Published in final edited form as:
*Cell*. 2012 August 17; 150(4): 855–866. doi:10.1016/j.cell.2012.08.001.

# A genome scale resource for *in vivo* tag-based protein function exploration in *C. elegans*

**Mihail Sarov**[1,#,*], **John Murray**[2,7], **Kristin Schanze**[1,#], **Andrei Pozniakovski**[1], **Wei Niu**[3], **Karolin Angermann**[1,#], **Susanne Hasse**[1,#], **Michaela Rupprecht**[1,#], **Elisabeth Vinis**[1,#], **Matthew Tinney**[1,#], **Elicia Preston**[2,7], **Andrea Zinke**[1], **Susanne Enst**[1], **Tina Teichgraber**[1], **Judith Janette**[3], **Kadri Reis**[1,#], **Stephan Janosch**[1,#], **Siegfried Schloissnig**[1,#], **Radoslaw K. Ejsmont**[1], **Cindie Slightam**[4], **Xiao Xu**[4], **Stuart K. Kim**[4], **Valerie Reinke**[3], **A. Francis Stewart**[5], **Michael Snyder**[6], **Robert Waterston**[2], and **Anthony A. Hyman**[1]

[1]Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany

[#]TransgeneOmics Unit, Dresden 01307

[2]Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195

[3]Departments of Genetics, Yale University, New Haven CT 06520

[4]Departments of Developmental Biology, Stanford University Medical Center, Stanford, CA 94305

[5]University of Technology Dresden, Department of Genomics, Dresden 01307

[6]Genetics, Stanford University Medical Center, Stanford, CA 94305

## Abstract

Understanding the in vivo dynamics of protein localization and their physical interactions is important for many problems in Biology. To enable systematic protein function interrogation in a multicelluar context, we built a genome-scale transgenic platform for *in vivo* expression of fluorescent and affinity tagged proteins in *Caenorhabditis elegans* under endogenous *cis* regulatory control. The platform combines computer-assisted transgene design, massively parallel DNA engineering and next generation sequencing to generate a resource of 14637 genomic DNA transgenes, which covers 73% of the proteome. The multipurpose tag used allows any protein of interest to be localized *in vivo* or affinity purified using standard tag-based assays. We illustrate the utility of the resource by systematic chromatin immunopurification and automated 4D imaging, which produced detailed DNA binding and cell/tissue distribution maps for key transcription factor proteins

## Introduction

A major challenge of the post-genomic era is to understand how the instructions encoded in the genome are read into the extraordinary variety of molecular structures and biochemical



[*]Corresponding author: Mihail Sarov, Phone: + 49 351 210-2617, sarov@mpi-cbg.de; MPI-CBG, Pfotenhauerstraße 108, 01307 Dresden, Germany.
[7]Current address: Department of Genetics, University of Pennsylvania School of Medicine

reactions that make up cells and organisms (Saghatelian and Cravatt, 2005; Dunn and Kingston, 2007). Due to the enormous complexity of biological systems and the multiple, interconnected levels of gene expression regulation (transcription, splicing, translation and turnover control through 5' and 3' UTRs and posttranslational modifications) the identity, molecular function, and localization of gene expression products cannot be accurately predicted from the genome sequence alone. The recent developments in next generation sequencing and shotgun proteomics have made it possible to form a global view of gene expression at both RNA and protein level. However, describing the molecular properties of the individual proteins *in vivo*, such as their native expression dynamics, physical interactions and localization, is still difficult, especially in multicellular organisms.

The inherent variations in protein size, charge and folding present a challenge to the systematic generation of reagents that specifically recognize individual proteins such as antibodies, which are also of limited use for interrogation of protein function *in vivo*. A generic solution to this problem is the use of fluorescent and affinity tags that, when fused to the protein of interest, allow *in vivo* visualization and use of standard affinity reagents to identify interactions (Chalfie et al., 1994; Rigaut et al., 1999). In *Saccharomyces cerevisiae*, where homologous gene targeting is efficient, nearly all protein-coding genes have been tagged at their endogenous genomic loci, allowing proteome scale analysis with standard, reproducible and comparable tag based assays (Gavin et al., 2002; Ghaemmaghami et al., 2003; Huh et al., 2003; Gavin et al., 2006; Krogan et al., 2006). Development of comparable reagents for multicellular species would enhance our understanding of the diversity of cell form and function. However, technical challenges have so far precluded genome-wide tagging of animal genomes.

We have previously established the use of large genomic DNA (gDNA) constructs such as BACs and fosmids as transgenes for protein function discovery (Muyrers et al., 1999; Sarov et al., 2006; Poser et al., 2008). These constructs carry sufficiently large sections of gDNA to include all of the important coding and regulatory sequences for most genes (Ristevski, 2005). The major advance allowing engineering of these large constructs was the development of *in vivo* homologous recombination mediated cloning, or recombineering (Zhang et al., 1998; Copeland et al., 2001), which permits the insertion of a tag coding sequence at any position, independent of the presence of restriction sites. More recently we showed that the fidelity of recombineering is sufficient to allow multistep transgene engineering in liquid culture (Sarov et al., 2006; Poser et al., 2008; Ejsmont et al., 2009; 2011), thereby enabling the throughput required for genome scale recombineering.

The nematode *C. elegans* is an attractive system for genome-wide application of a tagged gDNA transgene approach for protein function discovery. It's transparency, rapid generation time, fixed anatomy and reproducible cell lineage (Hillier et al., 2005) make it possible to study protein localization in all cells of the animal. The compact and wellannotated genome is covered by a mapped fosmid gDNA clone library. Fosmid-sized clones are normally sufficient to rescue loss of function mutations, and have been used on a large scale for mutation mapping (Janke et al., 1997). We and others have developed protocols for recombineering based transgene construction in *C. elegans* (Dolphin and Hope, 2006; Sarov et al., 2006; Tursun et al., 2009).

Here we present a genome-wide application of high throughput recombineering to build an efficient platform for *in vivo* tag based protein function analysis under native regulatory control in *C. elegans* and illustrate its utility for analyses of both protein localization at cellular level through high resolution *in vivo* imaging and physical interactions through biochemical purification. We refer to the platform and analysis as the "*C. elegans* TransgeneOme."

## Results

### Generation of the *C. elegans* TransgeneOme resource

The use of large gDNA constructs as "third allele" transgenes (Sarov and Stewart, 2005) expressing a fluorescent and/or affinity tagged version of the protein of interest, provides a generic platform for function exploration under endogenous regulatory control (Figure 1-A). To facilitate the large scale application of this approach we generated a genome wide resource of fosmid gDNA transgenes for the protein coding genes in the *C. elegans* genome, tagged at the C terminus with a fluorescent and affinity cassette consisting of the Ty1 peptide, GFP and 3xFLAG(Figure 1B).

Of the approximately 20000 protein coding genes in the *C. elegans* genome, 16,403 are currently covered by a fosmid clone that can be engineered into a transgene (Figure 1C), which defined our target set - “the taggable genome”. We used a streamlined, liquid culture recombineering pipeline in a 96 well format (Figure 1-D, Supplementary figure 1) with the selection of a suitable fosmid clone, primer design and sample tracking handled automatically, which allowed us to process thousands of clones in parallel with a throughput of over 4000 constructs/month. The success rate of the transgene engineering, as estimated by the number of viable wells in selection, was between 92 and 99% on all 4 steps of the pipeline, resulting in a cumulative success rate of 84% (Figure 1-D and Supplemental Table 1). The 16% that failed were subjected to a second run through the pipeline and 88% of these were successful, indicating that most of the initial failures were technical, rather than systemic. After the two rounds of recombineering we had successfully processed 98% of the taggable genome (16102 of 16403 genes).

To evaluate the accuracy of the engineering, we developed a high throughput validation method based on next generation sequencing (NGS). After pooling the 96 well pates from the final step, we isolated fosmid DNA from the pools, followed by fragmentation, barcoding, massively parallel paired end sequencing and mapping to the expected construct sequences. The correct tag insertion was confirmed by paired end reads spanning through the 5' and 3' tag/gene junctions for 14637 of the transgenes (90.7% of the viable wells or 89% of the taggable genome (Figure 1 E-G). As transgene engineering was carried out in liquid, the 96 well plate cultures are non-clonal and carry over products from the intermediate steps were also detected, however for most constructs they were present as a minor fraction only (Figure 1 G). To evaluate the fidelity of recombineering at the single clone level we isolated colonies on selective agar from 1205 NGS validated wells evenly distributed throughout the resource and Sanger sequenced the full extent of tag plus 100 bp of the flanking genomic sequence. Of the 918 clones for which we obtained data of sufficient quality throughout the entire region 83.1% exactly matched the expected region, 7.4% had mismatches (due to either sequencing noise or point mutations caused by PCR), 4.4% were intermediary products, and 5.1% were cross-contaminants from other wells (Figure 1H). Sequencing of a second colony from the 155 failed wells resulted in exact match for 59% of the 141 successfully sequenced clones.

We also validated the correct insertion of the *unc-119* marker cassette at the single clone level (Figure 1H). Of the 1067 reads of sufficient quality 89% matched the expected junction, 10% contained mixed sequences downstream of the insertion point, indicating a coexistence of the final constructs with carryover of intermediary products within the same cell, which occurs normally at low rates in recombineering and only less than 1% did not align to the expected region.

The TransgeneOme resource can be explored through a dedicated web application available at http://transgeneome.mpi-cbg.de, which provides full access to all construct related

information, including the validation data, sequence maps in FASTA, genbank or GFF format, as well as genome browser views and links to Wormbase (see Movie 1). To make the resource easily accessible for researchers that are more familiar with other model systems searching is possible with either the *C. elegans* gene name, or ortholog gene names from *C. briggsae, D. melanogaster, M. musculus* or *H. sapiens*.

## Stable gDNA transgenesis delivers native levels, patterns and dynamics of tagged protein expression

To generate worm lines expressing the tagged proteins of interest we used micro particle bombardment (Wilm et al., 1999; Praitis et al., 2001), which provides high throughput and can result in low copy stable integration of plasmid transgenes into the genome. We adapted the method for use with the larger fosmid transgenes (Figure 2A) and we generated 254 stable lines for 230 genes (Supplemental table 2). The localization data can be explored using the TransgeneOme web application (https://transgeneome.mpi-cbg.de) either by gene name or through a localization ontology browser, using standard, Wormbase compatible terms for developmental stage, anatomy and subcellular localization (see also Movie 1).

Using next generation sequencing we estimated the copy number of the regions covered by the transgenes in 33 lines and we found that despite their large size the full length of the fosmid transgenes can be integrated into the genome, with most of the transgenes present at a low copy number (Figure 2B and Supplemental figure 2). Western blot analysis with anti-GFP antibody showed that the tagged proteins are expressed at the expected molecular sizes, including known or predicted alternative isoforms (Figure 2D). We also compared the transgene derived protein expression levels to that of the endogenous gene using western blot with protein specific antibodies (Figure 2E). In all 3 cases (AIR-1, SPD-2, BUB-1), the tagged protein was expressed at roughly the same level as the endogenous counterpart.

Using live GFP fluorescence microscopy we observed tagged protein expression in many cell types and all life stages, including the germ line and the early embryo, where transgene silencing is often a problem in *C. elegans* (Kelly and Fire, 1998). When multiple expressing lines were obtained per gene, the expression patterns appeared similar. Quantitative image analysis of embryonic expression (see bellow) from two independent lines for five genes showed similar reproducibility as that seen between embryos from the same line (mean r=0.82, fraction of replicated cells = 0.97, Supplemental table 3).

On a subcellular level, we observed specific localization to a wide variety of compartments (Figure 3A), including cell membrane; cortex; centrosomes; spindle; kinetochores; mitochondria; throughout the nucleus or in specific subcompartments – nuclear envelope, nucleolus, heterochromatin or specific foci of unknown function; throughout the cytoplasm or in P granules (germ line specific cytoplasmic RNA-protein complexes); cytoskeleton and specialized structures such as muscle fibers. For 112 of the 230 analyzed proteins localization data form other experiments (including antibody staining) was available and all but 3 of the patterns matched the Wormbase description. Even though our test set was biased towards gene expression regulators and structural proteins, which are relatively well studied, over half of the patterns were new for *C. elegans*. Most of the patterns for which there was no data available in Wormbase (107 of 118) were consistent with the known or predicted function of the proteins or the localization of their orthologs in other species. However, there were also unexpected patterns that provide insight into the function of previously unstudied proteins or improve or correct earlier functional predictions. For example, M04F3.5 (Figure 3A), is the only *C. elegans* protein with an I BAR domain, a domain that has been shown to induce plasma membrane folding (Saarikangas et al., 2009). Consistent with this function we observe M04F3.5 at specific membrane foci in many tissues, particularly in neurons. An example of unexpected localization was SIR-2.2, a homolog of the yeast Sir2p NAD-

dependent histone deacetylase, which has not been previously localized in *C. elegans* but the phenotypic effects of it's loss in genetic screens have been attributed to it's presumed role as a chromatin structure regulator (Pothof et al., 2003; Bates et al., 2006). Surprisingly, instead of the nucleus SIR-2.2::GFP localizes to large cytoplasmic organelles, which we identified as mitochondria by costaining (Figure 3A and Supplemental Figure 3). Mitochondrial sirtuins with NAD-dependent deacetylase activity have been described in human cells (Schwer et al., 2002; Huang et al., 2010), and SIR- 2.2 is likely a functional homolog in *C. elegans*. This hypothesis is supported by the fact that of the 4 worm sirtuins SIR-2.2 is most similar to the human mitochondrial sirtuin hSirt4.

The tagged proteins also correctly reproduced examples of known dynamic subcellular localization patterns, both throughout the cell cycle and in response to signals. For example, the Aurora B homolog AIR-1 showed dynamic relocalization from chromatin to the spindle midzone, consistent with its function in both chromosome segregation and cytokinesis; the mitotic checkpoint protein BUB-1 localizes to the condensing chromosomes; P granule components like GLH-1 condense from the cytoplasm to droplet like particles (P granules) that are segregated to the germline precursor cell in the first cell division. We also observed specific relocalization dynamics in response to signaling: the *C. elegans* Notch homolog LIN-12 was generally localized to cell membranes throughout the embryo, but a nuclear signal was also detectable for several cells (Figure 3C). Despite its known importance in development (Greenwald, 1998), Notch transition to the nucleus as a result of signaling has not been previously observed *in vivo* in *C. elegans*, however it is known that the intracellular domain of LIN- 12 localizes to the nucleus (Struhl et al., 1993) and cells with active Notch signaling are known from genetic studies (Priess, 2005). Using automated lineage tracing (Murray et al., 2008) we identified nuclear signal in ABplaaa, ABplpap, and the lineages derived from Eal and Epl, all of which have being previously described as targets of Notch signaling(Priess, 2005). In addition, nuclear LIN-12::GFP signal was detectable for the MSpap lineage, which was has not been previously shown to be regulated by Notch signaling (Figure 3D).

## Protein localization dynamics reveal additional levels of gene expression controls in comparison with promoter::reporter constructs

As gene expression reporters the fosmid transgenes have the advantage of preserving the native *cis* regulatory elements, including intronic and 3'UTR sequences. The tagged proteins expression is further controlled by the specific rates of translation and protein degradation. We wanted to know how that would reflect in the overall gene expression pattern when compared with promoter reporters that lack some of the *cis* regulatory elements and do not reflect any of the posttranscriptional regulatory mechanisms.

Several imaging techniques have recently been developed to automatically and quantitatively assign fluorescent signals to specific cells by mapping to an anatomy atlas or lineage tracing(Bao et al., 2006; Murray et al., 2008; Long et al., 2009). We used lineage tracing to compare the expression pattern of 12 genes (*cnd-1, die-1, egl-5, elt-2, hlh-1, hnd-1, lin-39, med-2, nhr-2, pal-1, pha-4, tbx-8*) as either tagged proteins or promoter::histone::GFP reporters (Figure 4). In replicate embryos for each of the reporter types the patterns were 94% reproducible (Supplemental Table 3). However, in comparison between the two types of reporters only 62% of the cells expressing in one type were shared in the other. For some genes with poor correlation between the two types of reporters (*med-2, nhr-2, hnd-1, die-1*), manual curation of the data suggest that the patterns were similar and the differences were due to the low levels of expression, dropping to and bellow the limit of the automatic detection method. However, there were clear differences in pattern for several genes, which appear to be biologically significant.

Some genes were expressed in more cells as tagged protein than as promoter reporter, suggesting that the fosmid transgene contains additional regulatory elements. For example PAL-1 was expressed as a tagged protein in posterior ectodermal precursors, derived from the AB lineage in addition to the expression pattern in the C and D lineages revealed by the promoter reporter (Figure 4B). Other examples include LIN-39 (Figure 4A) in a subset of the C lineage and TBX-8 in a subset of the MS lineage.

For other genes, the tagged protein was only expressed in a subset of the cells expressing the promoter fusion, which at least in some cases appears to reflect lineagedependent differences in post-translational control. For example, the CND-1 protein was initially expressed in the same cells as the promoter fusion, but subsequently disappeared in many of them (Figure 4C). Most of the cells with persistent expression were precursors of D-type motor neurons, which have been reported to be defective in cnd-1 mutants (Hallam et al., 2000). Similarly, for the master pharyngeal regulator pha-4 both the promoter reporter and the tagged protein were initially expressed in broad, primarily pharynx-generating lineages, but the protein was then specifically lost in the few cells born from these lineages that adopt a non-pharyngeal fate. A surprising example of lineal regulation of turnover occurs for PAL-1. Both PAL-1 constructs were expressed throughout the D lineage, which produces only muscle. The tagged PAL-1 protein was then selectively lost in two cells at the D8 stage (when D has 8 descendants) and retained in the other six (Figure 4D). This loss was rapid, occurring in less than 10 minutes (Figure 4F), suggesting that an active mechanism exists to remove PAL-1 specifically in these lineages. This is surprising because all D lineage-derived cells become muscle, including the two that lose PAL-1:GFP, and implies that important molecular differences exist between different muscle cells in this lineage.

For the MyoD ortholog hlh-1 (Chen et al., 1994; Fukushige and Krause, 2005), the tagged protein was completely restricted to muscle precursors, while the promoter reporter was expressed both in muscle precursors and most other cells in the MS lineage, including pharynx, coelomocyte and other mesodermal precursors (Figure 4G). In this case, we saw late muscle expression but no broad early expression of the tagged protein, suggesting that *hlh-1* is transcribed in MS, but its translation or accumulation is limited to muscle precursors, supporting the model for *hlh-1* regulation established in previous work (Krause, 1990). Other genes, including the Hox gene lin-39 (Figure 4H) showed similar types of restriction. These are likely to include differences in translational regulation between cells, but could also include transcriptional regulatory elements outside the promoter region.

## High resolution mapping of transcription factor protein distribution

The ability to rapidly generate transgenic lines for any of the thousands of tagged genes opens up the possibility for systematic exploration of protein function for large sets of proteins. As part of the modENCODE initiative we used this approach to systematically explore transcription factor (TF) protein function in *C. elegans* (Celniker et al., 2009; Gerstein et al., 2010; Zhong et al., 2010; Niu et al., 2011). Using tag based Chromatin IP coupled with next generation sequencing we identified hundreds of thousands of TF binding sites (Figure 4A and Supplemental table 4) at various stages of development. All the generated binding data is available to the community at www.modencode.org and the Wormbase genome browser.

Understanding TF function from the DNA binding patterns alone is difficult if not impossible without knowledge of their expression patterns *in vivo*. We used automated lineage tracing (Murray et al., 2008) to generate quantitative 4D expression maps of 28 TF proteins, at 1.5 minute temporal resolution during embryonic development up to the 350 cell stage (Figure 5B and http://epic.gs.washington.edu). Many of these patterns are either new or add a higher spatiotemporal resolution to otherwise know localizations. Due to the

invariant cell lineage in *C. elegans*, the independently obtained patterns can be easily superimposed in both time and space. For example, overlay of the expression patterns of the HOX transcription factors LIN-39, MAB-5 and EGL-5 in the C muscle lineages around 100 cells stage shows the expected anterior-posterior distribution pattern (Figure 5C). We observed similar anterior-posterior gradients in other lineages (Figure 5C and Supplemental figure 4), however the expression domains of the different lineages did not overlap, reinforcing the previously established idea that in *C. elegans* Hox gene expression patterns are determined primarily by lineage specific mechanisms rather than by position (Cowing and Kenyon, 1996).

## Global cell cycle and tissue dynamics of TF expression suggest a developmental timer circuit

The lineage mapping dataset covers hundreds of cell cycles, which allowed us to look at some global properties of TF localization dynamics. Most of the studied TF proteins appear weakly with chromatin in mitosis, as they rapidly diffuse from the nucleus at nuclear envelope breakdown (Figure 4D) and are then imported back with protein specific rates that are consistent throughout the various lineages expressing the factors (supplemental figure 5). One interesting observation was that the expression or nuclear localization onset appears to be synchronized with the cell cycle for some TF proteins. For example, the onset of FKH-4::GFP expression did not occur in the middle or end of a cell cycle for any of the expressing cells (Figure 6B). The initial accumulation kinetics was very reproducible in all cells expressing FKH-4 and always began rapidly and immediately after mitosis in multiple lineages. Other proteins (CND-1, CEH-6, NHR- 25) showed similar cell cycle synchronization at the expression onset, suggesting this may be a common mechanism to ensure appropriate temporal expression (Figure 6C).

Some proteins had dynamic expression patterns across large lineages or even the entire embryo. A striking example of lineage-dependent differences in dynamics is PAL-1 (Figure 6A,B). Previous work has established that after being inherited maternally, PAL-1 protein is then restricted by posttranscriptional mechanisms to the posterior lineages C and D, where it specifies proper lineage identity and reactivates its own zygotic expression (Hunter and Kenyon, 1996). We observed the maternal expression, and loss in the AB lineage as expected (Figure 6A). Surprisingly, we saw essentially identical early expression levels in the C lineage as well as the more anterior E and MS lineages (Figure 6B). Only later, after the presumed onset of zygotic expression, does expression become stronger in C. These embryos developed normally, indicating that high PAL-1 is not sufficient to activate C and D lineage fate. Mutants defective in E and MS specification produce excess C-like cells (Bowerman et al., 1992).

Some genes showed temporally dynamic expression synchronously in all cells. A striking example is FKH-4::GFP (Figure 6C,D Movie S2), which accumulates in all nuclei of the embryo after they have undergone 5 cell divisions (except for the EMS lineage where protein accumulates after 4 cell divisions). Protein levels peak about 30 minutes after their onset and return to background levels in all cells within about an hour (~2 cell cycles). Reporters for nhr-2, die-1 were broadly expressed with peaks 1–2 cell cycles after that of fkh-4, suggesting they could also be under temporal control (Figure 6D and Movie S3). It's important to note that an analysis of promoter::reporter constructs would not have revealed these dynamics. For example, a promoter reporter construct for nhr-2, analyzed by the same methods, had a similar time of expression onset but the signal persisted and gradually increased throughout development rather than peaking at a precise developmental time (Supplemental figure 5).

## Discussion

The *C. elegans* TransgeneOme provides a platform for large-scale, tag-based protein function exploration under endogenous regulatory control in one of the beststudied multicellular model organisms. The dramatic scale up of DNA engineering and validation throughput presented here will be useful in other large scale transgenic and synthetic biology approaches and we are currently working to extend this approach to other model systems (Hofemeister et al., 2010; Ciotta et al., 2011; Ejsmont et al., 2011). The development of comprehensive resources of standard reagents and assays make it easy to reproduce and compare results from independent studies and enables a quantitative Systems Biology approach towards integration of the primary datasets.

We show the utility of the TransgeneOme platform for interrogation of transcription factor protein function *in vivo*. We observed rapid protein localization and turnover dynamics, including nuclear translocation (LIN-12/Notch) and cell specific protein degradation that would not have been detectable with other methods. Cell fates in Nematodes are primarily controlled by the cell origin (lineage), as illustrated by our comparative analysis of the spatial distribution of the Hox TF proteins at the 100 cells stage. Ensuring correct timing and synchrony in a system with primarily lineal mode of development may require an internal molecular timer circuit. Some of the TF protein expression patterns we observed - broad, synchronous, with rapid onset and offset hint towards such a mechanism. More work would be required to determine if the factors we identified are actively regulating temporal identity or merely responding to other temporal regulators. Interestingly, for several TFs we observed rapid increases of detectable nuclear protein within a few minutes following the cell division that produces the expressing cells, suggesting that cell division itself may be a mechanism for synchronizing temporal identity with gene expression.

The transgenes have many applications beyond protein localization. The patterns can serve as sensors in loss of gene function, drug screening or other system perturbations, which will dramatically expand the accessible phenotypic space. The tag we used is compatible with previously established methods that use protein purification coupled with mass spectrometry for protein-protein interactions discovery (Polanowska et al., 2004; Cheeseman and Desai, 2005; Poser et al., 2008; Hutchins et al., 2010). For proteins that physically interact with DNA, tag based chromatin IP is an efficient approach to map their binding sites, without the need to develop highly specific antibodies. We are systematically applying this approach to describe transcription factor binding sites *in vivo*. The properties of the emerging TF networks (Zhong et al., 2010; Niu et al., 2011), show a good correlation between the known TF function and their putative target genes. We have now extended these resources to hundreds of thousands of binding sites, which are accessible to the community through Wormbase and Modmine, and are continuously updated.

As any other tool the TransgeneOme still has potential limitations that should be taken into account when the results are interpreted. The endogenous expression levels for some proteins could be under the detection limit for GFP fluorescence microscopy or they may not be expressed at all under laboratory conditions. All genes in the resource were tagged at the C terminus, which may affect the function or stability of some proteins. Our results and previous work in *C. elegans* and other model systems has shown that such cases are rare – the transgenes have expression patterns consistent with their function and can rescue loss of function phenotypes (Voutev et al., 2009; Petersen et al., 2011). However, if required an N terminal tag (Supplemental figure 1) can be used instead. The fosmid transgenes provide very reproducible expression and appear more resistant to germline silencing, however multiple lines should be compared if conclusions are drawn based on expression pattern alone. Another issue is the still incomplete genome coverage of the resource, due to the lack

of suitable gDNA clones for close to 4000 genes. However, almost all of them can be fully contained in a fosmidsized construct. Extending the available gDNA libraries and alternative approaches such as targeted clone retrieval (Nedelkova et al., 2011) can bring us closer to a comprehensive resource. Recently, several new methods for site specific and targeted homologous recombination have been developed in *C. elegans* (Robert and Bessereau, 2007; Frøkjær-Jensen et al., 2008). Other tools such Zn finger or TALEN nucleases(de Souza, 2011) and the phiC31 integrase (Wirth et al., 2007), which are successfully used for transgenesis in other model systems should also be applicable to *C. elegans*. Changes in the TransgeneOme resource that would make it compatible with these techniques could be easily engineered. The ability to further modify the transgenes using recombineering adds to the value of the resource. For example the transgenes can be engineered to contain specific mutations for structure-function analysis or investigation of the role of *cis* regulatory sequences.

The availability of the TransgeneOme resource cuts down the time for generation of tagged protein expressing lines to just a few weeks, and opens up the road towards a proteome wide map of protein function in *C. elegans*. While the generation of thousands of transgenic worm lines is still a challenge, a distributed community-wide effort makes it feasible in a relatively short time. By providing this resource to the community, we hope to stimulate cooperation towards this common goal.

## Experimental procedures

### Bioinformatics

The clones containing genes of interest were chosen such that the ratio of upstream to downstream sequences is as close to 2:1 as possible. For genes part of operons the same rules were applied to the entire operon. The Sanger sequencing results were aligned to the theoretical construct using Supermatcher from EMBOSS package (Rice et al., 2000). Paired-end sequencing data was mapped to the expected sequence of the tagged genes in single-end mode using Mosaik.

### Transgene construction

Fosmid transgenes were constructed by liquid culture recombineering as previously described (Sarov et al., 2006), but the unc-119 marker was inserted into the fosmid backbone in the last step of the pipeline and all steps were performed in a 96 well plate format. (See Supplemental experimental procedures and Supplemental figure 1). Sequence maps and the validation data for all constructs is available at http://transgeneome.mpi-cbg.de

### Strain generation

Transgenic strains were made using microparticle bombardment essentially as described (Praitis et al., 2001), except that 20–50 μg of fosmid DNA was used per transformation. Each 100mm worm plate was bombarded twice with the same DNA construct using the Biorad Biolistic PDS with Hepta adapter.

### Expression analysis

We crossed the GFP protein-fusion reporter into a lineaging strain, RW10226, which expresses a histone H2B::mCherry fusion from a pie-1 promoter in the germline, and a histone H1.1::mCherry fusion from the his-72 promoter in all somatic embryonic cells. We used a Zeiss LSM510 microscope to collect 31 z-planes at 1-micron intervals every 1 or 1.5 minutes from the 4-cell stage through the onset of movement. Quantitation, analysis and

display of the resulting images were otherwise performed as described previously (Murray et al., 2008). The promoter reporter lines were published elsewhere (Murray et al., 2012)

### Transcription factor binding sites mapping by ChIP/Seq

ChIP/Seq was performed in duplicates as previously described (Zhong et al., 2010). In brief, staged worms were treated with formaldehyde, sonicated and the GFP tagged TFs were affinity purified using polyclonal anti GFP antibody. The copurified DNA was sequenced on the Illumina platform and the data was analyzed using Peakseq (Rozowsky et al., 2009). All of the binding data is available from the modENCODE website (See also supplemental table 4).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The TransgeneOme Project was partially funded by the Max Planck Society Initiative “BAC TransgeneOmics”. The ModENCODE project was funded by NIH grants to M.Sn., R.H.W. and J.I.M. J.I.M. was a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

## References

Bao Z, Murray J, Boyle T, Ooi S, Sandel M, Waterston R. Automated cell lineage tracing in Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2006; 103:2707–2712. [PubMed: 16477039]

Bates E, Victor M, Jones A, Shi Y, Hart A. Differential contributions of Caenorhabditis elegans histone deacetylases to huntingtin polyglutamine toxicity. Journal of Neuroscience. 2006; 26:2830. [PubMed: 16525063]

Bowerman B, Eaton BA, Priess JR. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. Cell. 1992; 68:1061–1075. [PubMed: 1547503]

Celniker SE, Dillon LAL, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, Macalpine DM, et al. Unlocking the secrets of the genome. Nature. 2009; 459:927. [PubMed: 19536255]

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994; 263:802–805. [PubMed: 8303295]

Cheeseman IM, Desai A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. Sci STKE. 2005; 2005:pl1. [PubMed: 15644491]

Chen L, Krause M, Sepanski M, Fire A. The Caenorhabditis elegans MYOD homologue HLH-1 is essential for proper muscle function and complete morphogenesis. Development. 1994; 120:1631–1641. [PubMed: 8050369]

Ciotta G, Hofemeister H, Maresca M, Fu J, Sarov M, Anastassiadis K, Stewart AF. Recombineering BAC transgenes for protein tagging. Methods. 2011; 53:113–119. [PubMed: 20868752]

Copeland N, Jenkins N, Court D. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet. 2001; 2:769–779. [PubMed: 11584293]

Cowing D, Kenyon C. Correct Hox gene expression established independently of position in Caenorhabditis elegans. Nature. 1996; 382:353–356. [PubMed: 8684464]

de Souza N. Primer: genome editing with engineered nucleases. Nat Methods. 2011; 9:27–27. [PubMed: 22312638]

Dolphin CT, Hope IA. Caenorhabditis elegans reporter fusion genes generated by seamless modification of large genomic DNA clones. Nucleic Acids Res. 2006; 34:e72. [PubMed: 16717278]

Dunn RK, Kingston RE. Gene regulation in the postgenomic era: technology takes the wheel. Mol Cell. 2007; 28:708–714. [PubMed: 18082595]

Ejsmont RK, Ahlfeld P, Pozniakovsky A, Stewart AF, Tomancak P, Sarov M. Recombination-mediated genetic engineering of large genomic DNA transgenes. Methods Mol Biol. 2011; 772:445–458. [PubMed: 22065454]

Ejsmont RK, Sarov M, Winkler S, Lipinski KA, Tomancak P. A toolkit for high-throughput, cross-species gene engineering in Drosophila. Nat Methods. 2009; 6:435–437. [PubMed: 19465918]

Frøkjær-Jensen C, Wayne Davis M, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, Grunnet M, Jorgensen EM. Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet. 2008; 40:1375–1383. [PubMed: 18953339]

Fukushige T, Krause M. The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. Development. 2005; 132:1795–1805. [PubMed: 15772130]

Gavin A, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen L, Bastuck S, Dumpelfeld B, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006; 440:631–636. [PubMed: 16429126]

Gavin A, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick J, Michon A, Cruciat C, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. 2002; 415:141–147. [PubMed: 11805826]

Gerstein MB, Lu ZJ, van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010; 330:1775–1787. [PubMed: 21177976]

Ghaemmaghami S, Huh W, Bower K, Howson R, Belle A, Dephoure N, O'Shea E, Weissman J. Global analysis of protein expression in yeast. Nature. 2003; 425:737–741. [PubMed: 14562106]

Greenwald I. LIN-12/Notch signaling: lessons from worms and flies. Genes Dev. 1998; 12:1751–1762. [PubMed: 9637676]

Hallam S, Singer E, Waring D, Jin Y. The C. elegans NeuroD homolog cnd-1 functions in multiple aspects of motor neuron fate specification. Development. 2000; 127:4239–4252. [PubMed: 10976055]

Hillier L, Coulson A, Murray J, Bao Z, Sulston J, Waterston R. Genomics in C. elegans: so many genes, such a little worm. Genome Res. 2005; 15:1651–1660. [PubMed: 16339362]

Hofemeister H, Ciotta G, Fu J, Seibert PM, Schulz A, Maresca M, Sarov M, Anastassiadis K, Stewart AF. Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis. Methods. 2010

Huang J, Hirschey M, Shimazu T, Ho L, Verdin E. Mitochondrial sirtuins. Biochimica Et Biophysica Acta (BBA)-Proteins & Proteomics. 2010; 1804:1645–1651.

Huh W, Falvo J, Gerke L, Carroll A, Howson R, Weissman J, O'Shea E. Global analysis of protein localization in budding yeast. Nature. 2003; 425:686–691. [PubMed: 14562095]

Hunter CP, Kenyon C. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. Cell. 1996; 87:217–226. [PubMed: 8861906]

Hutchins JRA, Toyoda Y, Hegemann B, Poser I, Hériché J-K, Sykora MM, Augsburg M, Hudecz O, Buschhorn BA, Bulkescher J, et al. Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science. 2010; 328:593–599. [PubMed: 20360068]

Janke DL, Schein JE, Ha T, Franz NW, O'Neil NJ, Vatcher GP, Stewart HI, Kuervers LM, Baillie DL, Rose AM. Interpreting a sequenced genome: toward a cosmid transgenic library of Caenorhabditis elegans. Genome Res. 1997; 7:974–985. [PubMed: 9331368]

Kelly W, Fire A. Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development. 1998; 125:2451. [PubMed: 9609828]

Krause M. CeMyoD accumulation defines the body wall muscle cell fate during C. elegans embryogenesis. Cell. 1990; 63:907–919. [PubMed: 2175254]

Krogan N, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006; 440:637–643. [PubMed: 16554755]

Long F, Peng H, Liu X, Kim S, Myers E. A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat Methods. 2009

Murray J, Bao Z, Boyle T, Boeck M, Mericle B, Nicholas T, Zhao Z, Sandel M, Waterston R. Automated analysis of embryonic gene expression with cellular resolution in C. elegans. Nat Methods. 2008

Murray JI, Boyle TJ, Preston E, Vafeados D, Mericle B, Weisdepp P, Zhao Z, Bao Z, Boeck M, Waterston RH. Multidimensional regulation of gene expression in the C. elegans embryo. Genome Research. 2012

Muyrers JP, Zhang Y, Testa G, Stewart AF. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 1999; 27:1555–1557. [PubMed: 10037821]

Nedelkova M, Maresca M, Fu J, Rostovskaya M, Chenna R, Thiede C, Anastasiadis K, Sarov M, Stewart AF. Targeted isolation of cloned genomic regions by recombineering for haplotype phasing and isogenic targeting. Nucleic Acids Res. 2011

Niu W, Lu ZJ, Zhong M, Sarov M, Murray JI, Brdlik CM, Janette J, Chen C, Alves P, Preston E, et al. Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. Genome Res. 2011; 21:245–254. [PubMed: 21177963]

Petersen SC, Watson JD, Richmond JE, Sarov M, Walthall WW, Miller DM. A transcriptional program promotes remodeling of GABAergic synapses in Caenorhabditis elegans. J Neurosci. 2011; 31:15362–15375. [PubMed: 22031882]

Polanowska J, Martin J, Fisher R, Scopa T, Rae I, Boulton S. Tandem immunoaffinity purification of protein complexes from Caenorhabditis elegans. Biotechniques. 2004; 36:778–780. 782. [PubMed: 15152596]

Poser I, Sarov M, Hutchins JRA, Hériché J-K, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods. 2008; 5:409–415. [PubMed: 18391959]

Pothof J, van Haaften G, Thijssen K, Kamath R, Fraser A, Ahringer J, Plasterk R, Tijsterman M. Identification of genes that protect the C. elegans genome against mutations by genome-wide RNAi. Genes Dev. 2003; 17:443. [PubMed: 12600937]

Praitis V, Casey E, Collar D, Austin J. Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics. 2001; 157:1217–1226. [PubMed: 11238406]

Priess J. Notch signaling in the C. elegans embryo. WormBook, the C Elegans Research Community, Ed. 2005; 10

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000; 16:276–277. [PubMed: 10827456]

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 1999; 17:1030–1032. [PubMed: 10504710]

Ristevski S. Making better transgenic models: conditional, temporal, and spatial approaches. Mol Biotechnol. 2005; 29:153–163. [PubMed: 15699570]

Robert V, Bessereau J-L. Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. Embo J. 2007; 26:170–183. [PubMed: 17159906]

Saarikangas J, Zhao H, Pykäläinen A, Laurinmäki P, Mattila PK, Kinnunen PKJ, Butcher SJ, Lappalainen P. Molecular mechanisms of membrane deformation by I-BAR domain proteins. Curr Biol. 2009; 19:95–107. [PubMed: 19150238]

Saghatelian A, Cravatt BF. Assignment of protein function in the postgenomic era. Nat Chem Biol. 2005; 1:130–142. [PubMed: 16408016]

Sarov M, Stewart AF. The best control for the specificity of RNAi. Trends in Biotechnology. 2005; 23:446–448. [PubMed: 15979179]

Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nat Methods. 2006; 3:839–844. [PubMed: 16990816]

Schwer B, North BJ, Frye RA, Ott M, Verdin E. The human silent information regulator (Sir)2 homologue hSIRT3 is a mitochondrial nicotinamide adenine dinucleotide–dependent deacetylase. The Journal of Cell Biology. 2002; 158:647. [PubMed: 12186850]

Struhl G, Fitzgerald K, Greenwald I. Intrinsic activity of the Lin-12 and Notch intracellular domains in vivo. Cell. 1993; 74:331–345. [PubMed: 8343960]

Tursun B, Cochella L, Carrera I, Hobert O. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in C. elegans. PLoS ONE. 2009; 4:e4625. [PubMed: 19259264]

Voutev R, Keating R, Hubbard EJA, Vallier LG. Characterization of the Caenorhabditis elegans Islet LIM-homeodomain ortholog, lim-7. FEBS Lett. 2009; 583:456–464. [PubMed: 19116151]

Wilm T, Demel P, Koop H, Schnabel H, Schnabel R. Ballistic transformation of Caenorhabditis elegans. Gene. 1999; 229:31–35. [PubMed: 10095101]

Wirth D, Gama-Norton L, Riemer P, Sandhu U, Schucht R, Hauser H. Road to precision: recombinase-based targeting technologies for genome engineering. Curr Opin Biotechnol. 2007; 18:411–419. [PubMed: 17904350]

Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 1998; 20:123–128. [PubMed: 9771703]

Zhong M, Niu W, Lu ZJ, Sarov M, Murray JI, Janette J, Raha D, Sheaffer KL, Lam HYK, Preston E, et al. Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. PLoS Genet. 2010; 6:e1000848. [PubMed: 20174564]

## Highlights

- A genome wide resource for *in vivo* expression of tagged proteins was engineered The tagged gene alleles provide native protein expression and localization patterns
- Tag based ChIP provides genome wide DNA binding site maps for key transcription factors
- Live 4D tracing reveals rapid transcription factor protein localization Dynamics

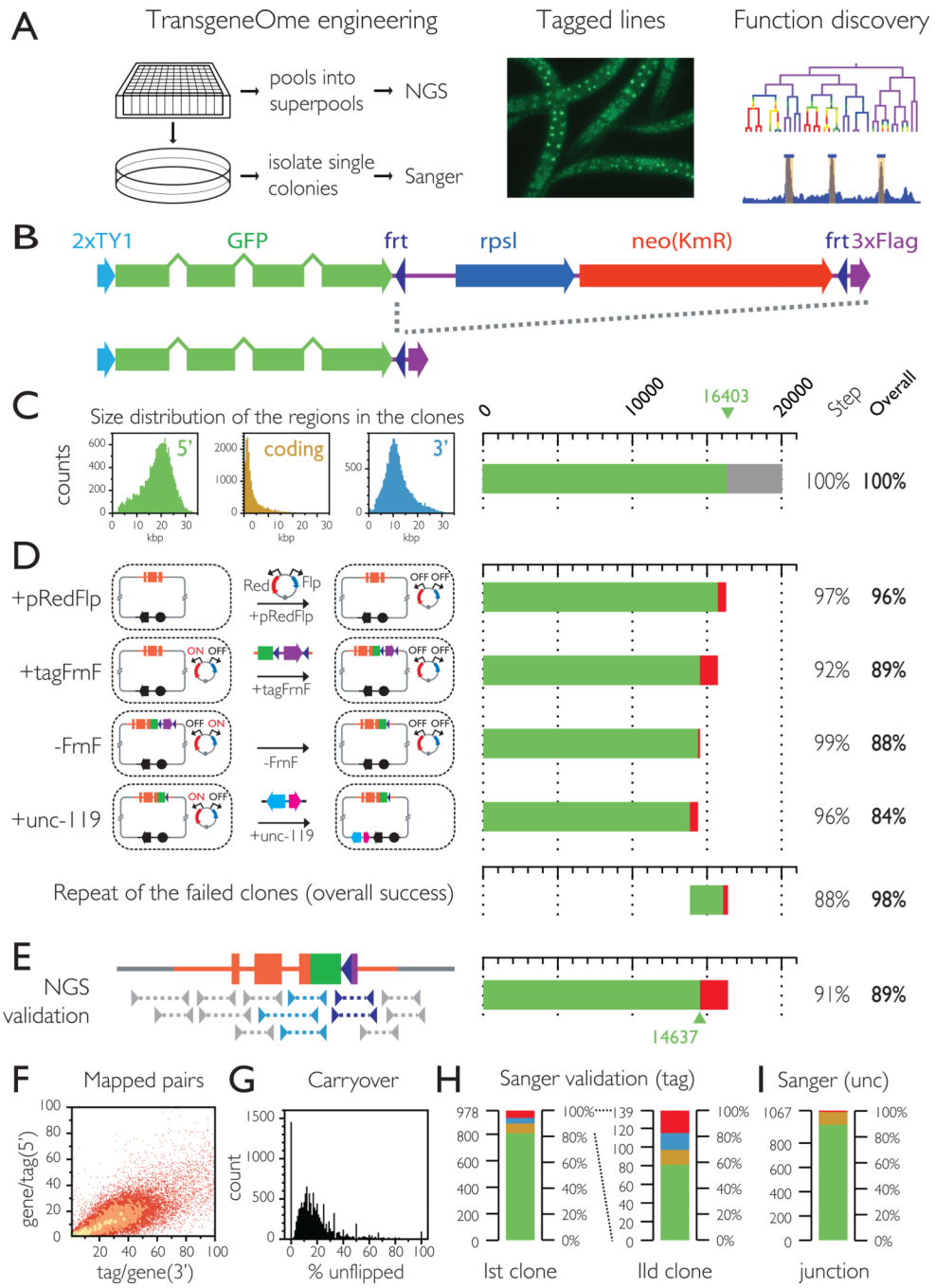


**Figure 1.**

Generation of the *C. elegans* TransgeneOme

**A**. Transgenic strategy for protein function discovery. Fosmid transgenes are generated in a high throughput 96 well pipeline, and validated by either high throughput NGS mapping or at the single clone level by Sanger sequencing. The gDNA transgenes are stably integrated into the worm genome and the lines expressing the tagged protein of interest under its native regulatory controls are used for tag based functional analysis.

**B**. Multipurpose tagging cassette. The cassette used for the construction of the TransgeneOme resource consists of 2 copies of the flexible linker peptide TY1, GFP, an

FRT flanked selection (neo) and counter selection (rpsl) cassette and the affinity tag 3xFlag. In the final construct the selection marker is removed by Flp/FRt recombination.
**C**. Size distribution of regions included in the fosmid transgene. The fosmid were selected so that the lengths of the 5' and 3' regions were in 2:1 ratio, as shown in the size distribution histograms on the left. Suitable clones were found for 16403 genes, or about 80% of the genome and that defined the target set for the *C. elegans* TransgeneOme.
**D**. Recombineering pipeline for fosmid transgene production. Diagram of the DNA engineering steps of the recombineering pipeline (described in detail in Supplementary figure 1) is showed on the left with the efficiency at each step indicated in the bar graph on the right. The original host cells containing the fosmids are made recombination proficient by transformation of the multipurpose recombineering plasmid pRedFlp. The tag is then inserted by recombineering just in front of the stop codon as a cassette (tagFrnF) with a FRT flanked selection/counterselection marker; The selection marker is then removed by Flp recombination, to leave only the tagging cassette minimizing any effect on regulation from downstream 3'UTR elements. The single remaining FRT site is positioned in the sequence such that it codes for a short flexible linker between the GFP and the 3XFlag. Finally a genetic selection marker for worm transgenesis (unc-119, cyan) is inserted as a cassette with an E. coli selection marker (nat, magenta) in the vector backbone depicted in black. For the repeat tagging of all failed clones only the cumulative success rate is shown.
**E**. High throughput NGS validation. After pooling the clones the correct transgene engineering was validated by paired end sequencing of the fosmid DNA. Constructs covered by a spanning read on both the 5' (light blue) and the 3' (dark blue) end were considered validated.
**F**. Validation depth. Constructs with at least one spanning pair were considered validated, but over 98% of the junctions were covered by more than one pair (with a average of 29).
**G**. Fraction of constructs with unflipped selection cassette. The constructs are engineered in liquid culture the final clone pools can contain carryover products from the intermediary steps, but they are only present as a minor fraction.
**H,I** Sanger sequencing validation at the single clone level. **H** Tagging validation: the tag region of isolated single clones was sequenced by Sanger sequencing. Color code: green – exact match; yellow – clones with mismatches (either sequencing noise or true mutations); blue – unflipped clones. 83% of the clones exactly matched the expected sequence. Sequencing of a second clone for the failed constructs resulted in exact match in 59% of the cases. **I** The *unc-119* marker insertion was validated for the same clones by sequencing the insertion junction. Color code: green-exact match; yellow – mixed sequence indicating carryover of the unmodified construct; red – no match. (See also Supplemental figure 1, Supplemental table 1 and Movie S1)

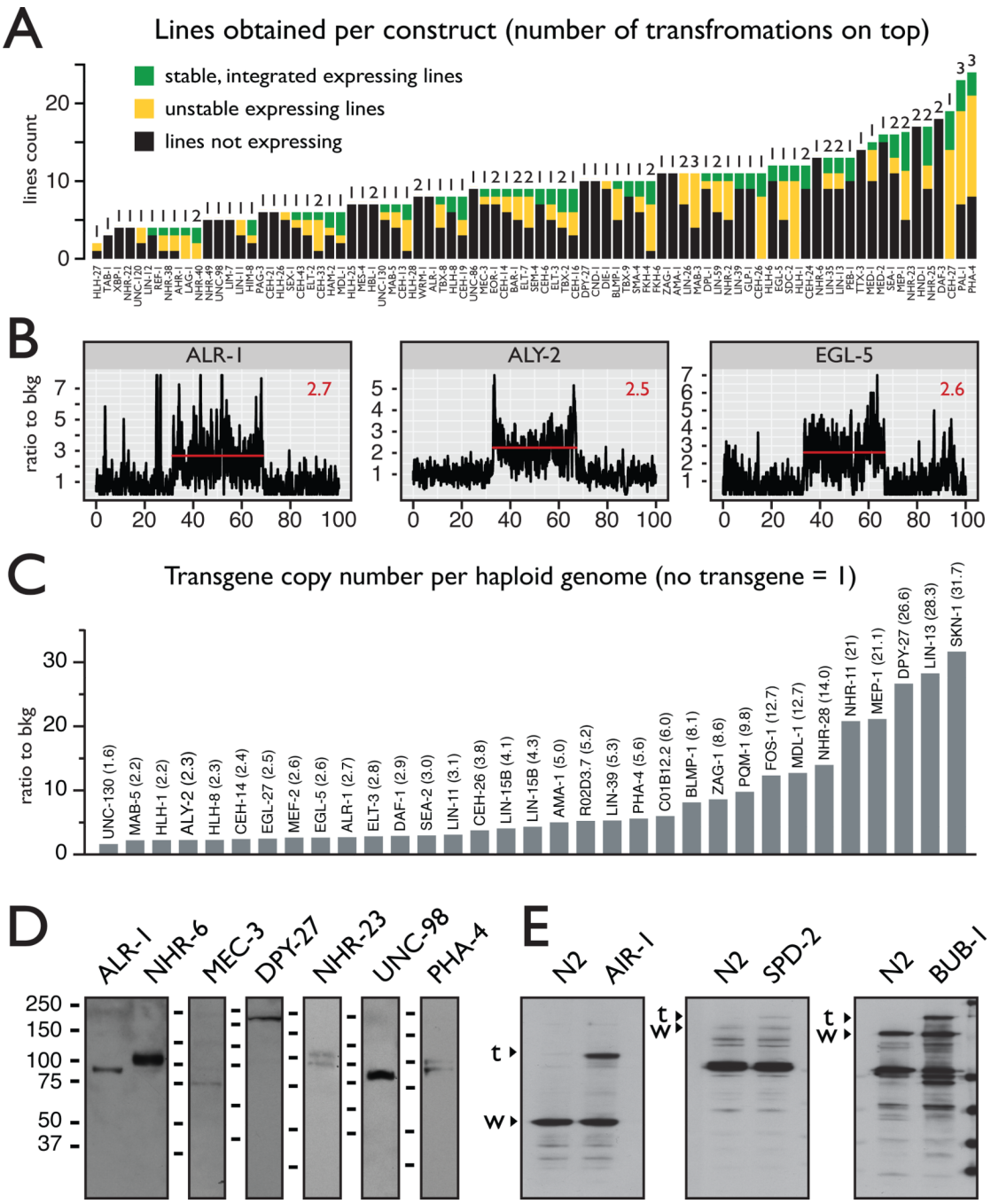


**Figure 2.**

Low copy integration of full-length fosmid transgenes into the genome results in correct protein expression regulation.

**A**. Transgenic line generation success rates The success rate of transformation by micro particle bombardment is shown as total number of lines generated for each construct. Black – fluorescence not observed; yellow – unstable fluorescent lines; green – stable fluorescent line. The number of transformation experiments is indicated on top of each column. In most cases a single transformation was sufficient to obtain at least one stable line.

**B.,C**. The full extent of the gDNA transgenes can be integrated into the genome. Deep sequencing of gDNA from the transgenic strains shows elevated signals in the regions covered by the fosmids (indicated with a red line in panel **B**). The average level for the transgene region, normalized to the neighboring genomic region is shown in panel **C**. Most of the transgenes are integrated at a low copy number (See also Supplemental figure 2).
**D**. The gDNA transgenes deliver tagged protein expression at the expected molecular weights. Tag based protein detection by α-FLAG western blot shows that the tagged proteins are expressed at the expected molecular weights, including specific protein isoforms. The expected sizes for all proteins including the 33 kDA tag are as follows: alr- 1, OP200 - 75 kD; nhr-6, OP90 - 483, 89, 104 kD; mec-3, OP55 - 65, 70 kD; dpy-27, OP32 - 202kD; nhr-23, OP43 - 75, 83, 97, 100kD; unc-98, OP85 - 69kD; pha-4, OP37 - 80, 84, 91kD.
**E**. The tagged proteins are expressed at endogenous levels Western blot with protein specific antibodies shows that the tagged (labeled with “t”) and the endogenous (labeled with “w”) isoforms are expressed at comparable levels, indicating that the tagged proteins are expressed at physiological levels.

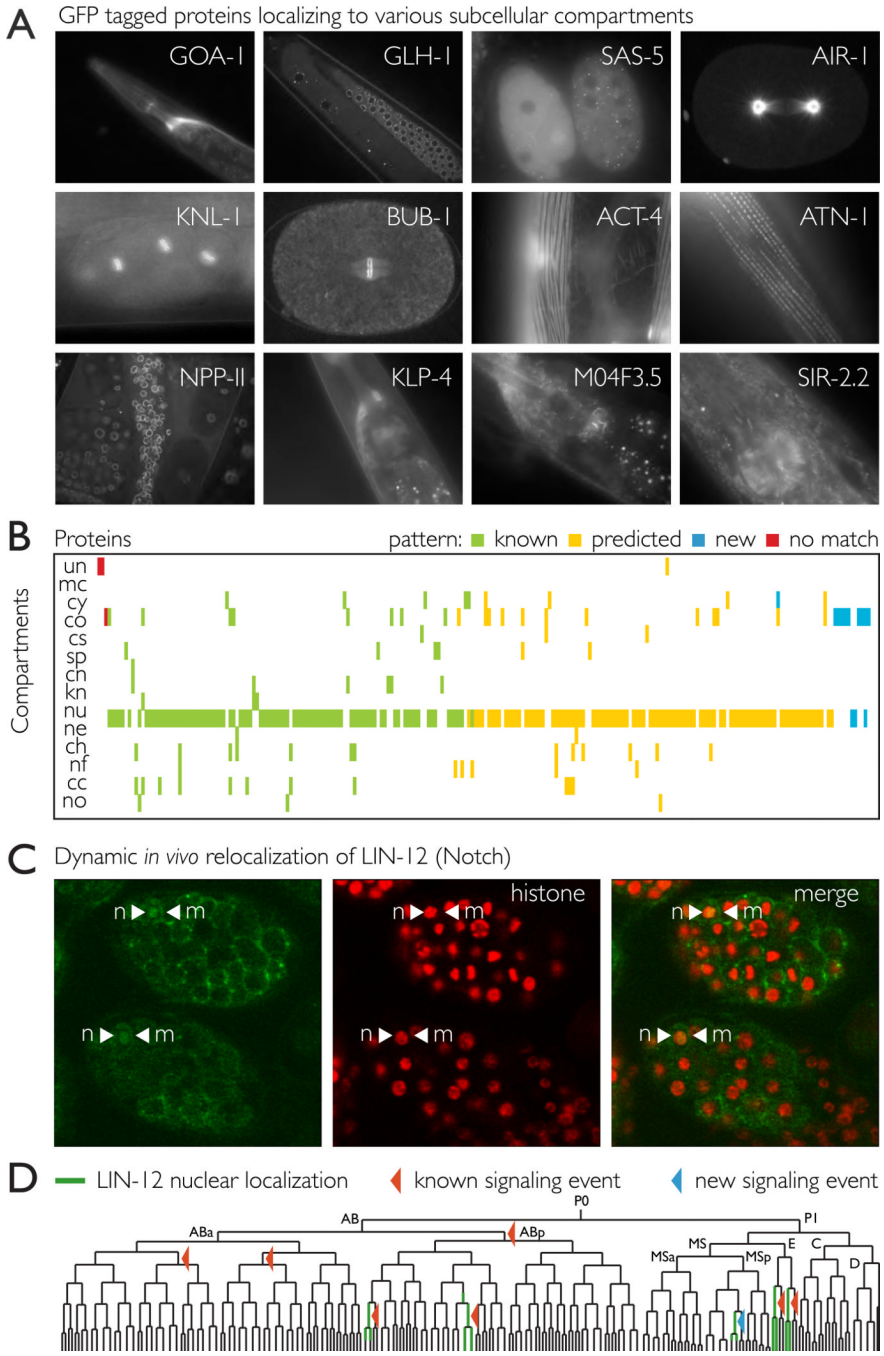


**Figure 3.**

The tagged proteins correctly localize to various subcellular compartments in accordance with their functional role.

**A**. Examples of specific subcellular localizations that reflect the molecular function of the protein *in vivo*. GOA-1 (Go/Gi class G protein alpha subunit) localizes to membranes. GLH-1 localizes to droplet like particles in the germ line cytoplasm (P granules); SAS-5 and AIR-1 (Aurora A kinase) localize to centrosomes; AIR-1 is also found on the spindle microtubules; KNL-1 and BUB-1 are kinetochores components and localize to the condensed chromosomes in mitosis; ACT-4 (actin) and ATN-1 (actinin), are part of the

muscle fiber structures;NPP-11 is nuclear pore complex protein; KLP-4 is a kinesin like protein which based on homology is expected to play a role in neuronal function and can be seen on axon microtubules; M04F3.5 is likely involved in formation of plasma membrane protrusions and localizes to specific membrane sub domains; SIR- 2.2 is a homolog of the mammalian mitochondrial sirtuins and localizes to mitochondria (Supplemental figure 3).
**B**. The observed subcellular localization patterns closely match the distributions described in the literature. 230 proteins (x axis) were localized to various compartments (y axis) Color code: green - matching localization patterns previously described in *C. elegans*; yellow - predicted pattern based on functional annotations or homology; blue – new patterns for *C. elegans* for which the localization could not be predicted (there is either no clear ortholog or it's localization is unknown); red – patterns that differ from the previously described localizations. Abbreviations: un – unknown, mc – membrane/cortex, cy – cytoplasm, co – cytoplasmic organelles, cs – cytoskeleton, sp – spindle, cn – centrosome, kn – kinetochore, nu – nucleus, ne – nuclear envelope, ch – chromatin, nf – nuclear foci, cc – condensed chromosomes, no – nucleolus. The full data is available in Supplemental table 2.
**C.,D**. Dynamic *in vivo* relocalization of LIN-12 Notch in response to signaling. The Notch homolog LIN-12 localizes to membranes in most of the cells in the developing embryo (indicated by arrowheads), but translocation to the nucleus is detectable in several cells. The translocation events marked on the lineage with red arrowheads correspond to know Notch signaling events. The blue arrowheads mark lineages where active Notch signaling has not been previously reported. The observed signals are significantly above the background signal in the sister cells without Notch signaling (Supplemental table 3)

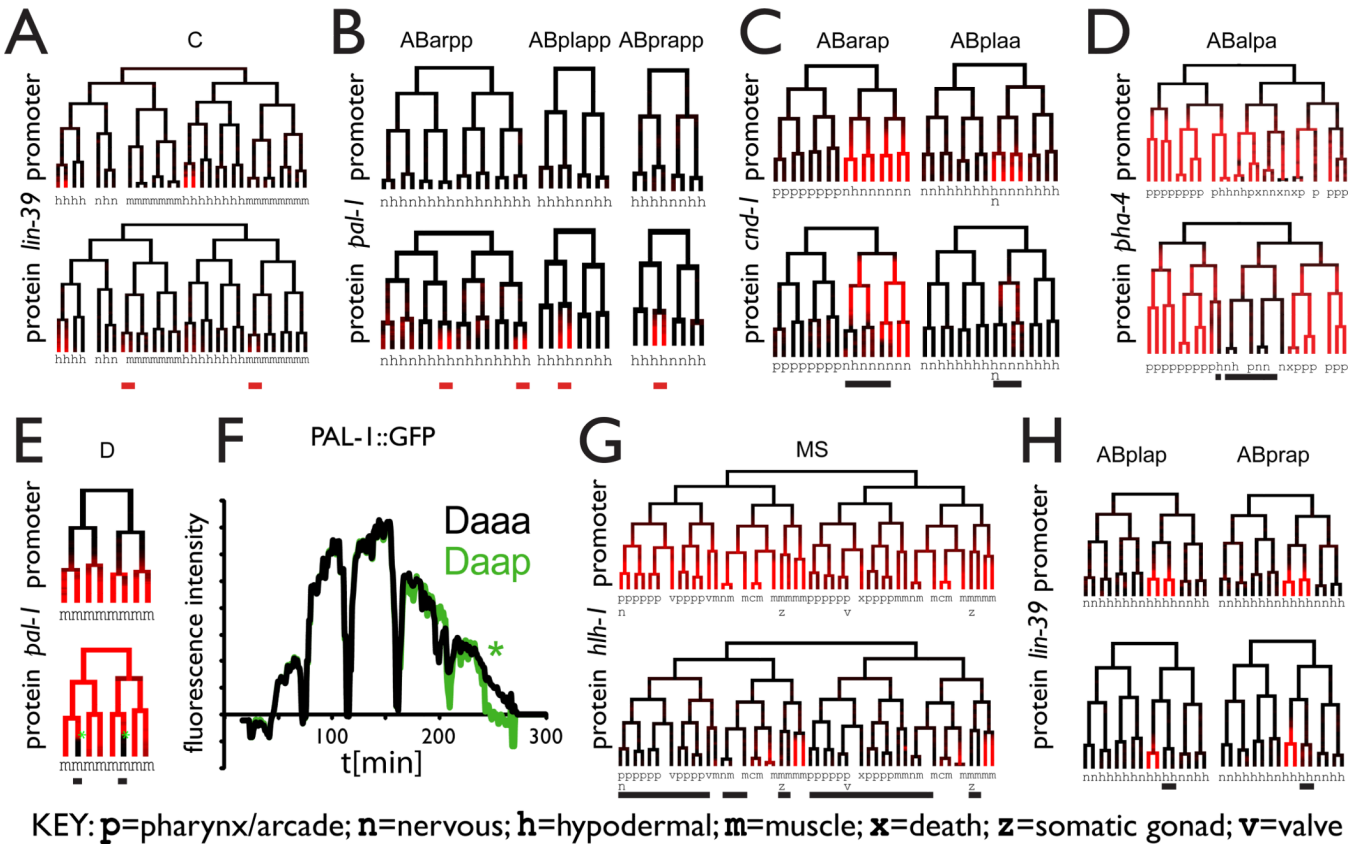


**Figure 4.**

The TransgeneOme derived protein localization patterns reveal additional levels of regulation when compared to promoter:GFP reporters

**A,B** Some genes express the protein fusion in lineages not expressing the promoter reporter (red bars), implicating transcriptional control elements outside of the 5'UTR region present in the promoter reporter.

**C-E** Other proteins showed expression in the same lineages as the promoter reporter but were subsequently lost in specific sub-lineages (black bars)

**E,F**. Rapid loss of PAL-1 protein occurs in two cells of the D lineage in less than 10 minutes, despite the fact that all cells give rise to a muscle fate.

**G,H**. Some genes show initial protein accumulation in only a subset of the promoter reporter expressing cells (black bars). See also Supplemental table 3.

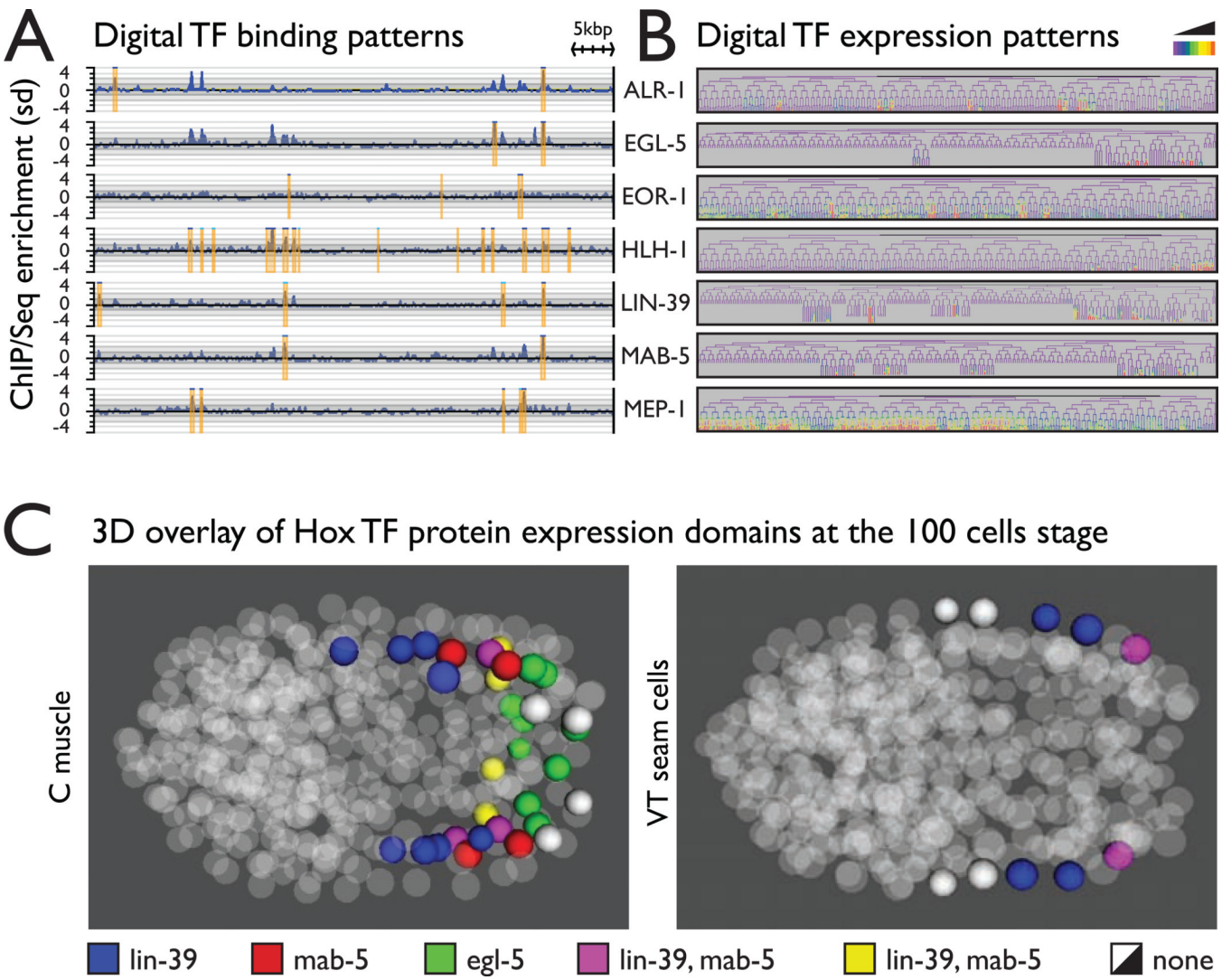


**Figure 5.**

Systematic high resolution mapping of TF protein localization at at molecular level by Chromatin IP and at cellular level by in vivo microscopy

**A**. Examples of TF binding maps (the full data is available at wormbase.org and modENCODe.org)

**B**. Examples of spatiotemporal mapping of TF expression on cellular/tissue level during embryonic development. The full spatiotemporal data of embryonic TF expression is available at http://epic.gs.washington.edu/.

**C**. Combinatorial expression map of Hox protein expression in the C muscle and the VT seam cells lineages at the 100 cells stage. The patterns follow an anteriorposterior gradient as expected. However the gradients in the two lineages are spatially shifted, suggesting that the formation of the gradients is driven primarily by lineage mechanisms rather then by position. The lineage maps and examples from other lineages are shown in Supplementary figure 4.

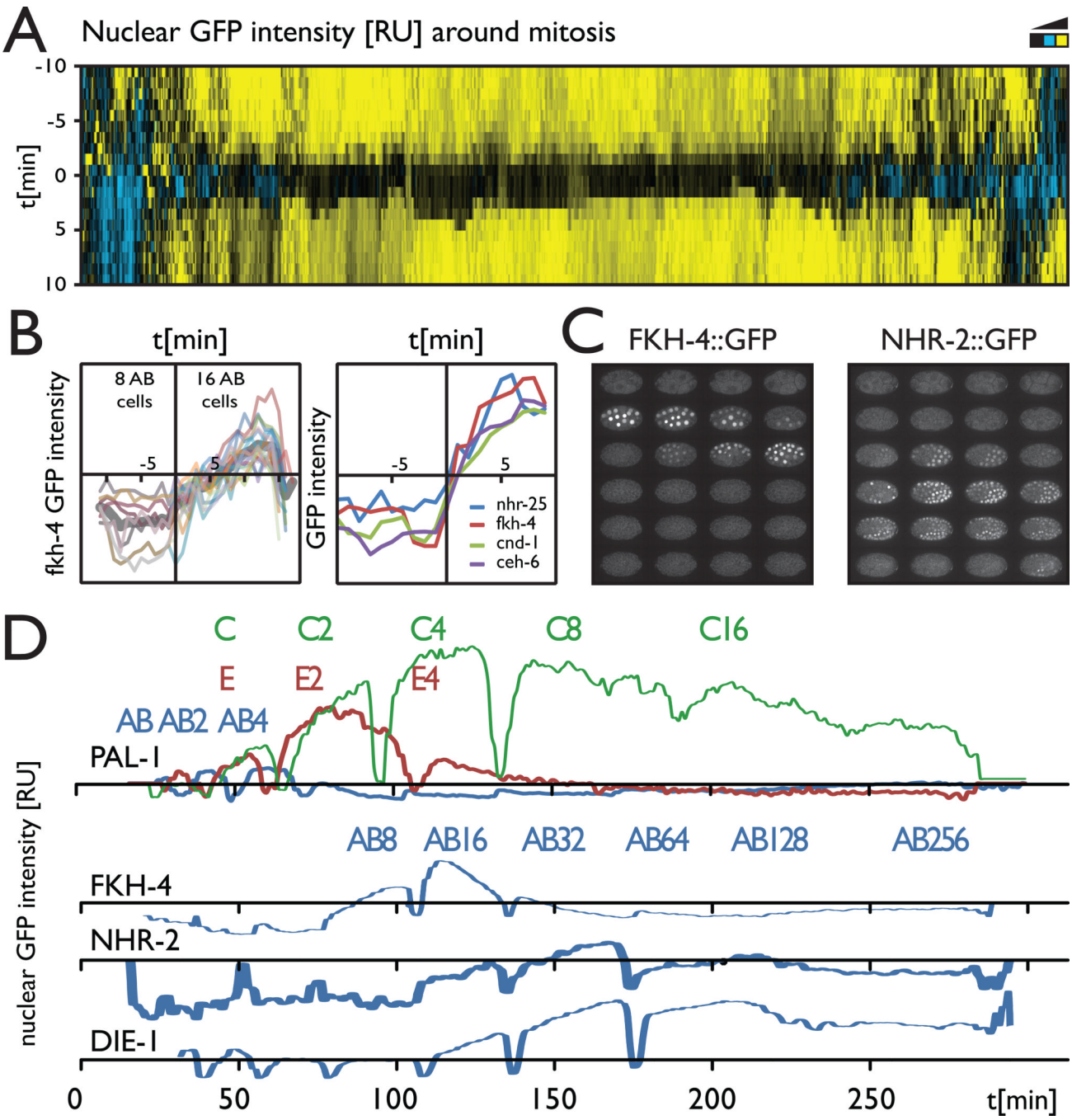


**Figure 6.**

Synchronized and rapid TF protein expression dynamics in multiple cells suggest a global circuit controlling developmental timing

**A**. A global map of TF protein nuclear localization around mitosis. The amount of TF protein in the nucleus for all cell division events in the TF expression dataset is plotted as a heat map centered on the chromosome segregation time point. Most TF proteins rapidly diffuse at nuclear envelope breakdown, indicating they are weakly associated with chromatin.

**B**. Synchronization of the expression onset with the cell cycle for FKH-4. The localization dynamics for all expressing cells are very similar (thin lines individual cells, thick line – average pattern).
**C**. Other proteins for which the expression onset appears synchronized with the cell cycle.
**D**. Expression patterns of FKH-4 and NHR-2 in the early embryo (see also sup videos).
**E**. Average expression of PAL::GFP over time for the founder lineages AB, E and C, showing graded differences in time and level of maximal expression going from posterior to anterior (C->E->AB); and average expression of AB lineage cells for FKH-4, NHR-2, and DIE-1, showing a potentially sequential global temporal regulation. (See supplemental figure 5 and supplemental movies 2 and 3 for more information)