Caenorhabditis elegans AF4/FMR2 family homolog *affl-2* is required for heat shock induced gene expression

Sophie J. Walton[*], Han Wang[*], Porfirio Quintero-Cadena[*], Alex Bateman[†], and Paul W. Sternberg[*,1]

[*] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125.

[†] European Molecular Biology Laboratory, European Bioinformatics Institute, (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

[1] Correspondence should be addressed to:

Paul W. Sternberg

Email: pws@caltech.edu

Phone: 626-395-2181

Address: Mail Code 156-29, 1200 E. California Blvd., California Institute of Technology, Pasadena, CA 91125

Keywords: *affl-1*, *affl-2*, heat shock response, *C. elegans*, AF4/FMR2

Sophie J. Walton ORCID: 0000-0003-1320-1525

Han Wang ORCID: 0000-0002-1933-5762

Porfirio Quintero-Cadena ORCID: 0000-0003-0067-5844

Alex Bateman ORCID: 0000-0002-6982-4660

Paul W. Sternberg ORCID: 0000-0002-7699-0173

Running title: *affl-2* is essential for heat shock response

Abstract

To mitigate the deleterious effects of temperature increases on cellular organization and proteotoxicity, organisms have developed mechanisms to respond to heat stress. In eukaryotes, HSF1 is the master regulator of the heat shock transcriptional response, but the heat shock response pathway is not yet fully understood. From a forward genetic screen for suppressors of heat shock induced gene expression in *C. elegans*, we identified a new allele of *hsf-1* that alters its DNA-binding domain, and three additional alleles of *sup-45*, a previously uncharacterized genetic locus. We identified *sup-45* as one of the two hitherto unknown *C. elegans* orthologs of the human AF4/FMR2 family proteins, which are involved in regulation of transcriptional elongation rate. We thus renamed *sup-45* as *affl-2* (AF4/FMR2-Like). *affl-2* mutants are egg-laying defective and dumpy, but worms lacking its sole paralog (*affl-1*) appear wild-type. AFFL-2 is a broadly expressed nuclear protein, and nuclear localization of AFFL-2 is necessary for its role in heat shock response. *affl-2* and its paralog are not essential for proper HSF-1 expression and localization after heat shock, which suggests that *affl-2* may function downstream or parallel of *hsf-1*. Our characterization of *affl-2* provides insights into the complex processes of transcriptional elongation and regulating heat shock induced gene expression to protect against heat stress.

Introduction

Heat is a universal source of stress in nature, which has detrimental effects including disrupting cellular organization and upsetting proteostasis (Morimoto 1998). One way organisms restore homeostasis after heat stress is through rapid transcriptional changes to upregulate genes that assist with combating damaging effects of heat (Morimoto 1998; Hajdu-Cronin *et al.* 2004; Richter *et al.* 2010). In eukaryotes, heat shock induced transcription is initiated when transcription factors known as heat shock factors (HSFs) are activated and bind to heat shock elements (HSEs) in promoters (Morimoto 1998), and HSF1 has been identified as the primary regulator of heat shock induced transcription in eukaryotes (Åkerfelt *et al.* 2010; Richter *et al.* 2010). The current model for transcriptional control of heat shock response by HSF1 is as follows: under normal conditions chaperones sequester HSF1 and upon heat stress the chaperones disassociate with HSF1 so HSF1 is free to upregulate other genes (Richter *et al.* 2010; Voellmy and Boellmann 2007). However, the complete regulatory system that is responsible for the precise transcriptional control of heat shock response is not yet fully understood (Richter *et al.* 2010).

Although regulation of initiation in heat shock induced transcription has been well characterized, it is known that transcription is also tightly regulated at the elongation and termination steps (Kuras and Struhl 1999; Sims *et al.* 2004; Saunders *et al.* 2006; Lenasi and Barboric 2010). For some genes involved in development and stress, RNA Polymerase II becomes paused at the promoter-proximal region, and escape from the paused state becomes a rate limiting process in transcription (Levine 2011; Lin *et al.* 2011; Zhou *et al.* 2012; Luo *et al.* 2012a) In metazoans, Positive Transcription Elongation Factor beta (P-TEFb), a heterodimeric

kinase complex composed of CDK-9 and a Cyclin T1 Partner, causes RNA Polymerase II and associated factors to become phosphorylated to stimulate promoter escape (Levine 2011; Zhou *et al.* 2012; Luo *et al.* 2012b). When performing this function, P-TEFb participates in the super elongation complex (SEC): a multi-subunit complex composed of a P-TEFb, an AF4/FMR family protein (AFF1 or AFF4), a Pol II elongation factor (ELL, ELL2 or ELL3), a partner protein (EAF1 or EAF2), and an ENL family protein (ENL or AF9) (Lin *et al.* 2010; Luo *et al.* 2012b). Although both AFF4 and AFF1 serve as scaffolds in the SEC, they have been found to work in different processes; e.g., AFF4 has been found to be more important for HSP70 transcription and AFF1 has been linked to promoting HIV transcription (Lu *et al.* 2015).

Caenorhabitis elegans has been used as a multicellular *in vivo* model to perform genetic analysis of homologs of components of the SEC. The *C. elegans* homolog of the ELL gene family (*ell-1*) and its partner (*eaf-1)* are necessary for embryonic development and heat shock induced transcription (Cai *et al.* 2011). In addition, *cdk-9*, *cit-1.1*, and *cit-1.2*, homologs of CDK9 and cyclin T1, are necessary for phosphorylation of Ser2 in RNA polymerase II in the soma (Shim *et al.* 2002; Bowman *et al.* 2013). However, a *C. elegans* version of the AF4/FMR2 family proteins, which serve as the scaffolds for the SEC, has not yet been identified.

In a genetic screen for suppressors of heat shock induced gene expression, we identified a new reduction of function *hsf-1* allele and cloned a *C. elegans* AF4/FMR2 homolog, *affl-2* (AF4/FMR2-Like). *affl-2* encodes a previously uncharacterized protein that is predicted to have two nuclear localization signals and a collection of disordered residues at its N terminus. Indeed, AFFL-2 is a broadly expressed nuclear protein, and its nuclear localization is necessary for its role in heat shock response. In addition to being defective in heat shock response, *affl-2* mutants

87 are dumpy, egg-laying defective, and some have protruding intestines from their vulvas. *affl-2* is
88 not necessary for the proper localization and expression of HSF-1 pre or post heat shock,
89 suggesting that *affl-2* may function downstream of *hsf-1*. Our identification and characterization
90 of *affl-2* furthers our understanding of heat shock response induced transcriptional regulation in
91 *C. elegans* and validates the power of *C. elegans* for genetic analysis of general transcriptional
92 control.
93

Materials and Methods

C. elegans cultivation

C. elegans were grown using the methods described in (Brenner 1974). Strains were maintained on NGM agar plates at room temperature (20 °C) and fed OP50, a slow-growing strain of *Escherichia coli*. A list of strains used in this study can be found in Supplementary Experimental Procedures.

Genomic Editing

We made *affl-1* mutants by inserting the STOP-IN cassette in the 5' end of the coding sequence of *affl-1* using CRISPR/Cas9 with a co-conversion marker (Wang *et al.* 2018). We injected N2 worms to create *affl-1 (sy1202)* single mutants, and we injected *affl-2(sy975)* worms to create *affl-2(sy975) affl-1(sy1220)* double mutants. *sy1202* and *sy1220* had the same molecular change in the *affl-1* gene. The sequence change is a 43bp insertion with 3-frame stop codon and is near the 5' end of the gene *Y55B1BR.1/affl-1:*

5' flanking seq: CCGTACCCGTAGAATGCTTGAAGAAATGGCCGGCC
Insertion: GGGAAGTTTGTCCAGAGCAGAGGTGACTAAGTGATAA
3' flanking seq: TCGTGGGAACTAAACCATTGAGCCAGCTTCCTCGAAG

Primers for genotyping and sequencing can be found in the Supplementary Experimental Procedures.

Generation of Transgenic Lines

Methods to generate transgenic animals were adapted from Mello and Fire (1995). The *affl-2* driver strain was constructed by injecting 25 ng/µl of pSJW003 along with 40 ng/µl of P*unc-122::rfp* and 35 ng/µl of 1 kb DNA ladder (NEB) into the GFP effector strain *syIs300*, and then outcrossing the resulting worms to add the GFP::H2B effector strain *syIs407* in place of *syIs300* (Wang *et al.* 2017). All *affl-2* rescue variants were constructed by injecting 10 ng/µL of

the plasmid containing the rescue construct along with 80 ng/µl of pBluescript and 10 ng/µL of the co-injection marker plasmid KP1368(P*myo-2::nls::mCherry*) into the strain PS8082 (*syIs231 II; affl-2 (sy975) III*). A list of transgenic lines and plasmids used can be found in the Supplementary Experimental Procedures.

lin-3c overexpression assays

We used pumping quiescence as a readout for expression of heat shock driven *lin-3c*, for cessation of pumping is characteristic of *lin-3c* overexpression induced quiescence. We adapted our *lin-3c* overexpression assay from Van Buskirk and Sternberg (2007). 15-30 L4 animals were picked onto NGM agar plates that were seeded 48-72 hours prior. 16-20 hours later, plates with adult animals were parafilmed and placed in a 33 °C water bath for 15 minutes. We used a 15-minute heat shock rather than a 30-minute heat shock because we wanted to be able to detect weaker suppressors. Plates were then left in 20 °C with their lids open to recover for 3 hours before scoring for pumping quiescence. By this time, all worms would have recovered from the mild heat shock and would thus exhibit only *lin-3c* overexpression dependent quiescence. Pumping quiescence was scored using a stereomicroscope on 25-50x magnification, and quiescence was defined as a cessation of movement of the pharyngeal bulb.

Isolation of suppressors

EMS Mutagenesis

Mutagenesis was performed on about 500 late L4 hermaphrodites (PS7244) as described by Brenner (1974). In particular, worms were incubated in a solution of 4 mL M9 with 20 µL EMS (Sigma) for 4 hours. At the end of the 4 hours, we washed the worms three times each with 3 mL of M9 to dilute the EMS. We then plated the mutagenized worms on a seeded NGM agar

plate outside the OP50 lawn and left them to recover for at least 30 minutes before plating the P0 worms.

To screen a synchronized F2 population, we treated the F1 adults with alkaline hypochlorite (bleach) treatment to isolate the eggs of the F2 generation (Protocol B from Porta-de-la-Riva *et al.* 2012). After bleaching treatment, we immediately plated the F2 generation eggs. These steps ensured that all of our F2 animals reached adulthood at roughly the same time.

We performed the *lin-3c* overexpression assay as described above on adult F2 worms. We isolated worms who did not exhibit pumping quiescence onto separate plates, and we screened their progeny (F3 generation) to ensure that the phenotype was stable. Mutants isolated from different P_0 plates were deemed independent.

Complementation Testing with *hsf-1* and *affl-2* Mutants

To identify *hsf-1* and *affl-2* mutants, we performed complementation testing with *hsf-1(sy441), affl-2(sy509),* and *affl-2(sy978)* mutants. Note that *affl-2* was originally named *sup-45.* *hsf-1(sy441), affl-2(sy509),* or *affl-2(sy978)* hermaphrodites were crossed with *syIs197* males. We crossed the resultant male cross progeny into suppressors, and we performed the *lin*-3c overexpression assay on F1 cross progeny of the cross to assay complementation.

SNP Mapping

We used the polymorphic Hawaiian strain CB4856 to perform SNP mapping of our suppressor loci (Doitsidou *et al.* 2010). Our SNP mapping strain was PS7421, which we created by outcrossing PS7244 ten times to CB4856. We followed the Hobert's Lab protocol for worm genomic DNA for SNP Mapping (pooled samples) (Oliver Hobert Lab) to prepare genomic DNA from the progeny of 50 suppressors and 50 non suppressors crossed with PS7421. For identifying *Y55B1BR.2* we sequenced PS75971 (*syIs231 II; affl-2(sy978) III*) crossed into

PS7421 and we sequenced PS8082 (*syIs231 II; affl-2 (sy975) III*) in a N2 background.We analyzed our sequencing results using MiModD mapping software to identify putative mutations (Maier *et al.* 2014).

Protein Sequence Analysis

Y55B1BR.2 similarity to AF4 was found in the first iteration of a JackHMMER iterative search on the EMBL-EBI webserver (Potter *et al.* 2018). Further rounds of searching revealed homologues in a wide range of eukaryotic species. The search was performed against Reference Proteomes using the phmmer algorithm with the following settings:

HMMER Options: -E 1 --domE 1 --incE 0.01 --incdomE 0.03 --mx BLOSUM62 --pextend 0.4 --popen 0.02 --seqdb uniprotrefprot

We used the MUSCLE alignment tool with the default settings to create multiple sequence alignments (Edgar 2004). To determine the location of the predicted NLSs we used the cNLS mapper with a cut-off score of 0.5 and the option to search for bipartite NLSs with a long linker within terminal 60-amino acid regions (Kosugi *et al.* 2008, 2009a; b). We checked that these predictions were consistent with PSORT and NucPred. We used IUPred2A long disorder (Dosztányi *et al.* 2005; Dosztanyi *et al.* 2018) to predict disordered regions of AFFL-2, AFFL-1, AFF1, and AFF4. We used ANCHOR2 software to predict the presence of the disordered protein binding regions of AFFL-2, AFFL-1, AFF1, and AFF4 (Dosztányi *et al.* 2009; Dosztanyi *et al.* 2018). The alignment of the C-terminal homology domain (CHD) was generated using the MUSCLE alignment tool on a selected set of 16 eukaryotic homologues identified using Jackhmmer (Edgar 2004).

Molecular Biology

We used Roche Taq for genotyping PCR products < 1 kb, and we used NEB Phusion High Fidelity or NEB High Fidelity Q5 for PCR of all cloning inserts and genotyping PCR products that were > 1kb. NEB T4 Ligase was used to construct pSJW003. The NEBuilder HiFi DNA Assembly Master Mix was used for Gibson Assembly of pSJW005, pSJW035, pSJW036, pSJW040, and pSJW041. We used DH5α competent cells for all transformation, and we plated all transformed cells on carbenicillin LB plates. A detailed list of oligos and plasmids can be found in the Supplementary Experimental Procedures.

Microscopy

For images of plate phenotypes, one day old adult worms were filmed on plates seeded within 24 hours after the L4 larvae stage. We used a Wild Makroskop M420 dissecting scope at 32x or 40x magnification.

Images of *affl-2* expression and localization of HSF-1::GFP and AFFL-2::GFP were taken using a Plan Apochromat 10x or 63x/1.4 Oil DIC objective in a Zeiss Imager Z2 microscope with an Axiocam 506 Mono camera using ZEN Blue 2.3 software. Young adults were used for imaging, because their lower fat content made it easier to detect fluorescent proteins. Worms were anesthetized in 3 mM levamisole and mounted on 2 % agarose pads. Z-stacks were taken at 63x and maximum-intensity projections were generated using Fiji/ImageJ software (Schindelin *et al.* 2012). Z-stacks were used for images of HSF-1::GFP localization, and when indicated for AFFL-2::GFP localization.

Our method to image HSF-1::GFP was adapted from Morton and Lamitina (2013). To image HSF-1::GFP after heat shock, worms were mounted on slides and incubated in a preheated PCR machine at 35 °C for five minutes. The lid of the PCR machine was left open, for it did not reach 35 °C. Instead, slides were placed in a foil packet on top of the PCR wells. We found that a

five-minute heat shock was long enough to cause worms to form granules consistently and longer incubation caused the plates to dry out. Worms were imaged immediately after heat shock. Non heat shock controls were left on the benchtop (20 °C) for five minutes prior to imaging. In our image analysis, we selected nuclei manually, and chosen nuclei were segmented automatically. We detected granules in chosen nuclei automatically using the blob detector function from scikit-image (Walt *et al.* 2014). We used the mean intensity of the segmented nuclei and granules after background subtraction to determine HSF-1::GFP nuclear intensity.

To image AFFL-2::GFP after heat shock, worms were heat shocked following the same protocol as we used for the *lin-3c* overexpression assay. We choose to use this protocol rather than the one adapted from (Morton and Lamitina 2013) because we wanted to recreate the conditions of the worms were in during the genetic screen. After the heat shock worms were immediately mounted onto slides and imaged.

Data Analysis

Unless otherwise specified, data analysis was carried out using Python 3.7 with standard scientific libraries (Jones *et al.* 2001) and version 0.041 of the bebi103 package (Bois 2018).

Determining pumping quiesence frequency

We used a Bayesian modeling framework to estimate the frequency of worms pumping after heat shock in the *lin-3c* overexpression assay (Bois 2018)

Our prior distribution for ϕ was:

$$\phi \sim Beta\ (\alpha, \beta)$$

Where, α, β are parameters for the Beta distribution that we specified. Our likelihood for n, the number of worms exhibiting pumping quiesence was

$$n \sim Binomial\ (N, \phi)$$

where N is the number of worms in each experiment, n is the number pumping, and ϕ is the probability of pumping. We used Stan to sample from the posterior distribution and find confience intervals for the parameter estimates (Stan Development Team, 2018).

Quantification of HSF-1 Expression and Nuclear Granule Formation

We used non-parametric bootstrapping to estimate the mean of HSF-1 nuclear intensity before heat shock, HSF-1 granule intensity after heat shock, and estimate of the mean of the number of granules per nuclei formed in each strain. We used the permutation test to compute all reported p-values, for we did not have a good parametric model to explain the data generating processes.

Results

A genetic screen of pumping quiescence suppression yields new alleles involved in heat shock transcription response

To identify genes involved in heat shock response, we conducted a forward genetic screen to search for suppressors of pumping quiescence caused through P*hsp-16.41* (promoter of *hsp-16.41*) driven *lin-3* (Fig 1A). When overexpressed in adult animals, *lin*-3, encoding the *C. elegans* ortholog of the epidermal growth factor (EGF), causes a reversible state of quiescence that is characterized by cessation of feeding, locomotion, defecation, and decreased responsiveness (Van Buskirk and Sternberg 2007). *hsp-16.41*'s expression is induced by heat shock; therefore, placing *lin-3* under the control of P*hsp-16.41* gives us temporal control of quiescence. We screened for animals which are defective in downstream steps of quiescence or do not express the transgene properly by searching for pumping animals after heat shock (Fig 1A). We expected that some suppressors would be defective in heat shock induced transcription

of the *lin-3* transgene. Indeed, we found that one of the strong suppressors, *sy1198*, is recessive and was mapped to the chromosome I (data not shown). The gene *hsf-1* is also located on chromosome I, and it was identified from a similar screen using a transgene with a gain-of-function *goa-1* driven by the promoter of *hsp-16.2*, another heat shock response gene similar to *hsp-16.41* (Hadju-Cronin et. al, 2004). We found that *sy1198* and *hsf-1(sy441)* failed to complement for suppression of quiescence, which indicated that *sy1198* is an allele of *hsf-1*. By Sanger sequencing**,** we found *sy1198* is a T to A mutation in the *hsf-1* gene, which leads to a Leucine to Glutamine substitution (L93Q) in the predicted DNA binding domain of HSF-1 protein (Table 1, Fig 1b; Hadju-Cronin *et al.* 2004). As null mutants of *hsf-1* are lethal (Li *et al.* 2016), *sy1198* is likely to be a weak loss-of-function allele.

AFFL-1 and AFFL-2 are homologs of members of the AF4/FMR2 family

We used Hawaiian SNP mapping and whole genome sequencing to find candidates for causal mutations of the suppression of quiescence. We noticed that a group of suppressors which failed to complement one another had mutations in the uncharacterized gene *Y55B1BR.2*, and their mapping position (Fig S1) was similar to that of *sup-45*. Hadju-Cronin *et al.* identified *sup-45* mutants that are defective in heat shock induced transgene expression, but they were unable to clone the gene. We found that our *Y55B1BR.2* mutants failed to complement *sup-45(sy509)*, which confirmed that the suppressors are alleles of *sup-45*. We determined the molecular changes of our three *sup-45* alleles and two previously described *sup-45* using Sanger sequencing, and indeed all had mutations in *Y55B1BR.2* (Table 1). For the rest of our experiments, we used *sy975*, which contains a nonsense mutation at residue 456 (Fig 1C-D, Table 1).

We found that Y55B1BR.2 is predicted to have two nuclear localization signals (NLS), which suggested that it is a nuclear protein (see Methods). Additionally, Y55B1BR.2 has an adjacent paralog *Y55B1BR.1*, whose encoding gene's start codon is separated from the stop codon of *Y55B1BR.2* by only 489 nucleotides ("Y55B1BR.2 (gene) - WormBase□: Nematode Information Resource). Using JackHMMR (Potter *et al.* 2018), we found that Y55B1BR.2 and Y55B1BR.1 are homologs of AFF4, which is a member of the human AF4/FMR2 family. Based this homology we chose to name *Y55B1BR.1* as *affl-1* (AF4/FMR2 Like) and *Y55B1BR.2* as *affl-2*. The AF4/FMR2 family includes the proteins AFF4 and AFF1, which serve as scaffold proteins for multi-subunit super elongation complexes (SECs) that assist with releasing RNA polymerase from promoter-proximal pausing (He and Zhou 2011; Lu *et al.* 2014; Mück *et al.* 2016. AFF1 and AFF4 both consist of an intrinsically disordered N-terminus that interacts with other members of the SEC and a C-terminal homology domain (CHD) that is conserved among members of the AF4/FMR2 family (Chen and Cramer 2019). AFF4's binding sites to SEC partners have been studied and are diagramed in Fig 2B. Similarly, AFFL-2 is also predicted to have disordered residues, most of which are in its N-terminus (Fig 2A). Disordered proteins sometimes have protein binding domains that are disordered in isolation but become structured upon binding (Dosztányi *et al.* 2009), and AFFL-2 has three such predicted disordered protein binding regions in its N-terminus (Fig 2A).

Multiple Sequence Alignment of *H. sapiens* AFF1and AFF4 along with *C. elegans* AFFL-2 and AFFL-1 reveals that most of the similarity between the four proteins is in the conserved C-terminal Homology Domain (CHD) of the AF4/FMR2 family members (Fig 2C, Fig S2).The CHD of AFF4 has been shown to bind nucleic acids, form homodimers, and form heterodimers with the AFF1-CHD *in vitro* (Chen and Cramer 2019). Interestingly, our sequence

searches identified previously unknown homologues of AF4/FMR2 proteins in *Arabidopsis thaliana, Dictostelium discoideum, Acanthamoeba castellani* and sporadic yeast species. Gopalan and colleagues previously identified a homologue within *Schizosaccharomyces pombe* (Gopalan *et al.* 2018).

affl-2 mutants, but not _affl-1_ mutants, are Egl, Dpy, and deficient in heat shock response

affl-2 mutants have been found to be Egl (EGg Laying defect), Dpy (Dumpy) and deficient in heat shock induced transcription (Hajdu-Cronin *et al.* 2004). We also noticed that some *affl-2* mutants have their intestines protruding from their vulvas (Fig 1E). We did not observe such hernias in either *hsf-1(sy441)* or *hsf-1(sy1198)* worms.

We did not find any *affl-1*—the paralog of *affl-2*— mutants in our screen, so we made a null mutant of *affl-1* to see if it is also necessary for heat shock induced gene expression. *affl-1* mutants appear wild-type and do not have any of the morphological phenotypes characteristic of *affl-2* mutants (Fig 1E). We also made *affl-2(sy975) affl-1(sy1220)* double-mutant animals, which we found are Dpy, Egl, and have herniated intestines. We did not notice any obvious defects in the double mutants that are not present in *affl-2(sy975)* single mutants (Fig 1E).

affl-2 has been identified in two independent forward genetic screens using different transgenes: one driven by *Phsp-16.2* (where it was called *sup-45)* and the other by *Phsp-16.41*. It has already been demonstrated that *hsp-16.2* transcription is eliminated in *hsf-1* and *affl-2* mutants worms using qPCR (Hajdu-Cronin *et al.* 2004). Since *hsp-16.2* and *hsp-16.41* share the same regulatory sequence and are both induced by heat shock (Jones *et al.* 1986), we believe that *hsp-16.41* transcription is also likely eliminated in *hsf-1* and *affl-2* mutants. Thus, we decided to use heat shock inducible pumping quiescence, due to expression of *Phsp-16.41*, driven *lin-3* as a

readout for *hsp-16.41* expression. We estimated ϕ, the probability of a given worm pumping after heat shock, to see whether the *hsp-16.41* promoter is active under heat shock conditions in *affl-2, hsf-1, affl-1,* and *affl-2 affl-1* mutants (Fig 1F). We also ensured that all mutants pump at wild type levels prior to heat shock (Fig 1F). The estimates of ϕ for *hsp-16.41:lin-3c* and *affl-1(sy1202); hsp-16.41:lin-3c* were both 0, which indicates that there are no defects in heat shock induced *hsp-16.41* expression. The estimates of ϕ for wild type, P*hsp-16.41:lin-3c; affl-2(sy975),* P*hsp-16.41:lin-3c; hsf-1(sy441),* P*hsp-16.41:lin-3c; hsf-1(sy1198),* P*hsp-16.41: affl-2(sy975) affl-1(sy1220)* were all close to one, which indicates that these mutants were not expressing the *lin*-3c transgene after heat shock. These results confirmed that *hsf-1* and *affl-2*, but not *affl-1,* are necessary for heat shock induced *hsp-16.41* expression.

AFFL-2 is a broadly expressed nuclear protein

We cloned the first 3 kb of sequence upstream from *affl-2's* start codon as *affl*-2's 5' regulatory region and promoter. We used this sequence to create a cGAL driver (Wang et al. 2017), which we crossed with a GFP::H2B effector to create a transcriptional reporter for *affl-2*. GFP was visible in all tissues in worms of all stages, which indicates that *affl-2* is ubiquitously expressed (Fig 3A). To observe the subcellular localization of AFFL-2, we used our cloned *affl-2* promoter to drive AFFL-2 cDNA::GFP. *affl-2(sy975)* mutants with the transgene appear wild-type and are able to express the *hsp-16.41* promoter driven *lin*-3c transgene (Fig 3E). Therefore, we believe that our fusion protein is functional, and our cloned promoter for *affl-2* reflects its endogenous expression pattern. AFFL-2::GFP is exclusively located in the nucleus prior to heat shock, and we do not see any noticeable difference in AFFL-2::GFP localization or intensity between worms before and after heat shock (Fig 3B).

As shown previously, *affl-1* is not necessary for heat shock induced *hsp-16.41* transcription despite also being a homolog of AFF1 and AFF4 (Fig 1E). Additionally, our Multiple Sequence Alignment suggests that AFFL1 shares little similarity with the first 135 amino acids of AFFL2, and AFFL-1 is not predicted to have any NLS (Fig 2, see Methods). We thus decided to investigate the role of AFFL-2's N-terminus, which contains its predicted NLSs and the majority of its predicted disordered residues (Fig 1D, 2A) by creating alternative versions of AFFL-2 with modified N termini (Fig 3C). First, we created a modified AFFL-2::GFP in which we removed 129 amino acids from the N terminus of AFFL-2. This modification eliminated both predicted NLSs and the majority of the disordered residues. To test the necessity of the disordered residues independently from the NLS, we created a construct in which we substituted the deleted residues with the SV40 NLS. To test the role of the disordered nature of the domain independently of its sequence, we made another construct that included the artificial NLS and the 212 residue fused in sarcoma low complexity (FUS LC) domain. FUS is one of three RNA binding proteins with LC domains that when fused to DNA binding domains cause a variety of cancers (Arvand and Denny 2001; Guipaud *et al.* 2006; Lessnick and Ladanyi 2012). 84 % of the FUS LC domain consists of glycine, serine, glutamine, and tyrosine, and at high concentrations the domain has been shown to polymerize. Kwon *et. al.* (2013) also demonstrated that the FUS LC domain fused to the GAL4 DNA binding domain can induce transcriptional activation. As a control, we added a similar region but with 10 tyrosines mutated to serines, which was shown to prevent the GAL4 and FUS LC fusion transcriptional activation abilities (Kwon *et al.* 2013).

We found that in a wild-type background, all of the altered AFFL-2 proteins containing an NLS are observed in the nucleus (Fig 3D). The modified AFFL-2 lacking the artificial NLS is

primarily located in the nucleus, but we saw that some of the protein is present in the cytoplasm (Fig 3D, Fig S3). To test whether the nuclear localization of this construct is dependent on the presence of wild type AFFL-2, we introduced the modified AFFL-2 with N-terminal deletion in *affl-2(sy975)* animals. The localization of AFFL-2::N-terminal Deletion::GFP was similar in both a wild type and *affl-2(sy975)* background: some animals had it strictly localized to the nucleus while others had it in the cytoplasm as well (Fig S3). We did not expect this modified protein to localize to the nucleus, but this result suggests there is an alternative mechanism for its nuclear import. Introducing the modified AFFL-2 with the deletion, but no other alterations, did not rescue any morphological defects of *affl-2(sy975)* worms (Fig S4). However, the constructs with an artificial NLS rescued the morphological defects of *affl-2(s975)* mutants (Fig S4).

To determine the extent that the constructs rescued the heat shock induced gene expression defects of *affl-2(sy975),* we estimated the probability of worms exhibiting pumping quiescence due to heat shock induced *Phsp-16.41: lin-3c* expression (Fig 3E). The estimate of ϕ (the probability of pumping after heat shock) for *affl-2(sy975)* animals with a wild type copy of AFFL-2::GFP was $0^{0.0}_{0.0}$ (estimated mean with lower and upper subscripts denoting lower and upper bounds for estimated 95% confidence interval), which demonstrates that the full AFFL-2::GFP construct is functional. The estimates for ϕ for the deletion-only construct and the construct with the addition of the modified FUS LC domain were both relatively high ($0.511^{0.526}_{0.477}$, $0.406^{0.416}_{0.365}$), and the estimates for ϕ for constructs with only the artificial NLS and the FUS LC were much lower ($0.157^{0.163}_{0.131}$ and $0.166^{0.173}_{0.148}$, respectively). Even though the modified AFFL-2 with both predicted NLSs removed can be seen in the nucleus, adding back the artificial NLS significantly increased the performance of AFFL-2. This suggests that exclusive restriction of AFFL-2 in the nucleus is important for its function in heat shock response. Adding

back the low complexity FUS LC did not further increase the performance of the modified AFFL-2, but the modified (Tyrosine to Serine) FUS LC hindered the performance of AFFL-2. It is believed that FUS increases transcriptional activity by enhancing recruitment of RNA polymerase II, for mutants that bind RNA polymerase II better increase transcription (Kwon *et al*. 2013), and thus this result suggests that the role of AFFL-2 is not to enhance recruitment of RNA polymerase

affl-1 and affl-2 do not significantly influence HSF-1 localization and expression

Since *hsf-1* is an essential gene, we could not perform traditional epistasis experiments using the null mutants to determine whether, and if so how, *affl-2* and *hsf-1* genetically interact. Instead, we used an *hsf-1* translational reporter to determine if *affl-2* and/or *affl-2* are necessary for proper localization and expression of HSF-1. *C. elegans* HSF-1 is a ubiquitously expressed nuclear protein, and HSF-1 will aggregate to form nuclear stress granules after heat shock (Morton and Lamitina 2013). The HSF-1 foci do align with marks of active transcription and are dependent on the HSF1 DNA binding domain, but the putative sites of the foci are still unknown (Morton and Lamitina 2013). We quantified the formation of granules after heat shock using the *Phsf-1::*HSF-1::GFP transgene from Morton and Lamitina (2013) (Fig 4). We also quantified the intensity of the granules after heat shock and the intensity of HSF1::GFP prior to heat shock for all genotypes (Fig S5). We found that HSF-1 expression prior to heat shock is similar in all genotypes (Fig S5a), which demonstrates that neither *affl-2* nor *affl-1* are critical for regulating HSF1 expression. Although some differences of means of the number of granules (Fig 4b) and intensity (Fig S5b) of HSF-1::GFP between genotypes are statistically significant ($p < 0.05$), there is still too much overlap between the different distributions of HSF-1::GFP intensity for these differences to fully explain the highly non-overlapping differences between the quiescence

phenotype of the different strains. Surprisingly, *affl-1(sy1202)* worms have more granules per nucleus after heat shock compared to wild-type worms ($p = 0.0016$) and the most number of granules compared to the other mutants. *affl-2(sy975)* mutants formed slightly less granules after heat shock ($p = 0.034$) , and *affl-2(sy975) affl-1(sy1220)* mutants formed similar numbers of granules per nucleus as wild type ($p = 0.883$). These results do not explain why *affl-2(sy975)* worms are unable to express *hsp-16.41,* for some wild type worms had similar HSF1 granule numbers per nuclei and similar HSF-1::GFP levels. Since HSF-1 localization and expression is not significantly disrupted in *affl-2* mutants, we believe that AFFL-2 acts either downstream or parallel to HSF-1 to regulate heat shock induced transcription in *C. elegans*. This suggests a predicted role of *affl-2* in elongation, based on its homology with AF4/FMR2 family members.

Discussion

We have cloned and performed a genetic analysis of *affl-2*, a homolog of AF4/FMR2 family members and showed that *affl-2* is necessary for heat shock induced transcription. Through a forward genetic screen for suppressors of heat shock induced *lin-3* overexpression, we identified one new *hsf-1* and three *affl-2* alleles. To our knowledge, this is the first isolated viable *hsf-1* allele with an altered DNA binding domain. We found that *affl*-2 mutants are Dpy, Egl, and have herniated intestines, whereas animals lacking a functional *affl-1*—a homolog of AF4/FMR2 family members and the paralog of *affl-2*— appear wild type. We determined that *affl-2* is a ubiquitously expressed nuclear protein, and proper localization is necessary for its role in heat shock induced transcriptional response.

affl-2 mutants had been identified in another screen for suppressors of heat shock induced gene expression, and that screen also identified *hsf-1* and *cyl-1* as regulators of heat shock response (Hajdu-Cronin *et al.* 2004). As said above, we found an *hsf-1* mutant, but we did not recover any *cyl-1* mutants. However, Hajdu-Cronin *et al.* (2004) reported that their *cyl-1* mutants did not suppress the effects of a P*hsp-16.41* driven transgene, which indicate that *cyl-1* is not responsible for *hsp-16.41* transcription. Hadju-Cronin's efforts and ours illustrate the power of simple genetic screens for gene expression in *C. elegans* to find genes responsible for regulation of transcription.

Along with conservation in the CHD, we see a partial conservation of AFF4's AF9/ENL, ELL-1/2, and P-TEFb binding sites in the AFFL-2 N terminal sequence (Fig 2C). AFFL-2 is predicted computationally to have three candidates for binding sites, but we have not yet verified whether these are real. It is possible that these three sites are sufficient for AFFL-2 to bind to its partners in the SEC, and it is possible that AFFL-2 does not interact with all components of the

448 SEC that human AFF4/AFF1 have been found to bind. Our deletion removed the region of
449 AFFL-2 similar to AFF4's binding site to P-TEFb, and thus we expected it to be necessary for
450 the AFFL-2's function. Surprisingly, we found that replacing much of the disordered N-terminus
451 of AFFL-2 with an exogenous NLS restores protein function to about 80% of the wild-type
452 control, even though the modified AFFL-2 with its predicted NLSs at the N-terminus removed
453 still partially localizes to the nucleus. It is possible that the C-terminus of AFFL-2 may contain a
454 weak NLS that cannot be predicted by current software which allows AFFL-2 to partially
455 localize to the nucleus at low levels. Addition of an exogenous NLS could be necessary to
456 increase the concentration of nuclear AFFL-2 to improve its functioning, but does not restore
457 AFFL-2 activity to wild-type levels. Our deleted residues removed only one of the candidate
458 binding sites of AFFL-2, and it is possible that the other binding sites and disordered residues
459 can act redundantly to maintain AFFL-2 activity. However, a more thorough biochemical
460 investigation of AFFL-2 is needed to determine the role of different domains of the protein.
461 Despite AFFL-1 being an ortholog of AFF4/AFF1 as well, AFFL-1 is not necessary for
462 heat shock induced transcription and *affl-1* mutants appear wild type. AFFL-1 is not predicted to
463 have any Nuclear Localization Signals, which suggests that it may not even be a nuclear protein.
464 Since we do not have a phenotype for *affl-1* mutants we have no way to validate any expression
465 pattern or localization obtained using a fusion protein, for we cannot validate that the fusion
466 protein is functional. It is not clear what AFFL-1's role is, and if AFFL-1 has a role in
467 transcription or not. We have not fully investigated *affl-1* mutants to see if they have any
468 deficiencies in other processes besides heat shock induced transcription, and *affl-1* could play a
469 redundant role with another gene.

We used translational reporters to examine the roles of *affl-1* and *affl-2* on HSF-1 subcellular localization and expression. While we did find some differences in HSF-1 expression prior to heat shock and HSF-1 granule formation after heat shock, we are not confident that these differences can explain the phenotypes of the various mutants because the distributions of our measurements for different genotypes overlap. This aligns with our hypothesis that *affl-2* is necessary for proper elongation in transcription, not initiation, for it suggests that AFFL-2 acts downstream of granule formation. Furthermore, addition of the FUS LC domain does not significantly increase the performance of the modified AFFL-2, which suggests that AFFL-2 is not involved in recruiting RNA polymerase but is acts in a downstream step. However, there are no putative binding sites for the granules and it is unclear what their role in heat shock response is (Morton and Lamitina 2013).

As mentioned previously, AFFL-1 and AFFL-2 are homologs of mammalian AFF1 and AFF4, which serve as scaffolds for the super elongation complex (SEC). AFF1 and AFF4 serve as scaffolds in the super elongation complex (SEC), which regulates release from promoter-proximal pausing during transcriptional elongation using P-TEFb (He and Zhou 2011; Lu *et al.* 2014; Mück *et al.* 2016). AFF4 is responsible for heat shock induced HSP70 expression, which illustrates that its role in heat shock induced gene expression is conserved (Lu *et al.* 2015). In *C. elegans,* the P-TEFb complex has been shown to be necessary for embryonic development and expression of *hsp-16.2* (Schulze-Gahmen *et al.* 2013). Although *affl-2* and *affl-1* mutants survive past embryonic development, *affl-2* mutants are deficient in heat shock induced gene expression. Future work should investigate whether *affl-2* mutants are also deficient in expression of genes involved in embryonic development and whether they have the same defects in Ser2 phosphorylation as animals lacking a function P-TEFb complex (Schulze-Gahmen *et al.* 2013).It

493 is possible that *affl-2* mutants still have Ser2 phosphorylation, but at lower levels than wild type,
494 which could allow them able to survive through development.

495 Our results demonstrate that the *C. elegans* ortholog of AF4/FMR2 family members,
496 AFFL-2, is necessary for heat shock induced transcription. Our sequence analysis suggests that
497 AF4/FMR2 homologues are found more widely in nature than previously thought, highlighting
498 their importance. These results combined with previous work on other members of the SEC
499 suggest that *C. elegans* can be a powerful, multicellular model to understand transcriptional
500 elongation. Further study of *C. elegans* homologs of human AF4/FMR2 proteins will facilitate
501 our understanding of heat shock response as well as transcriptional elongation in general.
502
503

Acknowledgements

We thank Heenam Park for helping create *Y55B1BR.1* (*affl-1*) null mutants using CRISPR/Cas9, and we thank Jean Badroos, Jasmine S. Revanna, and Minyi Tan for technical assistance. We thank Hillel Schwartz, Jonathan Liu, and members of the Sternberg Lab for insightful discussions. We thank Sarah MacLean and the Sternberg Lab for comments on the manuscript. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This work was supported by NIH (K99GM126137 to H.W., U24HG002223 to P.W.S., and R240D023041). The Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Technology performed whole genome sequencing. S.J.W. was supported by the Caltech Student Faculty Programs, the family of Laurence J. Stuppy, Samuel N. Vodopia, and Carol J. Hasson.

Data Availability.

Strains and plasmids are available upon request. Code for data analysis and image processing can be found at: https://github.com/sophiejwalton/affl-2.git. Data will be made publicly available upon publication.

References

Åkerfelt M., R. I. Morimoto, and L. Sistonen, 2010 Heat shock factors: integrators of cell stress, development and lifespan. Nature Reviews Molecular Cell Biology 11: 545–555. https://doi.org/10.1038/nrm2938

Arvand A., and C. T. Denny, 2001 Biology of EWS/ETS fusions in Ewing's family tumors. Oncogene 20: 5747–5754. https://doi.org/10.1038/sj.onc.1204598

Bois J., 2018 *bebi103*.

Bowman E. A., C. R. Bowman, J. H. Ahn, and W. G. Kelly, 2013 Phosphorylation of RNA polymerase II is independent of P-TEFb in the C. elegans germline. Development 140: 3703–3713. https://doi.org/10.1242/dev.095778

Brenner S., 1974 The Genetics of Caenorhabditis Elegans. Genetics 77: 71–94.

Cai L., B. L. Phong, A. L. Fisher, and Z. Wang, 2011 Regulation of Fertility, Survival, and Cuticle Collagen Function by the Caenorhabditis elegans eaf-1 and ell-1 Genes. J. Biol. Chem. 286: 35915–35921. https://doi.org/10.1074/jbc.M111.270454

Chen Y., and P. Cramer, 2019 Structure of the super-elongation complex subunit AFF4 C-terminal homology domain reveals requirements for AFF homo- and heterodimerization. J. Biol. Chem. 294: 10663–10673. https://doi.org/10.1074/jbc.RA119.008577

Chou S., H. Upton, K. Bao, U. Schulze-Gahmen, A. J. Samelson, *et al.*, 2013 HIV-1 Tat recruits transcription elongation factors dispersed along a flexible AFF4 scaffold. PNAS 110: E123–E131. https://doi.org/10.1073/pnas.1216971110

543 Doitsidou M., R. J. Poole, S. Sarin, H. Bigelow, and O. Hobert, 2010 C. elegans Mutant
544 Identification with a One-Step Whole-Genome-Sequencing and SNP Mapping Strategy.
545 PLoS One 5. https://doi.org/10.1371/journal.pone.0015435

546 Dosztányi Z., V. Csizmok, P. Tompa, and I. Simon, 2005 IUPred: web server for the prediction
547 of intrinsically unstructured regions of proteins based on estimated energy content.
548 Bioinformatics 21: 3433–3434. https://doi.org/10.1093/bioinformatics/bti541

549 Dosztányi Z., B. Mészáros, and I. Simon, 2009 ANCHOR: web server for predicting protein
550 binding regions in disordered proteins. Bioinformatics 25: 2745–2746.
551 https://doi.org/10.1093/bioinformatics/btp518

552 Dosztanyi Z., B. Meszaros, and G. Erdos, 2018 *IUPred2a*. MTA-ELTE Momentum
553 Bioinformatics Research Group.

554 Edgar R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high
555 throughput. Nucleic Acids Res 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

556 Gopalan S., D. M. Gibbon, C. A. Banks, Y. Zhang, L. A. Florens, *et al.*, 2018
557 Schizosaccharomyces pombe Pol II transcription elongation factor ELL functions as part
558 of a rudimentary super elongation complex. Nucleic Acids Res 46: 10095–10105.
559 https://doi.org/10.1093/nar/gky713

560 Guipaud O., F. Guillonneau, V. Labas, D. Praseuth, J. Rossier, *et al.*, 2006 An in vitro enzymatic
561 assay coupled to proteomics analysis reveals a new DNA processing activity for Ewing
562 sarcoma and TAF(II)68 proteins. PROTEOMICS 6: 5962–5972.
563 https://doi.org/10.1002/pmic.200600259

564 Hajdu-Cronin Y. M., W. J. Chen, and P. W. Sternberg, 2004 The L-type cyclin CYL-1 and the
565 heat-shock-factor HSF-1 are required for heat-shock-induced protein expression in
566 Caenorhabditis elegans. Genetics 168: 1937–1949.
567 https://doi.org/10.1534/genetics.104.028423

568 He N., and Q. Zhou, 2011 New Insights into the Control of HIV-1 Transcription: When Tat
569 Meets the 7SK snRNP and Super Elongation Complex (SEC). J Neuroimmune
570 Pharmacol 6: 260–268. https://doi.org/10.1007/s11481-011-9267-6

571 Jones D., R. H. Russnak, R. J. Kay, and E. P. Candido, 1986 Structure, expression, and evolution
572 of a heat shock gene locus in Caenorhabditis elegans that is flanked by repetitive
573 elements. J. Biol. Chem. 261: 12006–12015.

574 Jones E., T. Oliphant, P. Peterson, and et. al, 2001 *SciPy: Open Source Scientific Tools for*
575 *Python*.

576 Kosugi S., M. Hasebe, T. Entani, S. Takayama, M. Tomita, *et al.*, 2008 Design of peptide
577 inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling.
578 Chem. Biol. 15: 940–949. https://doi.org/10.1016/j.chembiol.2008.07.019

579 Kosugi S., M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, *et al.*, 2009a Six
580 classes of nuclear localization signals specific to different binding grooves of importin
581 alpha. J. Biol. Chem. 284: 478–485. https://doi.org/10.1074/jbc.M807017200

582 Kosugi S., M. Hasebe, M. Tomita, and H. Yanagawa, 2009b Systematic identification of cell
583 cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite

584 motifs. Proc. Natl. Acad. Sci. U.S.A. 106: 10171–10176.
585 https://doi.org/10.1073/pnas.0900604106

586 Kuras L., and K. Struhl, 1999 Binding of TBP to promoters in vivo is stimulated by activators
587 and requires Pol II holoenzyme. Nature 399: 609–613. https://doi.org/10.1038/21239

588 Kwon I., M. Kato, S. Xiang, L. Wu, P. Theodoropoulos, *et al.*, 2013 Phosphorylation-Regulated
589 Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains. Cell
590 155: 1049–1060. https://doi.org/10.1016/j.cell.2013.10.033

591 Leach B. I., A. Kuntimaddi, C. R. Schmidt, T. Cierpicki, S. A. Johnson, *et al.*, 2013 Leukemia
592 Fusion Target AF9 Is an Intrinsically Disordered Transcriptional Regulator that Recruits
593 Multiple Partners via Coupled Folding and Binding. Structure 21: 176–183.
594 https://doi.org/10.1016/j.str.2012.11.011

595 Lenasi T., and M. Barboric, 2010 P-TEFb stimulates transcription elongation and pre-mRNA
596 splicing through multilateral mechanisms. RNA Biology 7: 145–150.
597 https://doi.org/10.4161/rna.7.2.11057

598 Lessnick S. L., and M. Ladanyi, 2012 Molecular Pathogenesis of Ewing Sarcoma: New
599 Therapeutic and Transcriptional Targets. Annu. Rev. Pathol. Mech. Dis. 7: 145–159.
600 https://doi.org/10.1146/annurev-pathol-011110-130237

601 Levine M., 2011 Paused RNA Polymerase II as a Developmental Checkpoint. Cell 145: 502–
602 511. https://doi.org/10.1016/j.cell.2011.04.021

603 Li J., L. Chauve, G. Phelps, R. M. Brielmann, and R. I. Morimoto, 2016 E2F coregulates an
604 essential HSF developmental program that is distinct from the heat-shock response.
605 Genes Dev. 30: 2062–2075. https://doi.org/10.1101/gad.283317.116

606 Lin C., E. R. Smith, H. Takahashi, K. C. Lai, S. Martin-Brown, *et al.*, 2010 AFF4, a Component
607 of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can
608 Link Transcription Elongation to Leukemia. Molecular Cell 37: 429–437.
609 https://doi.org/10.1016/j.molcel.2010.01.026

610 Lin C., A. S. Garrett, B. D. Kumar, E. R. Smith, M. Gogol, *et al.*, 2011 Dynamic transcriptional
611 events in embryonic stem cells mediated by the super elongation complex (SEC). Genes
612 Dev. 25: 1486–1498. https://doi.org/10.1101/gad.2059211

613 Lu H., Z. Li, Y. Xue, U. Schulze-Gahmen, J. R. Johnson, *et al.*, 2014 AFF1 is a ubiquitous P-
614 TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of
615 SECs for HIV transactivation. PNAS 111: E15–E24.
616 https://doi.org/10.1073/pnas.1318503111

617 Lu H., Z. Li, W. Zhang, U. Schulze-Gahmen, Y. Xue, *et al.*, 2015 Gene target specificity of the
618 Super Elongation Complex (SEC) family: how HIV-1 Tat employs selected SEC
619 members to activate viral transcription. Nucleic Acids Res 43: 5868–5879.
620 https://doi.org/10.1093/nar/gkv541

621 Luo Z., C. Lin, E. Guest, A. S. Garrett, N. Mohaghegh, *et al.*, 2012a The Super Elongation
622 Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and

623 Transcriptional Output. Molecular and Cellular Biology 32: 2608–2617.
624 https://doi.org/10.1128/MCB.00182-12

625 Luo Z., C. Lin, and A. Shilatifard, 2012b The super elongation complex (SEC) family in
626 transcriptional control. Nature Reviews Molecular Cell Biology 13: 543–547.
627 https://doi.org/10.1038/nrm3417

628 Maier W., K. Moos, M. Seifert, and R. Baumeister, 2014 *MiModD - Mutation Identification in*
629 *Model Organism Genomes*. SourceForge.net.

630 Morimoto R. I., 1998 Regulation of the heat shock transcriptional response: cross talk between a
631 family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev.
632 12: 3788–3796. https://doi.org/10.1101/gad.12.24.3788

633 Morton E. A., and T. Lamitina, 2013 C. elegans HSF-1 is an essential nuclear protein that forms
634 stress granule-like structures following heat shock. Aging Cell 12: 112–120.
635 https://doi.org/10.1111/acel.12024

636 Mück F., S. Bracharz, and R. Marschalek, 2016 DDX6 transfers P-TEFb kinase to the
637 AF4/AF4N (AFF1) super elongation complex. Am J Blood Res 6: 28–45.

638 Oliver Hobert Lab, Worm Genomic DNA Prep

639 Potter S. C., A. Luciani, S. R. Eddy, Y. Park, R. Lopez, *et al.*, 2018 HMMER web server: 2018
640 update. Nucleic Acids Res 46: W200–W204. https://doi.org/10.1093/nar/gky448

641 Qi S., Z. Li, U. Schulze-Gahmen, G. Stjepanovic, Q. Zhou, *et al.*, 2017 Structural basis for ELL2
642 and AFF4 activation of HIV-1 proviral transcription. Nature Communications 8: 14076.
643 https://doi.org/10.1038/ncomms14076

644 Richter K., M. Haslbeck, and J. Buchner, 2010 The Heat Shock Response: Life on the Verge of
645 Death. Molecular Cell 40: 253–266. https://doi.org/10.1016/j.molcel.2010.10.006

646 Saunders A., L. J. Core, and J. T. Lis, 2006 Breaking barriers to transcription elongation. Nat
647 Rev Mol Cell Biol 7: 557–567. https://doi.org/10.1038/nrm1981

648 Schindelin J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, *et al.*, 2012 Fiji: an open-
649 source platform for biological-image analysis. Nat Methods 9: 676–682.
650 https://doi.org/10.1038/nmeth.2019

651 Schulze-Gahmen U., H. Upton, A. Birnberg, K. Bao, S. Chou, *et al.*, 2013 The AFF4 scaffold
652 binds human P-TEFb adjacent to HIV Tat. eLife.

653 Schulze-Gahmen U., H. Lu, Q. Zhou, and T. Alber, 2014 AFF4 binding to Tat-P-TEFb indirectly
654 stimulates TAR recognition of super elongation complexes at the HIV promoter. eLife.

655 Schulze-Gahmen U., and J. H. Hurley, 2018 Structural mechanism for HIV-1 TAR loop
656 recognition by Tat and the super elongation complex. PNAS 115: 12973–12978.
657 https://doi.org/10.1073/pnas.1806438115

658 Shim E. Y., A. K. Walker, Y. Shi, and T. K. Blackwell, 2002 CDK-9/cyclin T (P-TEFb) is
659 required in two postinitiation pathways for transcription in the C. elegans embryo. Genes
660 Dev. 16: 2135–2146. https://doi.org/10.1101/gad.999002

661 Sims R. J., R. Belotserkovskaya, and D. Reinberg, 2004 Elongation by RNA polymerase II: the
662 short and long of it. Genes Dev. 18: 2437–2468. https://doi.org/10.1101/gad.1235904

663 Van Buskirk C., and P. W. Sternberg, 2007 Epidermal growth factor signaling induces
664 behavioral quiescence in *Caenorhabditis elegans*. Nature Neuroscience 10: 1300–1307.
665 https://doi.org/10.1038/nn1981

666 Voellmy R., and F. Boellmann, 2007 Chaperone Regulation of the Heat Shock Protein Response,
667 pp. 89–99 in *Molecular Aspects of the Stress Response: Chaperones, Membranes and*
668 *Networks*, Advances in Experimental Medicine and Biology. edited by Csermely P.,
669 Vígh L. Springer New York, New York, NY.

670 Walt S. van der, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, *et al.*, 2014
671 scikit-image: image processing in Python. PeerJ 2: e453.
672 https://doi.org/10.7717/peerj.453

673 Wang H., J. Liu, S. Gharib, C. M. Chai, E. M. Schwarz, *et al.*, 2017 cGAL, a temperature-robust
674 GAL4–UAS system for *Caenorhabditis elegans*. Nature Methods 14: 145–148.
675 https://doi.org/10.1038/nmeth.4109

676 Wang H., H. Park, J. Liu, and P. W. Sternberg, 2018 An Efficient Genome Editing Strategy To
677 Generate Putative Null Mutants in Caenorhabditis elegans Using CRISPR/Cas9. G3:
678 Genes, Genomes, Genetics 8: 3607–3616. https://doi.org/10.1534/g3.118.200662

679 Zhou Q., T. Li, and D. H. Price, 2012 RNA Polymerase II Elongation Control. Annu. Rev.
680 Biochem. 81: 119–143. https://doi.org/10.1146/annurev-biochem-052610-095910

681

682 **Supporting Information**

683 **Supplementary Data:**

689 **Supplementary Experimental Procedures:**

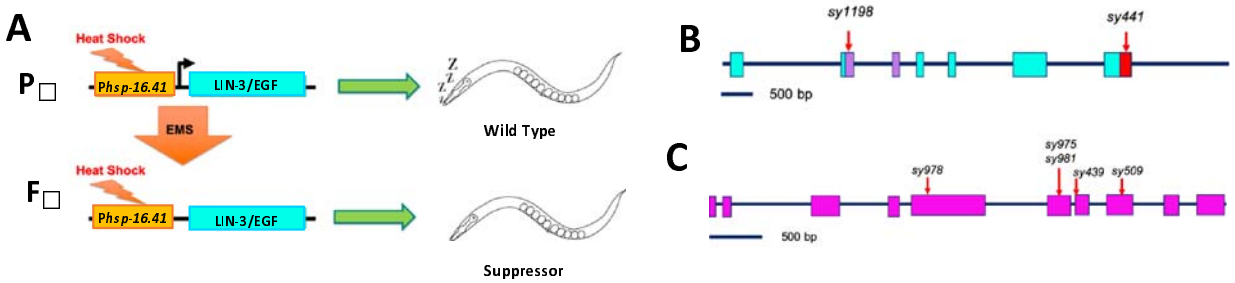
Tables

Table 1: DNA Sequence Changes of Alleles

Allele	Sequence change with 25 base pair flanking sequence around the mutation (bold letter)	Predicted Amino Acid Change
hsf-1(sy1198)	GACGACGACAAGCTTCCAGTATTTC **A**GATAAAATTGTGGAATATCGTAGAA	L93Q
affl-2(sy978)	GAAGCTCATTGAGCTTGCGAAGACC **T**GAGGGAAGCTGACGGATGTTGTGGA	R247*
affl-2(sy981)	GTATTTTTTGTACTTGAAGCCATCT **A**GAAAGCTGACAAGGAGCGGACAACT	W456*
affl-2(sy975)	TATTTTTTGTACTTGAAGCCATCTG **A**AAAGCTGACAAGGAGCGGACAACTC	W456*
affl-2(sy509)	AGTGGCTCACGGGACCCCCAAACCA **A**CC**CAC**GTCCGCCGGATCGACACCATCTGGT	Missense mutations starting at S581 R635* and from to frameshift caused by CC to CAC mutation.
affl-2(sy439)	ttttaaaaattgatgaataatttca **A**GTCGCCGTGTTTCAAGGTGTCCGC	Altered splice acceptor site before G482

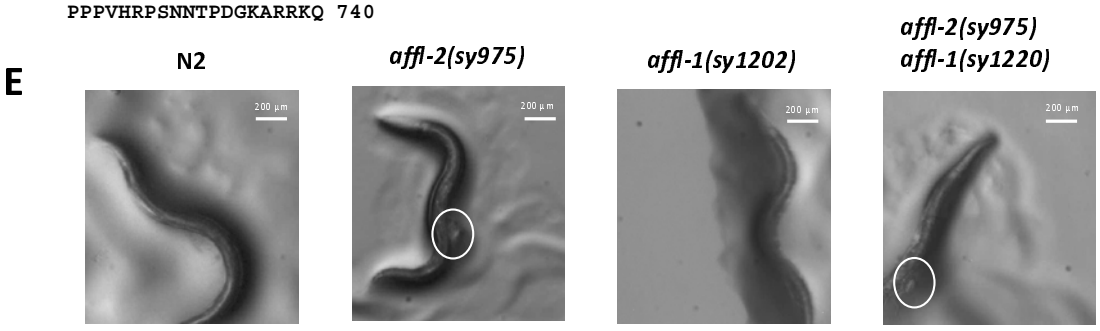
Table 1: DNA Sequence Changes of Alleles. Alleles *sy439* and *sy509* were both identified in Hadju-Cronin et. al (2004), where *affl-2* was named *sup-45*. All other alleles were identified in this screen. Bold letters indicate modified bp(s) in that allele. Uppercase represents exons and lowercase represents introns.

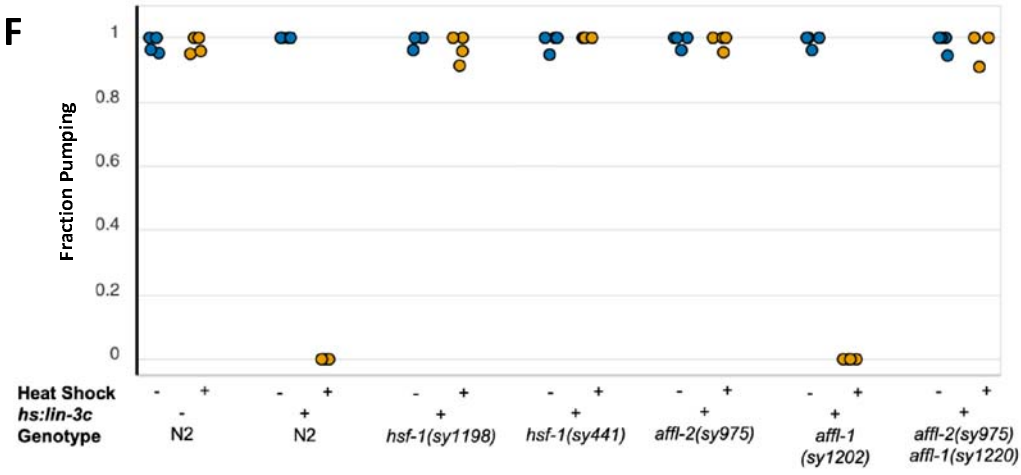
711 **Figures:**



D

```
MEEDSREDDFNGKIRRSLCEWLGPFEEFHDRVAKNQQTS RHGLIRSKMKPRPSPFVTLQAPPHESPSHVPVH 72
VALQSMKNLVDERITEITHHEFADLKKGMNNVKSSATSS TDKKRKIGEDEDSQIEVKRKKES RNYIPPSTPD 144
SGTHSTESDMIEEPNSDEVAAMLSIMKTLDTPKLSPLPKDFKVVCKEDEQPGPSSSKFPSDSSVSAASTSRE  216
                                  sy978 (R247*)
STPEPKLLGTLRIKALKPERVQKLIELAKT RGKLTDVVEVEKKVVEKVVEKQEKRVLEPLPPPPRPPSVPAV 288
PAVPTPSRDPILGPGSGSARNSPLPPQKPANKTVQSLQGLQGSSRTMSPSPRNIRSTTPSVPSRPPSALSTH 360
SHVETEVAAVATTSAPAEASAAVKPRFVCQWSHARLKKAQRVPFPDPSSTSTKTKGEFYHALAKDWKSKADN 432
                        sy975, sy981 (W456*)                sy436 (G482#)
SKDRVTRPLNYMLSSVFFVLEAI WKADKERTTQSKMQCASIYRDTYELL GVAVFQGVRDTDDSLAVHILPRV 504
KVIGQVMLAVMQYQMYIFRSEQAIKTMSRLDMREVSESIDIRPVSRASDSSSVHSQHLTVPTNKAVTSVAHG 576
 sy509 (S581+)
TPKP SPSAGSTPSGPTGAPVIPWIASIQACQNMVSMPQIVYDAYKSQLKTANGLMLASRYWEDSKTLSNIID 648
SGTFVKDVESIVGKSISMDMQFADLATFVLTAVGSLKAEYEEEQKQPAKPILAKVKKGLDLAIRTGTFHSDR 720
PPPVHRPSNNTPDGKARRKQ 740
```



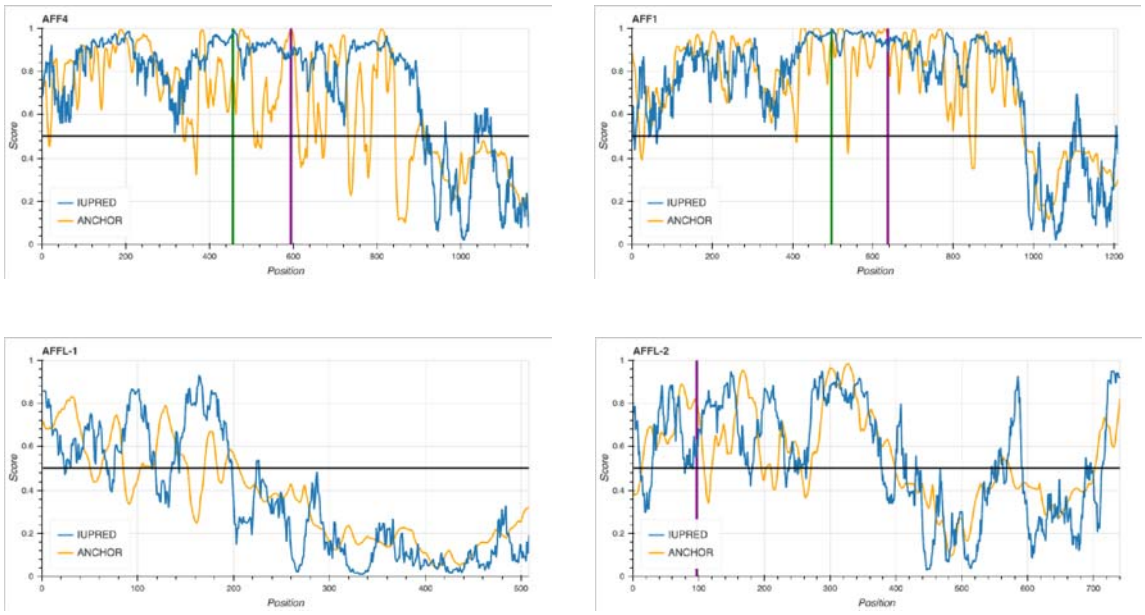


712

Figure 1. *hsf-1* and *affl-2* were discovered in a screen for suppressors of heat shock induced *lin-3c* overexpression. (A) Outline of screening process. The F2 progeny of mutagenized Phsp-16.41*:lin3-c* animals were screened for suppression of *lin-3c* overexpression induced pumping quiescence. Gene diagrams of *hsf-1* (B) and *affl-2* (C). Blocks are exons, lines are introns, and red arrows indicate position of molecular changes for alleles. Sequences for alleles can be found in table 1. (D) Annotated AFFL-2 sequence. Highlighted text in yellow corresponds to predicted NLSs, and disrupted residue due to allele change is underlined in red text. * indicates stop; # indicates splice site disruption prior to residue; + indicates first disrupted residue of missense mutation and frameshift caused by *sy509*. (E) Images of wild type (N2), *affl-2* mutant, *affl-1* mutant, and *affl-2 affl-1* double mutants. Protruding intestines from vulvas are circled. (F) Fraction of worms pumping before or after heat shock to induce *hsp-16.41: lin-3c* expression. Data are a proxy for gene expression (pumping quiescence indicates *hsp-16.41* expression). Each point represents a different trial.

728

A



729

B

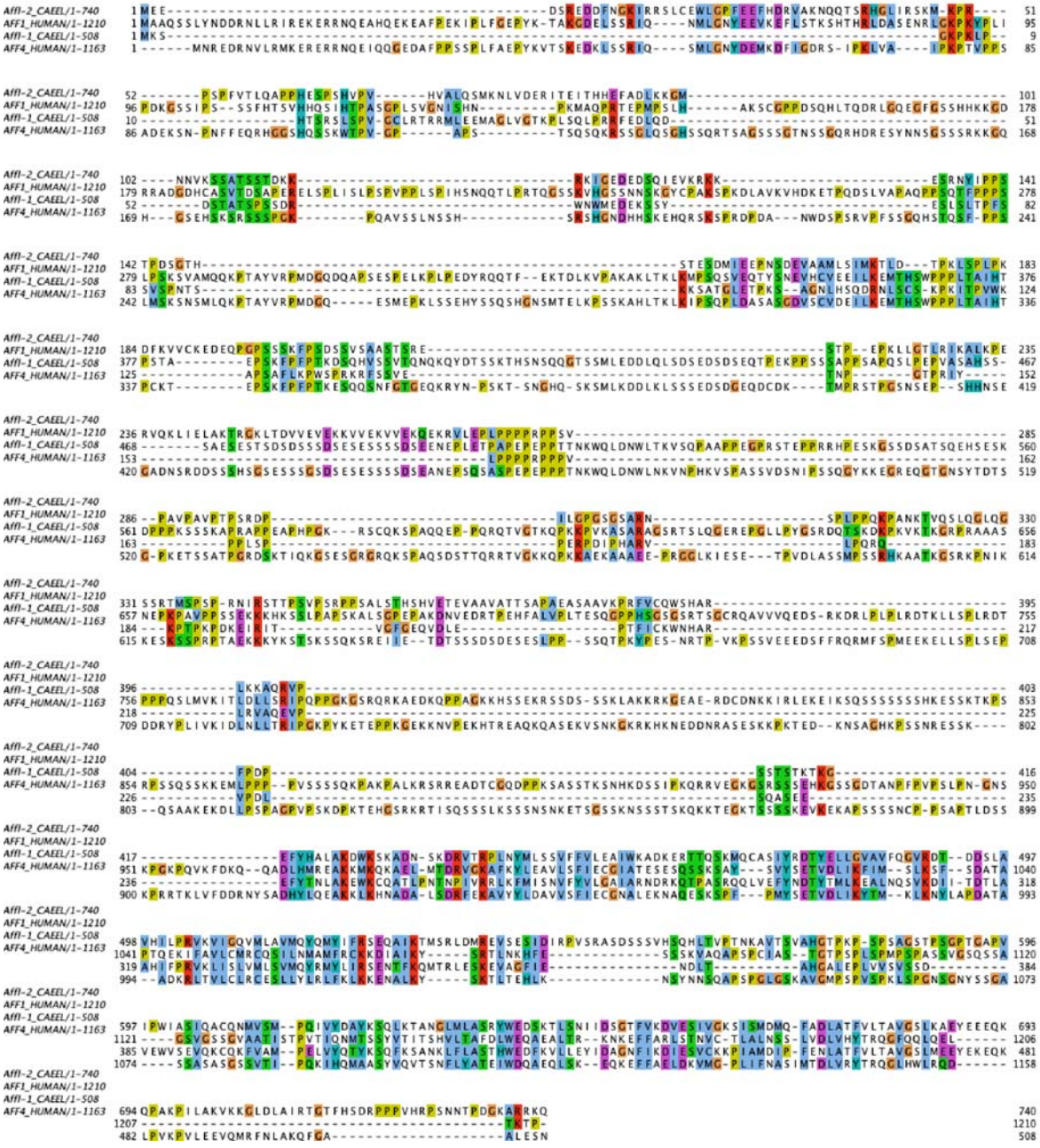
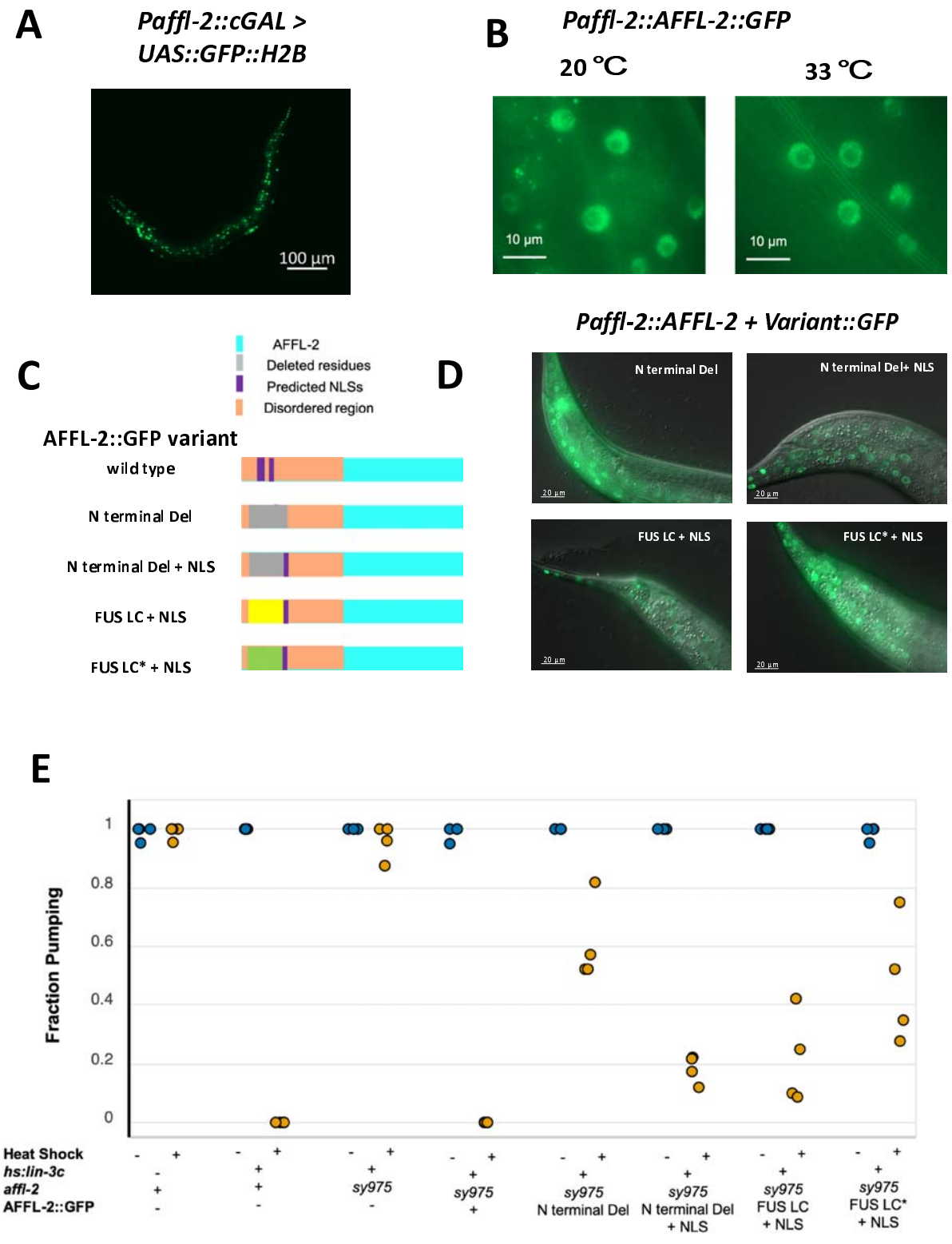
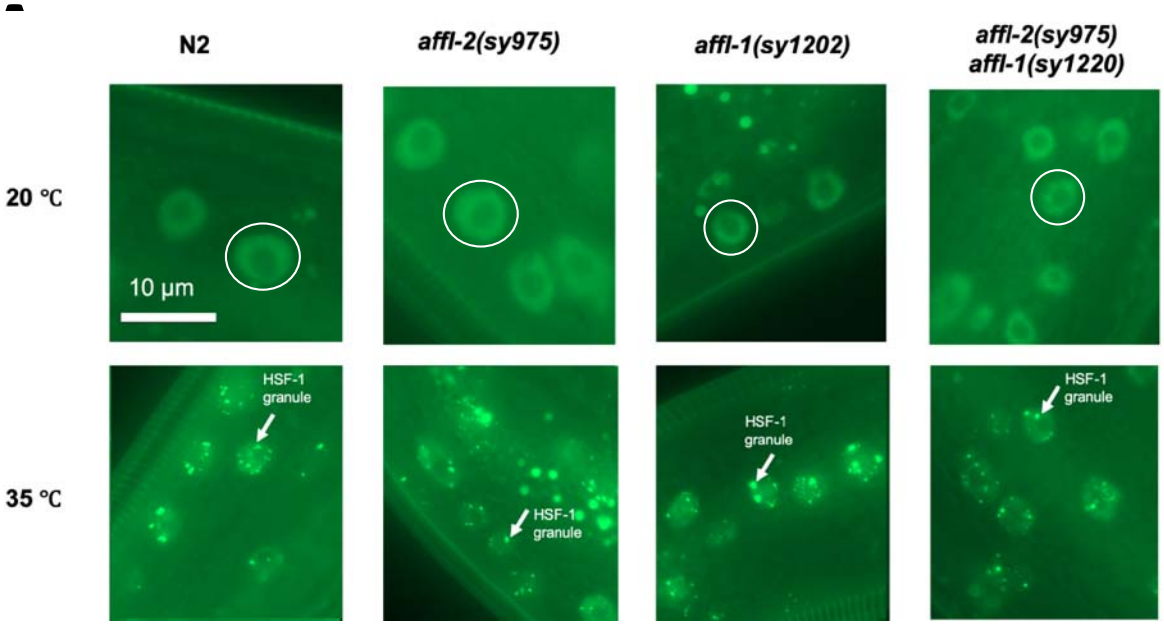


Figure 2. *affl*-2 is a homolog of the AF4/FMR2 family and predicted to be a disordered protein. (A) Plots of ANCHOR and IUPRED score per residue for AFF4, AFF1, AFFL-1, and AFFL-2. IUPRED predicts disordered residues, and higher scores indicate higher confidence that a residue is disordered (Dosztányi *et al.* 2005; Dosztanyi *et al.* 2018). ANCHOR predicts disordered protein binding regions, and higher scores indicate higher confidence that a disordered residue can participate in protein binding (Dosztányi *et al.* 2009).(B) Diagram of *H. sapiens* AFF4 with annotated binding sites (Chou *et al.* 2013; Leach *et al.* 2013; Schulze-Gahmen *et al.* 2013, 2014; Qi *et al.* 2017; Schulze-Gahmen and Hurley 2018; Chen and Cramer 2019). (C) Sequence alignment of AFFL-2, AFFl-1 (AFFL-2's paralog in *C. elegans*), and AFF1 and AFF4 in *H. sapiens*. The alignment was prepared using Jalview with ClustalX coloring to highlight conservation (Edgar 2004) , where conserved residues are colored as follows: hydrophobic (blue), positive charge (red), negative charge (magenta), polar (green), cysteines (pink), glycine (orange), proline (yellow) and aromatic (cyan). The UniProtKB accession identifiers for each sequence are listed here: Affl-2_CAEEL, Q95XW7; Affl-1_CAEEL, Q95XW6; AFF1_HUMAN, P51825; AFF4_HUMAN, Q9UHB7.

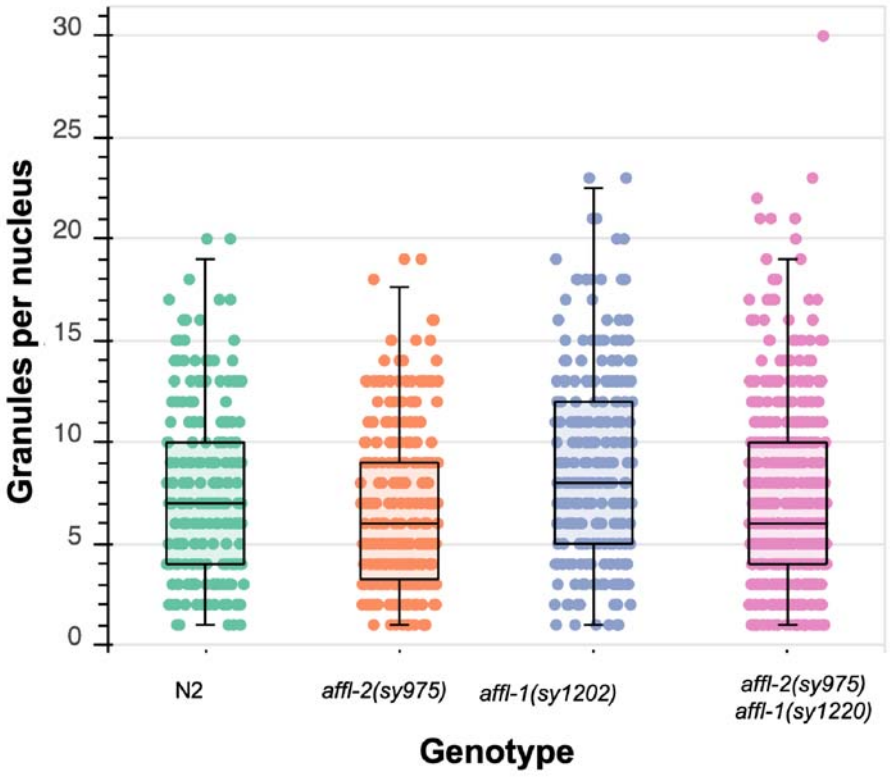


752

753 **Figure 3. AFFL-2 is a ubiquitously expressed nuclear protein.** A) Representative image of a
754 worm expressing *Paffl-2*::cGAL > *15xUAS::GFP::H2B,* which can be seen in the majority of
755 cells, indicating that *affl-2* is ubiquitously expressed. Note that the effector construct is
756 integrated, while the driver is an extrachromosomal array. B) AFFL-2::GFP at 20 °C and 33 °C.
757 Both images are taken of nuclei in the tail of young adults.
758 (C) Diagrams of AFFL-2 variants. Note that diagrams are not to scale, but are just representative
759 of the ordering of various elements. FUS LC* represents the modified FUS LC residues with
760 disordered residues mutated to more ordered ones. All constructs are driven by the *affl-2*
761 promoter (P*affl-2*). (D) Subcellular of localization of AFFL-2 variants. Animals in photos are
762 young wild type adults at room temperature. (E) Fraction of worms pumping before or after heat
763 shock to induce *hsp-16.41: lin-3c* expression. Data are a proxy for gene expression (pumping
764 quiescence indicates *hsp-16.41* expression). Each point represents a different trial.
765
766



B



767

768 **Figure 4. Analysis of the interaction of *hsf-1, affl-2,* and *affl-1*.** All genotypes contain *drSi13,*
769 which is the single-copy transgene with P*hsf-1::HSF-1::GFP*. (A) Representative images of
770 hypodermis nuclei in young adults. Subcellular localization of HSF-1::GFP with and without a
771 five-minute heat shock at 35 °C. Prior to heat shock, HSF-1::GFP is distributed throughout the
772 nucleus for all genotypes; after heat shock HSF-1::GFP forms nuclear granules. Images of nuclei
773 are from the hypodermis but are representative of HSF-1::GFP localization throughout the entire
774 animal. In top row, representative nuclei are circled. In bottom row, examples of nuclear
775 granules are indicated with arrows. (B) Quantification of HSF-1::GFP granules per nucleus after
776 heat shock for different genotypes. Each point represents one nucleus.
777