs Are the primary regulators of gene expression in the *C. elegans* germline. *Current Biology* **18**:1476–1482. doi:10.1016/j.cub.2008.08.013

Mouridi E, el Mouridi S, Alkhaldi F, Frøkjaer-Jensen C. 2022. Modular safe-harbor transgene insertion (MosTI) for targeted single-copy and extrachromosomal array integration in *C. elegans*. doi:10.1101/2022.04.19.488726

Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, Levy SF, Desai MM. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**:494–499. doi:10.1038/s41586-019-1749-3

Nonet ML. 2020. Efficient transgenesis in *Caenorhabditis elegans* using flp recombinase-mediated cassette exchange. *Genetics* **215**:903–921. doi:10.1534/genetics.120.303388

Nonet, M. (2021). Additional landing sites for recombination-mediated cassette exchange in *C. elegans*. *MicroPublication Biology*, *2021*. doi:10.17912/micropub.biology.000503

774 Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**:379–394.
775 doi:10.1038/nrg3927
776 Pontes-Quero S, Heredia L, Casquero-García V, Fernández-Chacón M, Luo W, Hermoso A, Bansal M, Garcia-
777 Gonzalez I, Sanchez-Muñoz MS, Perea JR, Galiana-Simal A, Rodriguez-Arabaolaza I, del Olmo-Cabrera S,
778 Rocha SF, Criado-Rodriguez LM, Giovinazzo G, Benedito R. 2017. Dual ifgMosaic: A versatile method for
779 multispectral and combinatorial mosaic gene-function analysis. *Cell* **170**:800-814.e18.
780 doi:10.1016/j.cell.2017.07.031
781 Priyadarshini M, Ni JZ, Vargas-Velazquez AM, Gu SG, Frøkjær-Jensen C. 2022. Reprogramming the piRNA
782 pathway for multiplexed and transgenerational gene silencing in *C. elegans*. *Nat Methods* **19**:187–194.
783 doi:10.1038/s41592-021-01369-z
784 Radman I, Greiss S, Chin JW. 2013. Efficient and rapid *C. elegans* transgenesis by bombardment and hygromycin B
785 selection. *PLoS One* **8**:e76019. doi:10.1371/journal.pone.0076019
786 Redemann S, Schloissnig S, Ernst S, Pozniakowsky A, Ayloo S, Hyman AA, Bringmann H. 2011. Codon
787 adaptation–based control of protein expression in *C. elegans*. *Nat Methods* **8**:250–252.
788 doi:10.1038/nmeth.1565
789 Reece-Hoyes JS, Pons C, Diallo A, Mori A, Shrestha S, Kadreppa S, Nelson J, DiPrima S, Dricot A, Lajoie BR,
790 Ribeiro PSM, Weirauch MT, Hill DE, Hughes TR, Myers CL, Walhout AJM. 2013. Extensive rewiring and
791 complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network. *Mol Cell* **51**:116–
792 127. doi:10.1016/j.molcel.2013.05.018
793 Reece-Hoyes JS, Shingles J, Dupuy D, Grove CA, Walhout AJ, Vidal M, Hope IA. 2007. Insight into transcription
794 factor gene duplication from *Caenorhabditis elegans* promoterome-driven expression patterns. *BMC Genomics*
795 **8**:27. doi:10.1186/1471-2164-8-27
796 Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang C-H, Petrov D, Winslow MM. 2017. A quantitative and
797 multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. *Nat Methods* **14**:737–742.
798 doi:10.1038/nmeth.4297
799 Sánchez-Rivera FJ, Diaz BJ, Kastenhuber ER, Schmidt H, Katti A, Kennedy M, Tem V, Ho Y-J, Leibold J,
800 Paffenholz S v., Barriga FM, Chu K, Goswami S, Wuest AN, Simon JM, Tsanov KM, Chakravarty D, Zhang
801 H, Leslie CS, Lowe SW, Dow LE. 2022. Base editing sensor libraries for high-throughput engineering and
802 functional analysis of cancer-associated single nucleotide variants. *Nat Biotechnol* **40**:862–873.
803 doi:10.1038/s41587-021-01172-3
804 Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, Hasse S, Rupprecht M, Vinis E, Tinney M,
805 Preston E, Zinke A, Enst S, Teichgraber T, Janette J, Reis K, Janosch S, Schloissnig S, Ejsmont RK, Slightam
806 C, Xu X, Kim SK, Reinke V, Stewart AF, Snyder M, Waterston RH, Hyman AA. 2012. A genome-scale
807 resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell* **150**:855–866.
808 doi:10.1016/j.cell.2012.08.001
809 Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M,
810 Davidson MW, Wang J. 2013. A bright monomeric green fluorescent protein derived from *Branchiostoma*
811 *lanceolatum*. *Nat Methods* **10**:407–409. doi:10.1038/nmeth.2413
812 Silva-García CG, Lanjuin A, Heintz C, Dutta S, Clark NM, Mair WB. 2019. Single-copy knock-in loci for defined
813 gene expression in *Caenorhabditis elegans*. *G3 Genes|Genomes|Genetics* **9**:2195–2198.
814 doi:10.1534/g3.119.400314
815 Smith AM, Durbic T, Oh J, Urbanus M, Proctor M, Heisler LE, Giaever G, Nislow C. 2011. Competitive genomic
816 screens of barcoded yeast libraries. *Journal of Visualized Experiments*. doi:10.3791/2864
817 Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. 2018. Simultaneous lineage
818 tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat Biotechnol* **36**:469–473.
819 doi:10.1038/nbt.4124
820 Stevenson ZC, Banse SA, Phillips PC. 2021. Genetic data compression and methods of use. US20210332387A1.
821 Stevenson ZC, Moerdyk-Schauwecker MJ, Jamison B, Phillips PC. 2020. Rapid self-selecting and clone-free
822 integration of transgenes into engineered CRISPR safe harbor locations in *Caenorhabditis elegans*. *G3*
823 *Genes|Genomes|Genetics* **10**:3775–3782. doi:10.1534/g3.120.401400
824 Stevenson ZC, Moerdyk-Schauwecker M, Banse S, Patel D, Lu H, Phillips P. 2022. High-throughput library
825 transgenesis in *Caenorhabditis elegans* via transgenic arrays resulting in diversity of integrated sequences
826 (TARDIS): Online supplemental collection. *Figshare*. doi:10.6084/m9.figshare.c.6264162
827 Stinchcomb DT, Shaw JE, Carr SH, Hirsh D. 1985. Extrachromosomal DNA transformation of *Caenorhabditis*
828 *elegans*. *Mol Cell Biol* **5**:3484–3496. doi:10.1128/MCB.5.12.3484

Tareen A, Kinney JB. 2020. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**:2272–2274. doi:10.1093/bioinformatics/btz921

Teterina AA, Coleman-Hulbert AL, Banse SA, Willis JH, Perez VI, Lithgow GJ, Driscoll M, Phillips PC. 2022. Genetic diversity estimates for the *Caenorhabditis* intervention testing program screening panel. *MicroPubl Biol* **2022**. doi:10.17912/micropub.biology.000518

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, Fisher DS, Sherlock G, Petrov DA. 2016. Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations inyeast. *Cell* **166**:1585-1596.e22. doi:10.1016/j.cell.2016.08.002

Venken KJT, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, Wan KH, Koriabine M, de Jong PJ, White KP, Bellen HJ, Hoskins RA. 2009. Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nat Methods* **6**:431–434. doi:10.1038/nmeth.1331

Venken KJT, He Y, Hoskins RA, Bellen HJ. 2006. P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science (1979)* **314**:1747–1751. doi:10.1126/science.1134426

Vicencio, J., Martínez-Fernández, C., Serrat, X., & Cerón, J. (2019). Efficient generation of endogenous fluorescent reporters by nested CRISPR in *Caenorhabditis elegans*. *Genetics*, *211*(4), 1143–1154. doi:10.1534/genetics.119.301965

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli A pietro, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D v., Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold G-L, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**:261–272. doi:10.1038/s41592-019-0686-2

Ward JD. 2015. Rapid and precise engineering of the *Caenorhabditis elegans* genome with lethal mutation co-conversion and inactivation of NHEJ repair. *Genetics* **199**:363–377. doi:10.1534/genetics.114.172361

Weismann A. 1893. The Germ-plasm: a theory of heredity. Translated by W. Newton Parker and Harriet Rönnfeldt. New York: Scribner. doi:10.5962/bhl.title.25196

Weissman TA, Pan YA. 2015. Brainbow: New resources and emerging biological applications for multicolor genetic labeling and analysis. *Genetics* **199**:293–306. doi:10.1534/genetics.114.172510

Wells JA. 1991. [18] Systematic mutational analyses of protein-protein interfaces. pp. 390–411. doi:10.1016/0076-6879(91)02020-A

Winslow MPDWIMCRZ. 2022. Compositions and methods for multiplexed quantitative analysis of cell lineages. 17/281919.

Woglar A, Yamaya K, Roelens B, Boettiger A, Köhler S, Villeneuve AM. 2020. Quantitative cytogenetics reveals molecular stoichiometry and longitudinal organization of meiotic chromosome axes and loops. *PLoS Biol* **18**:e3000817. doi:10.1371/journal.pbio.3000817

Wu J, Duggan A, Chalfie M. 2001. Inhibition of touch cell fate by *egl-44* and *egl-46* in *C. elegans*. *Genes Dev* **15**:789–802. doi:10.1101/gad.857401

Wu M-R, Nissim L, Stupp D, Pery E, Binder-Nissim A, Weisinger K, Enghuus C, Palacios SR, Humphrey M, Zhang Z, Maria Novoa E, Kellis M, Weiss R, Rabkin SD, Tabach Y, Lu TK. 2019. A high-throughput screening and computation platform for identifying synthetic promoters with enhanced cell-state specificity (SPECS). *Nat Commun* **10**:2880. doi:10.1038/s41467-019-10912-8

Xu H, Guo Y, Qiu L, Ran Y. 2022. Progress in soybean genetic transformation over the last decade. *Front Plant Sci* **13**. doi:10.3389/fpls.2022.900318

Zhang D, Tu S, Stubna M, Wu W-S, Huang W-C, Weng Z, Lee H-C. 2018. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science* **359**:587–592. doi:10.1126/science.aao2840

Zorita E, Cuscó P, Filion GJ. 2015. Starcode: sequence clustering based on all-pairs search. *Bioinformatics* **31**:1913–1919. doi:10.1093/bioinformatics/btv053

917 **Tables**

918 **Table 1**: Characteristics of injected promoters and presence in tested array line (PX819) and integrated lines derived from that array.

Promoter	Promoter size (bp)	Expected expression location	Array	Integrated
aha-1	330	neurons, hypodermis, intestine, pharynx (Jiang et al., 2001)	Y	Y
hlh-16	514	head neurons (Bertrand et al., 2011)	Y	N
ceh-40	965	dopaminergic neurons (Sarov et al., 2012)	Y	Y
ceh-10	1172	neurons, seam cells (Reece-Hoyes et al., 2007)	Y	Y
ahr-1	1387	ALM and RME neurons (Huang et al., 2004)	Y	N
mdl-1	2000	neurons, body wall, pharynx (Reece-Hoyes et al., 2007)	Y	Y
egl-43	2001	neurons, gonad (Hwang et al., 2007)	Y	N
ceh-20	2015	neurons, seam cells, vulva (Reece-Hoyes et al., 2007)	Y	Y
ceh-43	2096	neurons, anterior hypodermis (Reece-Hoyes et al., 2007)	Y	Y
daf-7	2524	head neurons, coelemocytes, pharynx (Klabonski et al., 2016)	Y	Y
lin-11	2857	neurons, uterus, vulva, head muscle (Gupta et al., 2003)	Y	Y
egl-46	4477	Neurons (Wu et al., 2001)	N	N
nhr-67	5545	neurons, excretory cell, rectal valve cell, vulva (Fernandes and Sternberg, 2007)	Y	Y

919

920

921

922

Supplemental Figures

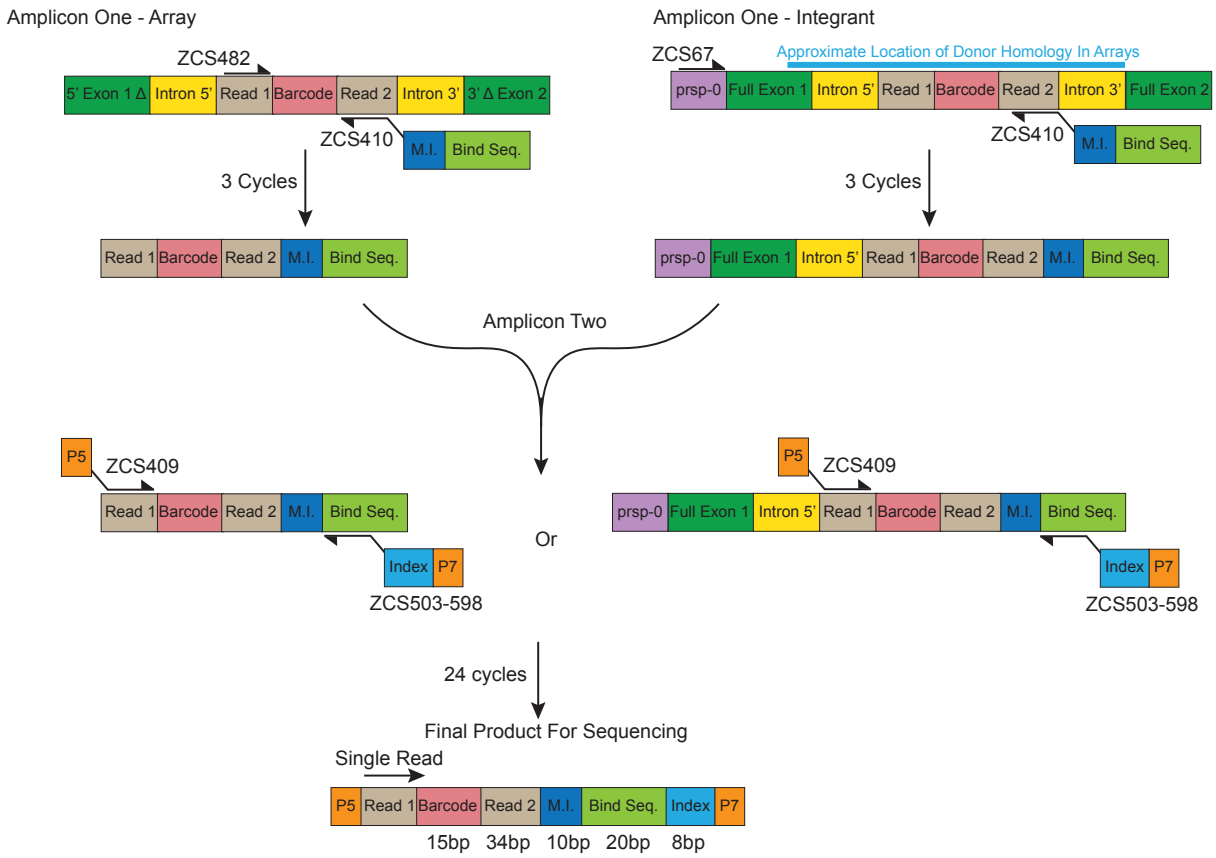


Figure 2 —figure supplement 1: Schematic layout for the two separate PCR processes for identification of barcode counts in arrays (Amplicon One-Array) and integrants (Amplicon One-Integrant).

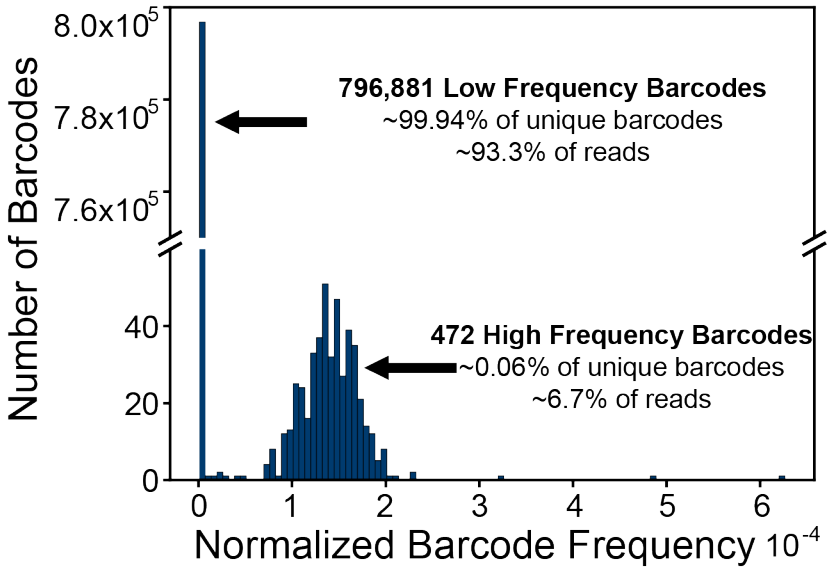


Figure 3—figure supplement 1: A) Barcode frequencies for Injection mix used for arrays 1, 2, and 3. There are approximately 1 million reads represented. In total, 797353 unique sequences were identified. A few of these unique sequences were represented at a higher frequency with a count cutoff greater than 50.

926

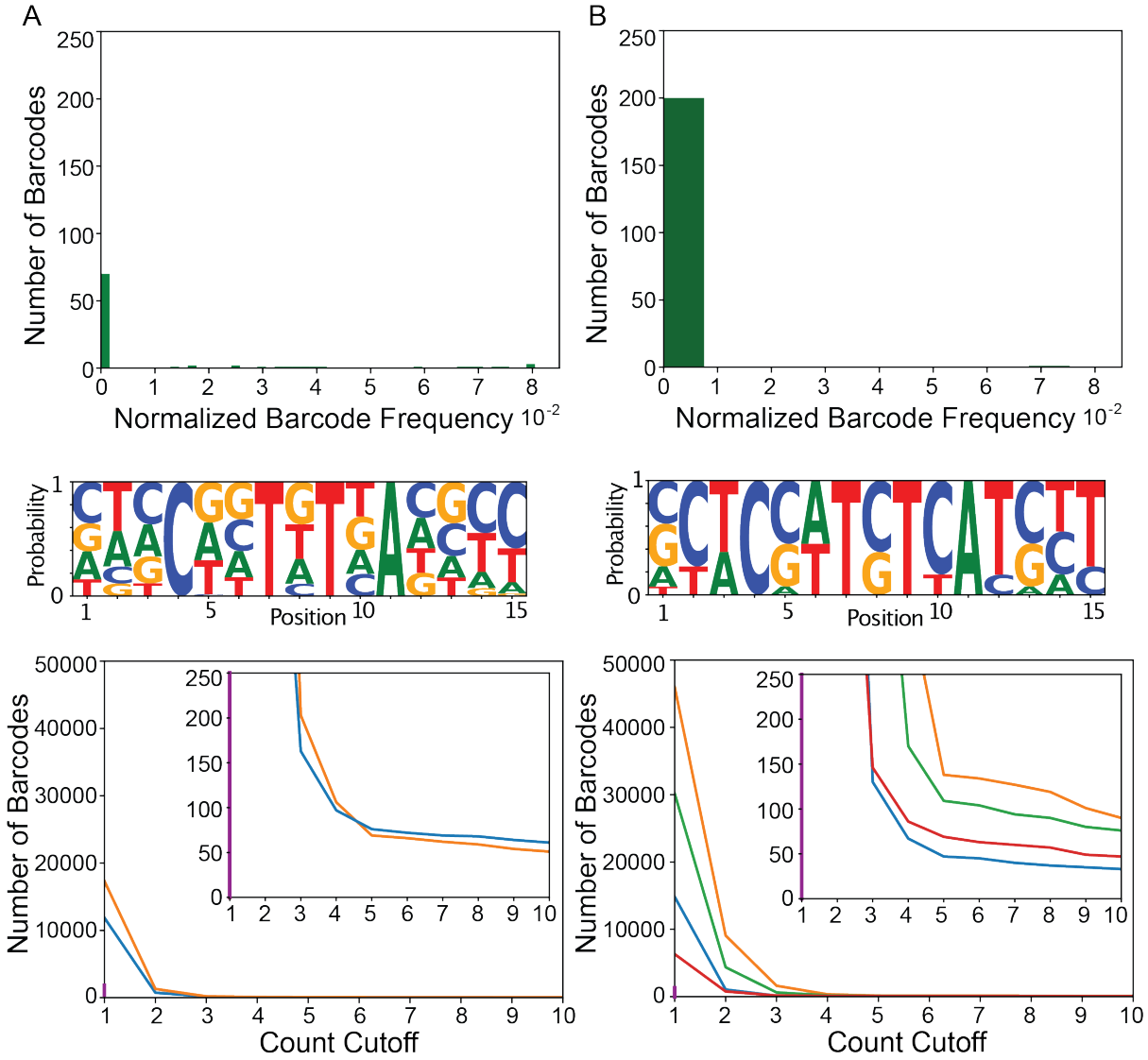


Figure 3—figure supplement 2: TARDIS Array 3. Two individual arrays were isolated from the same plate. Both show considerably less diversity than TARDIS arrays 1 and 2. Distribution of unique barcode frequencies, sequence logo base pair probability, and count cut off for A) TARDIS Array 3 profile 1 and B) TARDIS Array 3 profile 2.

927

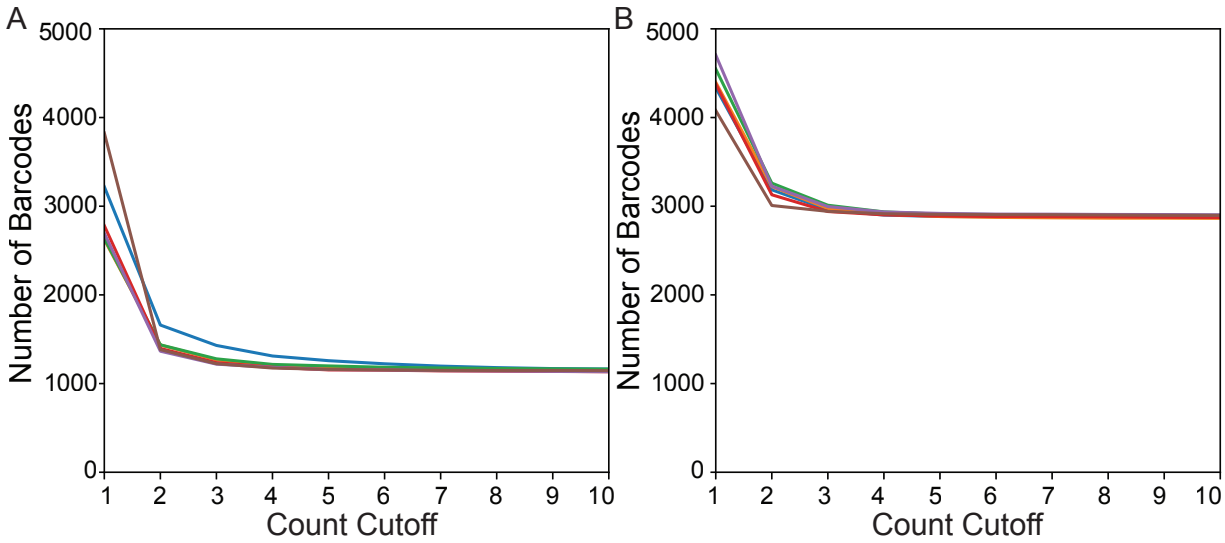


Figure 3—figure supplement 3: Determination of proper count cutoff for A) TARDIS Array 1 and B) TARDIS Array 2.

928

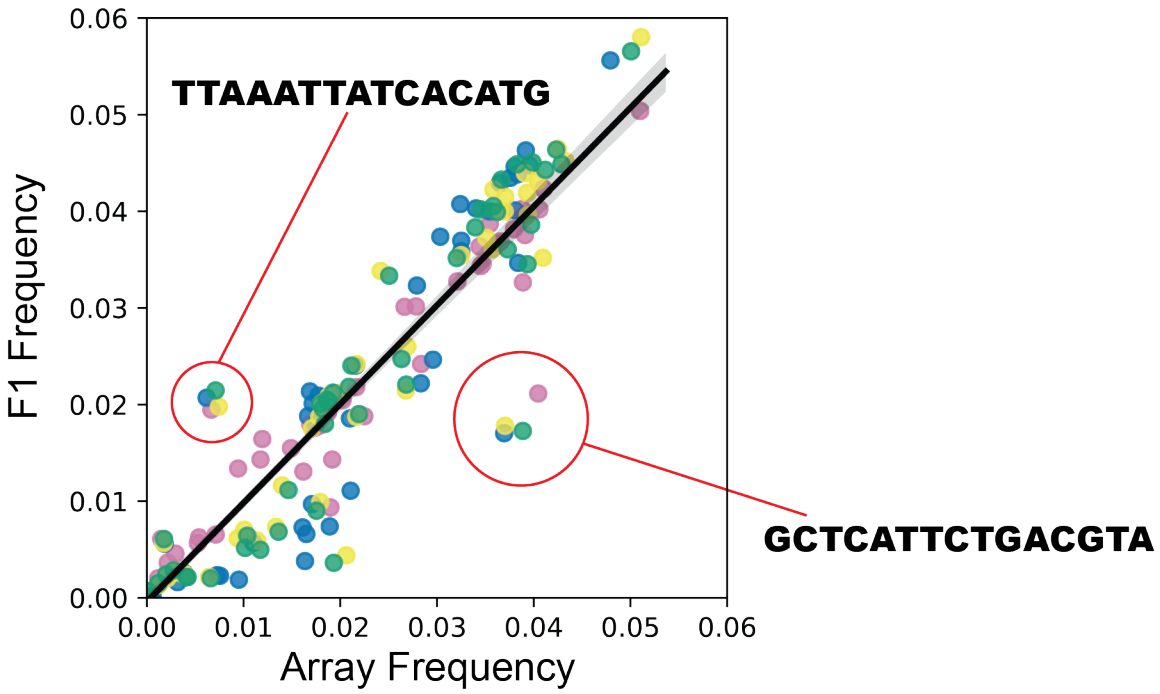
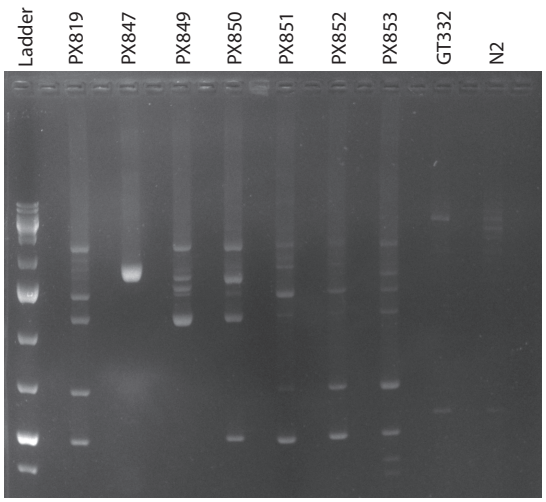


Figure 4—figure supplement 1: F1 integration events followed a consistent pattern, with replicated outlier barcode sequences. Generally, the same barcodes integrated at approximately the same frequency across the four replicates.

929

A.



B.

	aha-1p	*hlh-16p*	*ceh-40p*	*ceh-10p*	*ahr-1p*	*mdl-1p*	*egl-43p*	*ceh-20p*	*ceh-43p*	*daf-7p*	*lin-11p*	*egl-46p*	*nhr-67p*	
PX819	x	x	x	x	x	x	x	x	x	x	x		x	12/13
PX847							x						x	2/13
PX849			x	x	x	x	x		x	x	x	x	x	10/13
PX850	x		x	x	x	x	x	x	x	x	x	x	x	12/13
PX851	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
PX852	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
PX853	x	x	x	x	x	x	x	x	x	x	x	x	x	13/13
	5/7	4/7	6/7	6/7	6/7	6/7	7/7	5/7	6/7	6/7	6/7	5/7	7/7	

C.

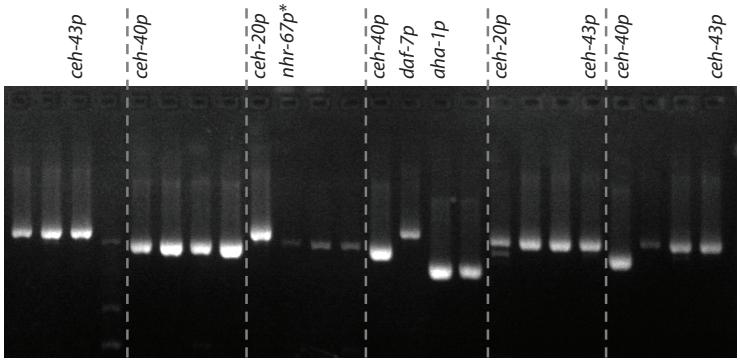


Figure 5—figure supplement 1: (A) Promoters from seven array bearing lines were amplified using universal primers and show distinct profiles. Note that not all promoters found in the arrays were detected in this screen. (B) Promoters contained in each array bearing line as determined by promoter specific PCR (C) Line PX819 was heat shocked to trigger integration and four hygromycin resistant progeny were singled from each of the 59/60 plates with hygromycin

resistant individuals. Singled worms were screened by PCR and select promoters were chosen for sequencing based on their size profile. PCR and sequencing result are shown for representative plates. *=Due to their size of the egl-46 and nhr-67 promoters do not reliably amplify with the universal primers. Therefore, samples with no or weak amplification were rescreened with primers specific to these two promoters (not shown).

Additional files

Supplementary file 1
A table of the plasmids used.

Supplementary file 2
A table of the primers used for cloning.

Supplementary file 3
A table of the primers used for strain confirmation by Sanger sequencing.

Supplementary file 4
A table of the primers and ultramer used to create the barcode donor homology.

Supplementary file 5
A table of the primers used for Illumina sequencing.

Supplementary file 6
A table of the primers used for promoter identification.

Supplementary file 7
A document with the complete genotypes for the barcode TARDIS arrays presented in the study.

Figure 5 – figure supplement 1 -source data 1 This file contains original gels used in creating figure 5 – figure supplement 1 parts A and C