


# A Twist-Box domain of the *C. elegans* Twist homolog, HLH-8, plays a complex role in transcriptional regulation

Michael J. Gruss, Colleen O'Callaghan, Molly Donnellan, Ann K. Corsi  \*

Department of Biology, The Catholic University of America, 620 Michigan Ave., NE, Washington, D.C. 20064 USA

\*Corresponding author: Ann K. Corsi, Department of Biology, The Catholic University of America, 620 Michigan Ave., NE, Washington, D.C. 20064 USA. Email: [corsi@cua.edu](mailto:corsi@cua.edu)

## Abstract

TWIST1 is a basic helix-loop-helix (bHLH) transcription factor in humans that functions in mesoderm differentiation. TWIST1 primarily regulates genes as a transcriptional repressor often through TWIST-Box domain-mediated protein–protein interactions. The TWIST-Box also can function as an activation domain requiring 3 conserved, equidistant amino acids (LXXXFXXR). Autosomal dominant mutations in *TWIST1*, including 2 reported in these conserved amino acids (F187L and R191M), lead to craniofacial defects in Saethre–Chotzen syndrome (SCS). *Caenorhabditis elegans* has a single TWIST1 homolog, HLH-8, that functions in the differentiation of the muscles responsible for egg laying and defecation. Null alleles in *hlh-8* lead to severely egg-laying defective and constipated animals due to defects in the corresponding muscles. TWIST1 and HLH-8 share sequence identity in their bHLH regions; however, the domain responsible for the transcriptional activity of HLH-8 is unknown. Sequence alignment suggests that HLH-8 has a TWIST-Box LXXXFXXR motif; however, its function also is unknown. CRISPR/Cas9 genome editing was utilized to generate a domain deletion and several missense mutations, including those analogous to SCS patients, in the 3 conserved HLH-8 amino acids to investigate their functional role. The TWIST-Box alleles did not phenocopy *hlh-8* null mutants. The strongest phenotype detected was a retentive (Ret) phenotype with late-stage embryos in the hermaphrodite uterus. Further, GFP reporters of HLH-8 downstream target genes (*arg-1::gfp* and *egl-15::gfp*) revealed tissue-specific, target-specific, and allele-specific defects. Overall, the TWIST-Box in HLH-8 is partially required for the protein's transcriptional activity, and the conserved amino acids contribute unequally to the domain's function.

**Keywords:** *C. elegans*, HLH-8, TWIST1, Saethre–Chotzen syndrome, Twist-Box domain, tissue-specific transcriptional regulation

## Introduction

During the development of a multicellular organism, genes are upregulated or downregulated in a spatial-temporal manner through the coordination of multiple transcription factors (TFs). One example of a tissue-specific TF is TWIST1, which encodes a basic helix-loop-helix (bHLH) TF that functions primarily as a repressor but also can activate transcription in mammals and is conserved across a variety of species ranging from nematodes to humans (Gitelman 1997; Castanon et al. 2001; Rodriguez et al. 2016). Twist family members in various species share a common feature of influencing the development of multiple mesodermally-derived tissues. An example of a tissue regulated by TWIST1 is the cranial sutures, where TWIST1 negatively regulates osteogenesis to preserve the potency of the mesenchymal stem cells of the sutural osteogenic front. When TWIST1 is defective, the stem cells differentiate, leading to the premature bone formation and suture fusion known as craniosynostosis (reviewed in Wilkie 1997; Twigg and Wilkie 2015). Craniosynostosis is characteristic of Saethre–Chotzen syndrome (SCS), which is an autosomal dominant, haploinsufficient disorder occurring in every 1/25,000 to 1/50,000 live births, with dozens of distinct mutations reported in *TWIST1* (reviewed in Barnes and Firulli 2009). Many of the missense mutations occur in the conserved bHLH domains. The bHLH family of TFs functions as dimers through the  $\alpha$ -helical HLH domain to form either homodimers, such as TWIST1–TWIST1, or

heterodimers with several different bHLH TFs, such as TWIST1–E12 and other E family bHLH proteins. Once dimers are formed, they bind to DNA via the basic domains (reviewed in Barnes and Firulli 2009) at DNA sequences called E-boxes (CANNTG) (Thisse et al. 1988). Once bound to DNA, bHLH TFs can influence transcription through a variety of mechanisms including interacting with other TFs, chromatin remodelers, and histone modifiers through 1 or more activation domains (ADs) (reviewed in de Martin et al. 2021).

The ADs of bHLH TFs differ significantly between both orthologs and paralogs. An example of this lack of conservation is seen when comparing the myogenic regulatory factors (MRFs) Myf5, MyoD1, Myogenin, and MRF4, which function together in skeletal muscle development (reviewed in Singh and Dilworth 2013; Zammit 2017). Like other bHLH proteins, these factors are highly conserved in their bHLH domains, but their amino and carboxyl terminal domains containing the ADs have much lower conservation (reviewed in Singh and Dilworth 2013). Domain swapping experiments among the MRFs have shown that the ADs provide the activity for each factor's specific function, e.g. muscle specification or differentiation (reviewed in Singh and Dilworth 2013). Further, swapping the C-terminal AD from invertebrate proteins such as HLH-1, the *C. elegans* MyoD homolog, into murine MyoD1 (mMyoD1) created chimeric MyoD proteins that were still capable of transactivation of mMyoD1 genes (Bergstrom and Tapscott 2001). These experiments support the suggestion that even with the low conservation in ADs, there

must be some evolutionary pressure to preserve specific amino acids and/or a protein structure to still allow transcriptional activation; in particular, many ADs across bHLHs are predicted to form alpha helices (Weintraub et al. 1991; Schwarz et al. 1992; Quong et al. 1993; Massari et al. 1996; McAndrew et al. 1998; Denis et al. 2014).

The C-terminus of mammalian TWIST1 contains a domain referred to as either the WR domain or the TWIST-Box. This domain contributes to TWIST1 transcriptional activation and repression depending on the target gene (reviewed in Franco et al. 2011). TWIST-Box-mediated repression appears to be primarily through direct protein–protein interactions. Bialek and colleagues showed that TWIST1 bound directly to the Runx2 osteogenic driver through a TWIST-Box-dependent interaction in the cranial sutures to delay suture closure (Bialek et al. 2004). Several other examples of TWIST-Box directed repression have been observed with the chondrogenic master regulator Sox9 (Gu et al. 2012), the p53 cell cycle/apoptotic regulator (Piccinin et al. 2012), and Snail1/2 epithelial-to-mesenchymal (EMT) TFs that are required to form the neural crest (NC) (Lander et al. 2013). In contrast, the TWIST-Box also has been shown to play a role in the transactivation of genes (Laursen et al. 2007; Franco et al. 2011) leading to the progression, increased EMT, and metastasis of various cancers that are over expressing TWIST1 (Gajula et al. 2013). For example, in breast cancer, the TWIST-Box cooperates with the NF- $\kappa$ B RELA subunit to transactivate interleukin 8 (IL8) leading to increased expression of metalloproteinases that drive invasion (Li et al. 2012). Finally, although a majority of SCS mutations are located in the highly conserved basic and HLH domains, several TWIST-Box mutations have been discovered in SCS patients (Kress et al. 2006; Seto et al. 2007; Peña et al. 2010).

The mammalian TWIST1 AD was first described by Laursen and colleagues who showed that mouse TWIST1/E12 heterodimeric complexes could transactivate a *Drosophila*-derived reporter construct and that TWIST-Box residues 165–206 were sufficient for the activity (Laursen et al. 2007). Amino acid alignment across multiple species revealed 3 equidistant highly conserved residues in the TWIST-Box, LXXXFXXXR, that were essential for transactivation by the TWIST1/E12 heterodimers, and Chou-Fasman secondary structure prediction suggested the TWIST-Box adopted an  $\alpha$ -helical conformation with the conserved amino acids on 1 face (Laursen et al. 2007) although some literature suggests that the TWIST-Box is an intrinsically disordered region (Kaustov et al. 2006; Rodriguez et al. 2016). When the entire TWIST-Box domain was deleted or when alanine substitutions to each individual or all 3 of the residues were made, an 80 to 100% decrease of reporter gene activity was observed leading to the conclusion that the TWIST-Box was an important AD in TWIST1 (Laursen et al. 2007).

*Caenorhabditis elegans*, the free-living nematode model organism, has only 1 TWIST-related homolog, HLH-8 (Harfe et al. 1998; Corsi et al. 2000). Similar to mammals and other higher organisms, HLH-8 regulates the differentiation of a subset of mesodermal cells. The *C. elegans* mesodermal cells include body wall muscles, coelomocytes, uterine muscles (UMs), vulval muscles (VMs), enteric muscles (EMs), pharyngeal muscles, and the somatic gonad (Sulston and Horvitz 1977; Horvitz and Sulston 1980). Of the 126 embryonic mesodermally-derived cells, only 4 of them express HLH-8: the anal depressor and 2 intestinal muscle cells (all EMs) and the M mesoblast cell, which gives rise to the VMs and UMs during postembryonic development (Harfe et al. 1998; Corsi et al. 2000). In *hlh-8* (*nr2061*) null animals, referred to here as *hlh-8* (*d*), these VMs and EMs fail to develop properly resulting in egg-laying

defective (Egl) and constipated (Con) phenotypes, respectively (Corsi et al. 2000). These phenotypes are straightforward to measure and can be used to determine the severity of individual *hlh-8* alleles (Kim et al. 2017; Gruss and Corsi 2022). In addition, several HLH-8 downstream target GFP-tagged transcriptional reporter genes can be used to assess HLH-8 function (Harfe et al. 1998; Zhao et al. 2007; Kim et al. 2017). To date, HLH-8 appears to be an activator and not a repressor; however, a transactivation domain for this protein has yet to be identified.

Outside of the bHLH domain, TWIST1 and HLH-8 share very low conservation of amino acids; in particular, the C terminus is nearly 3 times longer in HLH-8 compared to TWIST1. Therefore, depending on how the C termini of the 2 orthologs are aligned, opposite conclusions can be reached: either HLH-8 does not have a domain resembling a TWIST-Box (Qin et al. 2012) or HLH-8 has the features of a TWIST-Box since it has the conserved amino acids that are important for TWIST-Box activity (Laursen et al. 2007). Because we are interested in understanding how HLH-8 is functioning in transcription and in using it as a model for investigating genotype/phenotype correlations in patient mutations (Kim et al. 2017), we used CRISPR/Cas9 to make genomic mutations to investigate whether the LXXXFXXXR sequence in HLH-8 is a TWIST-Box similar to the domain in TWIST1. We hypothesized that an HLH-8 TWIST-Box might be necessary for all of the protein's transcriptional activity and deleting the domain might lead to a null phenotype. Surprisingly, the deletion animals had a much milder phenotype compared to a null allele (Corsi et al. 2000; Kim et al. 2017). To investigate the contribution of the conserved amino acids to the domain function, we took complementary strategies: first making alanine substitutions and then engineering known patient mutations, predicting both approaches would disrupt the function of the TWIST-Box in HLH-8. Again, we were surprised to find that the 3 amino acids did not contribute equally to the domain function with tissue and target gene specific differences depending on the substitution. Further, the patient mutations had distinct phenotypes compared to the corresponding alanine substitution suggesting only a partial loss-of-function of the domain's activity in the patients (Kress et al. 2006; Seto et al. 2007; Peña et al. 2010). Overall, we found that HLH-8 has a TWIST-Box that contributes to the tissue-specific transcriptional function of the protein.

## Materials and methods

### Protein modeling

Since there are no available crystal structures of TWIST1 or HLH-8, 2 protein modeling approaches were utilized to establish a template for each protein in order to predict potential structural/functional changes in HLH-8 mutants. Homology-based modeling was performed utilizing the known predicted structures of TWIST1 (UniProt# Q15672) and HLH-8 (UniProt# Q11094) according to the AlphaFold protein structural database (<https://alphafold.ebi.ac.uk/>). The RoseTTAFold Robetta software (<https://rosetta.bakerlab.org/>), which utilizes deep learning sequence predictions, was utilized to generate a model for each of the HLH-8 CRISPR/Cas9 edits. Once the HLH-8 template models were generated, they were saved as independent program database (PDB) files. Finally the PDB files were uploaded into the SWISS-MODEL program (<https://swissmodel.expasy.org/>) to make comparisons of 3D protein folding for wild type (WT) compared to mutant HLH-8. The reason for using the Robetta template rather than the AlphaFold homology-based template was to observe predicted structural changes outside of the mutated region

that were not possible when the remaining amino acids were constrained by mapping onto a wild-type template.

DynaMut2 (<http://biosig.unimelb.edu.au/dynamut2/>) software was used to predict how mutations would alter interactions with nearby amino acids and potentially explain phenotypic differences observed in the mutant animals. In brief, the Robetta-generated PDB files of the HLH-8 TWIST-Box mutant templates [e.g. HLH-8 (F99A)] were individually uploaded into the “wild type structure” box. From there, the revertant mutation (e.g. A99F) was entered into the “Mutation details” section. The resulting structures allowed visualization of the number of predicted bonds [aromatic, ionic, polar, hydrophobic, carbonyl, H-bonds, and van der Waals (VDWs)] for each mutant as well as the wild-type template at the corresponding position.

## Generation of *hlh-8* TWIST-Box domain alleles by CRISPR/Cas9

To generate all TWIST-Box mutant alleles, CRISPR/Cas9 gene editing (Paix et al. 2015) was used with a co-conversion *dpy-10* injection marker causing F1 edited progeny to have the dominant Roller (Rol) phenotype (Arribere et al. 2014). In total, 9 individual mutant alleles were constructed (Figs. 1–3, Supplementary Fig. 1) utilizing a common 5'-AGTCGGCTTTC AATATG TGG **AGG**-3' sgRNA (bold and underlined are the Phe99 and Arg103 SCS patient-mutated residues, respectively; the AGG also serves as the PAM site and is not included in the gRNA) to either delete the entire TWIST-Box domain (Figs. 1 and 2) or create missense mutations at 1 or all 3 of the evolutionarily conserved 95-LXXXFXXXR-103 TWIST-Box residues in the *C. elegans* HLH-8 homolog (Fig. 3, Supplementary Fig. 1). To ensure the Cas9 had minimum off-target effects on the genome, the sgRNA was analyzed utilizing [https://www.idtdna.com/site/order/designtool/index/CRISPR\\_CUSTOM](https://www.idtdna.com/site/order/designtool/index/CRISPR_CUSTOM) from the IDT website. The selected sgRNA had an on-target score of 65, but a strong 100 off-target score indicating that off-target editing was highly unlikely with only 3 potential off-target sites all with scores  $\leq 75$ .

In order to delete the HLH-8 TWIST-Box domain, a repair oligonucleotide was used: 5'-GCCATTCGATCTTCGACGAAAAACGTG GATACA**AAT**GGTAACAATGGATACACACCAATTGCTGGACCTTC-3' with Asp94 (in bold) and Gly104 (underlined) located at the boundaries of the 27 bp deletion. This oligo is expected to inhibit re-cutting by Cas9 since the PAM site (Arg103) is deleted. In order to verify the homozygous deletion mutants in subsequent generations, a PCR primer, underlined in Fig. 2c, was utilized to amplify a 412 bp band specific to the TWIST-Box deletion. From 6 injected worms, 89 rollers were isolated, allowed to lay progeny, and then processed for single worm lysis/PCR (Fay 2013). However, after screening through 10 rollers, triplicate independent  $\Delta 95$ –103 TWIST-Box deletion strains were identified.

For the missense mutants, alanine substitutions or SCS patient analogous alleles were generated as follows (see Table 1): AK183 *hlh-8* (th11) Leu95Ala: tta→gcc; AK188 *hlh-8* (th12) Phe99Ala: ttc→gcg; AK193 *hlh-8* (th13) Arg103Ala: agg→gcg, AK198: *hlh-8* (th14) Leu95Ala; Phe99A; Arg103A: tta→gcc + ttc→gcg + agg→gcg, AK213 *hlh-8* (th17) Leu95Phe: tta→ttc, AK219 *hlh-8* (th18) Phe99Leu: ttc→ctt, AK226 *hlh-8* (th19) Arg103Met: agg→atg, AK203 *hlh-8* (th15) Leu95Phe; Phe99Leu: tta→ttc + ttc→ctt. Since the PAM site occurs at Arg103, any strain with a missense mutation at that position (i.e. R103A, R103M, and Triple A) will have the PAM site altered. For the remaining strains, a silent Arg103 agg→gcg mutation was incorporated to avoid Cas9 recutting. To aid in screening through edited candidates, an additional silent mutation at the Asn100 codon aat→aac creates an AflIII

restriction site to use for RFLP analysis of PCR fragments (Supplementary Fig. 1) (Botstein et al. 1980).

The CRISPR/Cas9 injection mixes (Paix et al. 2015) contained the following in 20  $\mu$ L: 1.5  $\mu$ L Cas9 (10  $\mu$ g/ $\mu$ L), 0.5  $\mu$ L KCl (1 M), 0.75  $\mu$ L HEPES buffer (200 mM, pH 7.4), 0.3  $\mu$ L *dpy-10* crRNA (8  $\mu$ g/ $\mu$ L), 1.5  $\mu$ L Target crRNA (8  $\mu$ g/ $\mu$ L), 5  $\mu$ L tracrRNA (4  $\mu$ g/ $\mu$ L), 0.35  $\mu$ L *dpy-10* ssODN (600 ng/ $\mu$ L), 3  $\mu$ L of each Target gene ssODN (1  $\mu$ g/ $\mu$ L, dissolved in 5 mM Tris-HCL buffer pH 7.5), and H<sub>2</sub>O to 20  $\mu$ L. Two separate injection mixes were created for injection into 20 young adult N2 animals. One injection mix contained repair oligonucleotide sequences (ssODN) for the L95F, F99L, and R103M strains, while a second injection mix contained the repair oligonucleotide sequences for the L95A, F99A, R103A, Triple A, and L95F; F99L alleles. All repair nucleotides had 35 bp homology arms upstream (5'-GCCATTCGATCTTCGACGAAAAACGTGGATACAAT-3') of Leu95 and downstream (5'-GGTAACAATGGATACACCAATTGCTGGACCTTC-3') of Arg103. In brief, once the 20 animals were injected, they were singled on individual plates and screened for *dpy-10*-edited animals with the Rol phenotype that were isolated and processed for PCR and RFLP analysis after laying F2 progeny. Candidates were followed by PCR/RFLP until homozygous edits were isolated, which were confirmed via Sanger sequencing (Retrogen, San Diego, CA).

## Strains and genetic crosses

All strains were grown and maintained at 20°C unless otherwise indicated. All strains used are listed below and some in Table 1:

N2 wild type

AG316: *arg-1::gfp* (*ccls4444* II); *mIs10* (*[myo-2p::gfp + pes10p::gfp + gut-promoter::gfp]* V); *him-8* (*e1489*)

KM190: *egl-15::gfp* (*ayIs2* IV)

PD4443: *arg-1::gfp* (*ccls4443* IV)

PD4667: *hlh-8::gfp* (*ayIs7* IV)

PD4793: *mIs10* (*[myo-2p::gfp + pes10p::gfp + gut-promoter::gfp]* V)  
TY4236: *him-8*; *mIs10*

AK181: *egl-15::gfp*; *him-8*; *mIs10*

AK182: *him-8*; *hlh-8::gfp*; *mIs10*

AK183: *hlh-8* (*th11*) Leu95Ala

AK184: *arg-1::gfp*; *hlh-8* (*th11*) Leu95Ala

AK185: *egl-15::gfp*; *hlh-8* (*th11*) Leu95Ala

AK186: *hlh-8::gfp*; *hlh-8* (*th11*) Leu95Ala

AK187: *mIs10*; *hlh-8* (*th11*) Leu95Ala

AK188: *hlh-8* (*th12*) Phe99Ala

AK189: *arg-1::gfp*; *hlh-8* (*th12*) Phe99Ala

AK190: *egl-15::gfp*; *hlh-8* (*th12*) Phe99Ala

AK191: *hlh-8::gfp*; *hlh-8* (*th12*) Phe99Ala

AK192: *mIs10*; *hlh-8* (*th12*) Phe99Ala

AK193: *hlh-8* (*th13*) Arg103Ala

AK194: *arg-1::gfp*; *hlh-8* (*th13*) Arg103Ala

AK195: *egl-15::gfp*; *hlh-8* (*th13*) Arg103Ala

AK196: *hlh-8::gfp*; *hlh-8* (*th13*) Arg103Ala

AK197: *mIs10*; *hlh-8* (*th13*) Arg103Ala

AK198: *hlh-8* (*th14*) Leu95Ala; Phe99A; Arg103A

AK199: *arg-1::gfp*; *hlh-8* (*th14*) Leu95Ala; Phe99A; Arg103A

AK200: *egl-15::gfp*; *hlh-8* (*th14*) Leu95Ala; Phe99A; Arg103A

AK201: *hlh-8::gfp*; *hlh-8* (*th14*) Leu95Ala; Phe99A; Arg103A

AK202: *mIs10*; *hlh-8* (*th14*) Leu95Ala; Phe99A; Arg103A

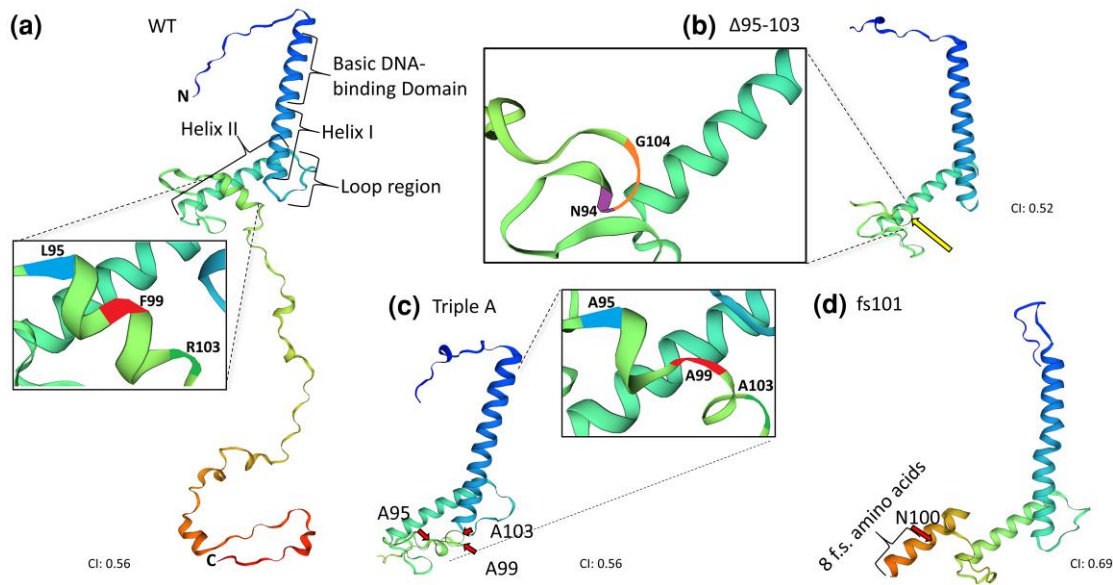
AK203: *hlh-8* (*th15*) Leu95Phe; Phe99Leu

AK204: *arg-1::gfp*; *hlh-8* (*th15*) Leu95Phe; Phe99Leu

AK205: *egl-15::gfp*; *hlh-8* (*th15*) Leu95Phe; Phe99Leu

AK206: *hlh-8::gfp*; *hlh-8* (*th15*) Leu95Phe; Phe99Leu

AK207: *mIs10*; *hlh-8* (*th15*) Leu95Phe; Phe99Leu



**Fig. 1.** The HLH-8 TWIST-Box LXXXFXXXR conserved motif is predicted to form an  $\alpha$ -helix. HLH-8 a) WT and b–d) mutant structures were modeled using a template from RoseTTAFold software in SWISS-MODEL. a) The N-terminal basic DNA binding domain of HLH-8 leads into Helix I followed by a loop into Helix II (labeled brackets). The evolutionarily-conserved amino acids in the putative TWIST-Box adopt an  $\alpha$ -helical secondary structure that spans from the conserved L95 to R103; the presumed location of each labeled amino acid on the helix can be seen in the magnified image (box). b) The predicted structure of the  $\Delta 95$ –103 TWIST-Box deletion is missing the helix (arrow); the magnified view (box) shows the location of the 2 amino acids surrounding the deletion N94 and G104. b, c) The C-terminal regions of the proteins are not shown. c) The Triple A (three arrows) mutant predicted structure suggests that the 3 alanine mutations at the conserved residues disrupt the proper  $\alpha$ -helical formation of the HLH-8 TWIST-Box with the magnified view (box) showing the A95, A99, and A103 locations. d) The NHEJ mutant, fs101, still is predicted to form an  $\alpha$ -helix although there are 8 amino acids (bracketed) after the N100 (arrow) that are not normally found in HLH-8 followed by a premature stop codon. Interestingly, these additional amino acids are predicted to adopt an  $\alpha$ -helical structure which may explain why the phenotype of the fs101 animals does not phenocopy the *hlh-8(A)* animals (see “Discussion”). All modeling was performed with independent RoseTTAFold HLH-8 templates utilizing the Robetta software (see “Methods”). The confidence score (CI) is indicated next to each model.

AK208: *hlh-8 (th16) Δ95–103*  
 AK209: *arg-1::gfp; hlh-8 (th16) Δ95–103*  
 AK210: *egl-15::gfp; hlh-8 (th16) Δ95–103*  
 AK211: *hlh-8::gfp; hlh-8 (th16) Δ95–103*  
 AK212: *mls10; hlh-8 (th16) Δ95–103*  
 AK213: *hlh-8 (th17) Leu95Phe*  
 AK214: *arg-1::gfp; hlh-8 (th17) Leu95Phe*  
 AK215: *egl-15::gfp; hlh-8 (th17) Leu95Phe*  
 AK216: *hlh-8::gfp; hlh-8 (th17) Leu95Phe*  
 AK217: *mls10; hlh-8 (th17) Leu95Phe*  
 AK219: *hlh-8 (th18) Phe99Leu*  
 AK220: *arg-1::gfp; hlh-8 (th18) Phe99Leu*  
 AK221: *arg-1::gfp; him-8; hlh-8 (th18) Phe99Leu*  
 AK222: *egl-15::gfp; hlh-8 (th18) Phe99Leu*  
 AK223: *hlh-8::gfp; hlh-8 (th18) Phe99Leu*  
 AK224: *mls10; hlh-8 (th18) Phe99Leu*  
 AK226: *hlh-8 (th19) Arg103Met*  
 AK227: *arg-1::gfp; hlh-8 (th19) Arg103Met*  
 AK228: *egl-15::gfp; hlh-8 (th19) Arg103Met*  
 AK229: *hlh-8::gfp; hlh-8 (th19) Arg103Met*  
 AK230: *mls10; hlh-8 (th19) Arg103Met*  
 AK242: *hlh-8 (th20) fs101*  
 AK243: *arg-1::gfp; hlh-8 (th20) fs101*  
 AK244: *egl-15::gfp; hlh-8 (th20) fs101*  
 AK245: *hlh-8::gfp; hlh-8 (th20) fs101*  
 AK246: *mls10; hlh-8 (th20) fs101*

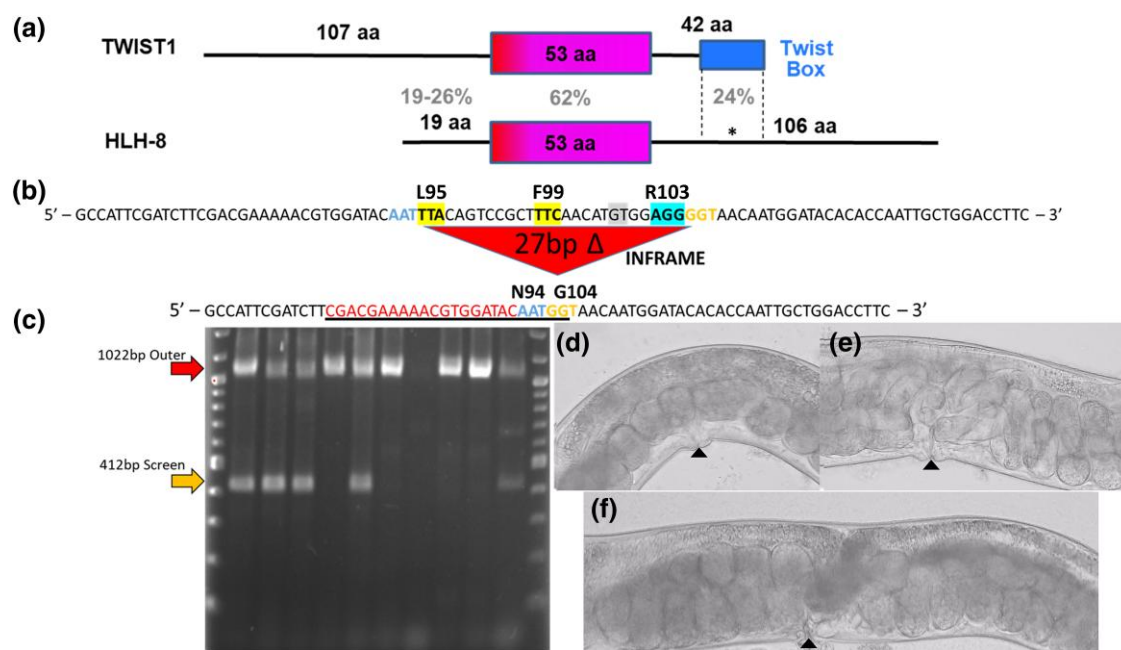
In order to make strains that were homozygous for the TWIST-Box mutants and the integrated *egl-15::gfp*, *arg-1::gfp*, *myo-2::gfp*, or *hlh-8::gfp* reporter strains, each reporter strain contained a *him-8 (e1489)* mutation to generate males for the crosses.

Since many of the mutants did not have an obvious phenotype, it was also convenient to use the *mls10 myo-2::gfp* reporter in addition to the other reporter constructs to distinguish cross progeny containing a green pharynx. We also used PCR/RFLP screening (identical to initial CRISPR/Cas9 screening method above) to follow the TWIST-Box mutations and ensure they were homozygous before examining the GFP phenotype. In cases where the GFP reporter was expressed in at least 1 tissue, we confirmed the reporter was homozygous by ensuring that 100% of the progeny expressed the GFP. In cases where the specific TWIST-Box mutant resulted in complete loss of GFP reporter expression, the candidate homozygous animals were backcrossed with *him-8; myo-2::gfp* to observe the reappearance of the GFP expression in 100% of the F1 progeny.

To determine whether any of the TWIST-Box heterozygotes had dominant phenotypes for GFP expression, the TWIST-Box homozygous mutant reporter strains (*egl-15::gfp* or *arg-1::gfp*) were crossed with the respective *myo-2::gfp*, *him-8*; *egl-15::gfp* or *arg-1::gfp* males. The *myo-2::gfp* expressing F1 cross progeny L4 animals were scored 24 h later for the GFP expression pattern.

## Microscopy and image capturing

For scoring of the TWIST-Box mutant allele's effect on *hlh-8* downstream target GFP reporter gene expression (*arg-1::gfp* or *egl-15::gfp*), 20–50 L4 animals were isolated and observed after 24 or 48 h with a compound epifluorescent microscope (Gruss and Corsi 2022). Images were taken with a XCAM Family 1080P HD Camera using TouView 3.7 software. For all GFP quantifications, the exposure time of the TWIST-Box mutant animals was normalized to the UV light exposure for the respective N2 transcriptional



**Fig. 2.** A TWIST-Box deletion does not phenocopy *hlh-8* (*A*). a) The HLH-8 and TWIST1 proteins have the highest amino acid identity (62%) in the basic DNA-binding and the dimer-forming helix-loop-helix domains (boxes labeled 53 aa). The TWIST-Box domain (box at the C terminus of TWIST1) has low amino acid identity (24%) between the 2 homologs. Despite the low identity, the 3 amino acids above the highlighted codons in b) are conserved across a variety of species. Note that the TWIST1 TWIST-Box domain is located at the C-terminus whereas the corresponding conserved amino acids in HLH-8 are not (\* in a). b) The experimental design for the  $\Delta 95$ –103 TWIST-Box allele in HLH-8 shows a PAM site coincident with R103 that was used to generate all alleles in this study with an upstream GT (highlighted) Cas9 endonuclease cleavage. The 27 bp deletion will cause N94 and G104 to become adjacent amino acids so a screening primer (underlined) was designed to identify candidates by PCR. c) An example gel from a PCR using 3 primers shows a 1022 bp band amplified from a primer pair outside the edited region (outer) and a 412 bp band amplified using the  $\Delta 95$ –103 screening primer. The outer lanes contain a 100 bp DNA ladder (NEB). d–f) d) A WT (N2) hermaphrodite has a single row of embryos in the uterus whereas e) *fs101* and f)  $\Delta 95$ –103 animals have more than 1 row of embryos at later stages of development, which characterizes the retentive (Ret) phenotype. All images depict the vulval/uterine region of the worms with the arrowhead marking the vulva.

reporter strain. In this case for all *egl-15::gfp*, *arg-1::gfp*, *myo-2::gfp*, or *hlh-8::gfp* scoring and image captures, the camera was set to the following restraints: target 9, exposure time 39 ms, and gain 59. Therefore, any alteration in the intensity of the GFP in the micrographs is not due to alteration of the camera or UV exposure time. Animals were placed in GFP categories based on the distinct intensity of GFP in the overall tissue (e.g. across all the VMs) with 0, no expression; 1, almost no expression but the outline of the cells visible; 2, all cells or 3 out of 4 of the muscle cells expressing distinct GFP; 3, bright GFP in all cells comparable to the intensity of the *myo-2::gfp* in the pharynx (Kim et al. 2017). For all reporter gene assay experiments and quantification, at least 3 independent experiments were performed.

## Phenotypic assays

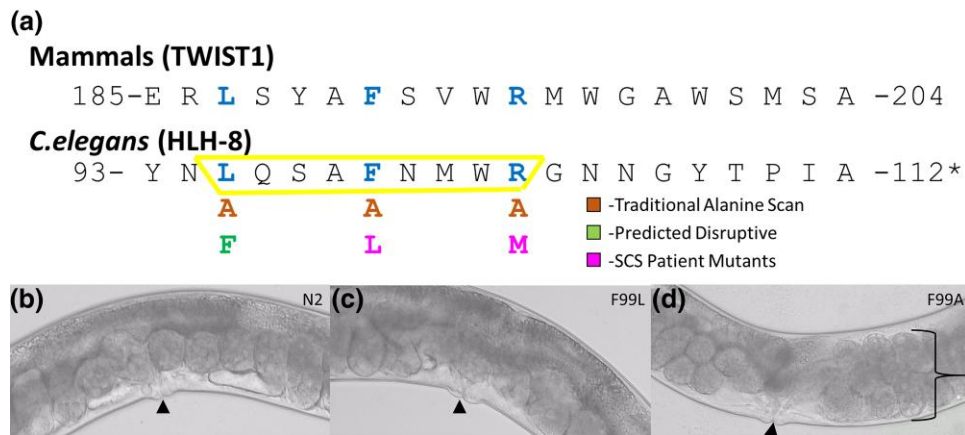
*myo-2::gfp* Ret assay: The degree of embryo retention in the uterus of TWIST-Box domain mutants was analyzed in the *myo-2::gfp* lines. The *myo-2::gfp* transgene is expressed in the developing pharynx only during late embryogenesis typically after being laid by the hermaphrodite. To characterize the degree of retention in the HLH-8 mutants, a group of 20–30 L4 hermaphrodites for each genotype was isolated to a plate and allowed to develop for 24 h or 48 h. Then, the animals were viewed with a compound epifluorescent scope. Each hermaphrodite was placed into 1 of 4 categories based on the number of *myo-2::gfp* expressing later stage embryos/larvae still residing within the adult uterus (0, 1–3, 4–6, or  $\geq 7$ ). WT animals are not expected to have any embryos with pharyngeal GFP expression and increasing numbers suggest more severe Ret phenotypes.

*arg-1::gfp* and *egl-15::gfp* expression assays: Expression of either HLH-8 target gene was analyzed utilizing an integrated, multi-copy transgene. The *arg-1* reporter gene contains 2.8 kb (Zhao et al. 2007) and the *egl-15* reporter contains 2.1 kb upstream of the start codon for each gene driving *gfp* (Harfe et al. 1998). L4 hermaphrodites that were homozygous or heterozygous for the TWIST-Box mutations were isolated and allowed to age 24 or 48 h. The percentage of the population expressing *arg-1::gfp* in either the head mesodermal cell (HMC), VMs, and ENT muscles (INT, SPH, DEP) or *egl-15::gfp* in the VMs was quantified based on the degree of intensity of green when compared to WT; *arg-1::gfp* or *egl-15::gfp* animals. The categories were defined by distinct, notable changes in the level of GFP to be either bright, medium, dim, or loss of GFP expression as defined above.

## Results

### Deep learning and homology-based modeling suggests the presence of an $\alpha$ -helical TWIST-Box domain in HLH-8

Previous structure/function studies in TWIST1, including Chou-Fasman secondary structure prediction, suggested the TWIST-Box forms an  $\alpha$ -helical structure (Laurson et al. 2007). Since HLH-8 shared the same spacing of the critical TWIST-Box amino acids, we used protein modeling to test whether the C. *elegans* Twist homolog also was likely to form an  $\alpha$ -helix and have a similar domain. Robetta RoseTTAFold modeling and AlphaFold homology-based modeling were both utilized to construct HLH-8 3D structures for analysis in the SWISS-MODEL software. The



**Fig. 3.** Some HLH-8 TWIST-Box missense mutations are Ret similar to  $\Delta 95$ –103 mutants. a) Alignment of the TWIST-Box domain with the conserved L95, F99, and R103 marked in bold. Below the HLH-8 sequence, the amino acid substitutions in this study are indicated. The trapezoid highlights the  $\Delta 95$ –103 TWIST-Box deletion (see Fig. 1). Each alanine substitution is indicated as well as the F99L and R103M SCS patient mutations (from Kress et al. 2006; Peña et al. 2010). The L95F allele is predicted to be disruptive. The HLH-8 TWIST-Box is not located at the C-terminus as it is in TWIST1 (\*). b) WT (N2), c) F99L, and d) F99A vulval region (arrowhead marking vulva opening) highlighting the Ret phenotype (bracketed stacked embryos) observed only in the F99A animals.

**Table 1.** CRISPR/Cas9 HLH-8 mutants generated in this study.

Strain name	Mutation made	Codon change	Reference name used in text
AK183: <i>hlh-8</i> (th11) Leu95Ala	Leu95Ala	tta→gcc	L95A
AK188: <i>hlh-8</i> (th12) Phe99Ala	Phe99Ala	ttc→gcg	F99A
AK193: <i>hlh-8</i> (th13) Arg103Ala	Arg103Ala	agg→gcg	R103A
AK198: <i>hlh-8</i> (th14) Leu95Ala; Phe99A; Arg103A	Leu95Ala; Phe99A; Arg103A	tta→gcc + ttc→gcg + agg→gcg	L95A; F99A; R103A (or Triple A)
AK203: <i>hlh-8</i> (th15) Leu95Phe; Phe99Leu	Leu95Phe; Phe99Leu	tta→ttc + ttc→ctt	L95F; F99L
AK208: <i>hlh-8</i> (th16) $\Delta$ Leu95-R103	$\Delta 95$ –103	27 bp in-frame deletion	$\Delta 95$ –103
AK213: <i>hlh-8</i> (th17) Leu95Phe	Leu95Phe	tta→ttc	L95F
AK219: <i>hlh-8</i> (th18) Phe99Leu	Phe99Leu	ttc→ctt	F99L
AK226: <i>hlh-8</i> (th19) Arg103Met	Arg103Met	agg→atg	R103M
AK242: <i>hlh-8</i> (th20) fs101	NHEJ, 87 bp $\Delta$ after Asn100	Frameshift/read-through into 3rd intron	fs101

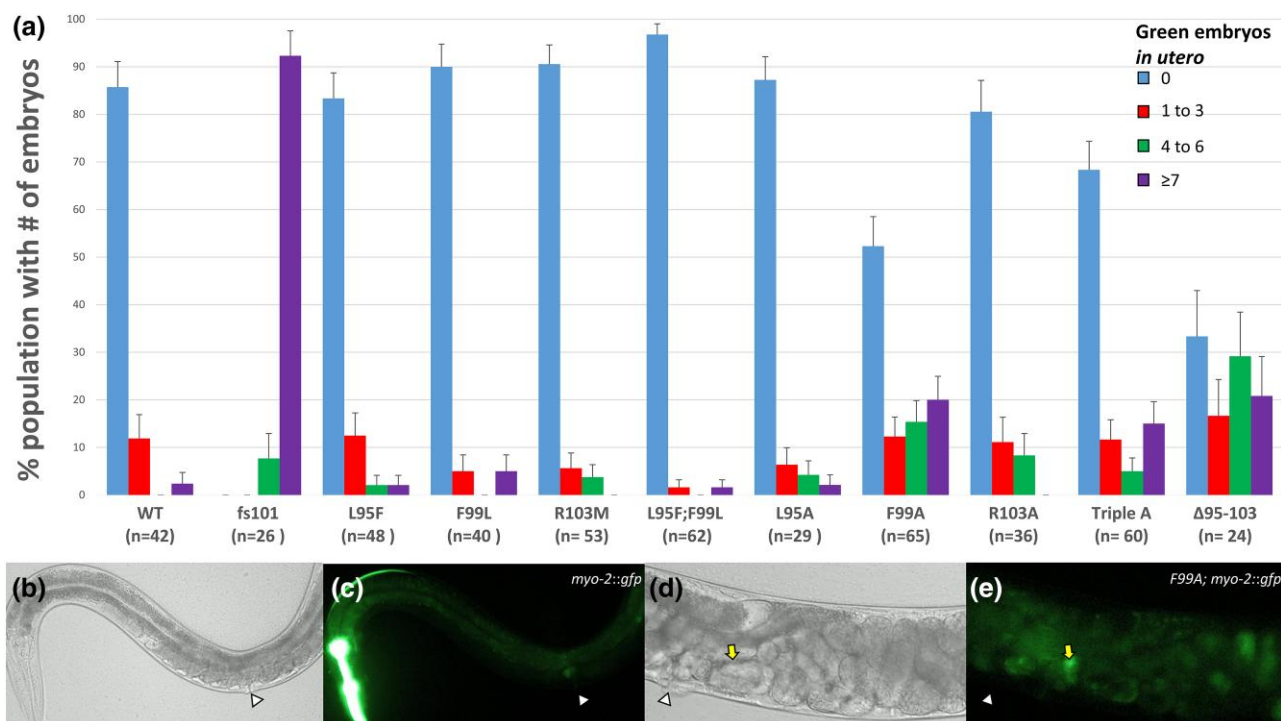
RoseTTAFold HLH-8 structural model predicts the 9 amino acids of the putative LXXXXXXR TWIST-Box form an  $\alpha$ -helix (Fig. 1a). If that 9 amino acid region is removed from the protein in silico (Fig. 1b), the  $\alpha$ -helix is lost entirely but the rest of the structure remains fairly similar to WT. In the TWIST1 studies, reporter gene expression was lost when this region was deleted and also when the 3 conserved residues were mutated to alanines (Laursen et al. 2007). To see how the HLH-8 putative domain would be altered when the corresponding alanine substitutions were made to the 3 highly conserved amino acids, we modeled a triple A mutant and, as expected, found the  $\alpha$ -helix to be disrupted (Fig. 1c). The protein modeling indicated that not only would these 9 amino acids in HLH-8 have the potential to be an  $\alpha$ -helical domain, as in TWIST1, but deleting the domain or making a triple A mutant would be predicted to have similar phenotypes.

### C. elegans HLH-8 TWIST-Box domain deletion animals retain embryos

Although the amino acid identity of the region of HLH-8 that most closely resembled the TWIST1 TWIST-Box is low (24%), the identity and spacing of the 3 critical amino acids LXXXXXXR were conserved (Fig. 2, a and b, Fig. 3a). In order to determine whether this potential domain was required for transcriptional activity, we utilized CRISPR/Cas9 gene editing to delete the domain in the genome (Fig. 2b). The deleted 9 amino acid region of HLH-8 starts at

Leu95 and ends at Arg103 to create the strain  $\Delta 95$ –103 and could be distinguished from the WT allele using a forward PCR primer containing the base pairs before/after the deleted sequence (Fig. 2c). See Table 1 for strain name abbreviations used throughout this paper. In addition to generating the  $\Delta 95$ –103 strain via homology directed repair (HDR), we also obtained an interesting non-homologous end-joining (NHEJ) mutant allele in the same region of the genome. DNA sequencing revealed that the NHEJ mutant was a frameshift mutation that occurred after the Asp100 (referred to here as fs101) and would be expected to preserve part of the TWIST-Box domain but would be missing the conserved Arg103. We modeled this mutant protein using the Robetta RoseTTAFold method and found that the first part of the domain is still predicted to form an alpha helix that is longer but contains amino acids not found in HLH-8 due to the frameshift (Fig. 1d). The fs101 allele was of interest since it had a phenotype that was distinct from either a null allele, *hlh-8* (*Δ*), or the  $\Delta 95$ –103 animals.

Previous work has shown that *hlh-8* (*Δ*) animals are egg-laying defective (Egl) and constipated (Con) (Corsi et al. 2000). We predicted that if the TWIST-Box domain was necessary for HLH-8 transcriptional activity, the  $\Delta 95$ –103 mutant animals also should be Egl and Con. However, both the  $\Delta 95$ –103 and fs101 worms do not phenocopy the *hlh-8* (*Δ*) worms. Instead they retain embryos (Ret), which is a less severe form of Egl (compare Fig. 2d to 2, e



**Fig. 4.** The F99A, Triple A, and  $\Delta 95-103$  missense alleles display the Ret phenotype. a) The percentage of animals in the population that retain embryos in the uterus 24 h post-L4 stage adults. The *myo-2::gfp* pharyngeal marker was crossed into each TWIST-Box mutant allele, and the animals were observed for embryos expressing the GFP to indicate that they have developed past the time that they are normally laid out in WT hermaphrodites. Animals with a Ret phenotype have more than 4 embryos in the uterus with a green pharynx (green or purple bars). The fs101 mutants have nearly 100% of the population that are Ret. b, c) WT (N2) animals typically do not have any pharyngeal GFP expression in the single row of embryos in the uterus seen in the Nomarski image b) and with GFP c). d, e) F99A TWIST-Box mutant animals also are Ret. Note the arrow pointing to one of the *myo-2::gfp* expressing embryos in the uterus. The arrowheads point to the vulva opening. Error bars are standard error of the proportion.

and f). The Ret phenotype is characterized by the presence of more than 1 row of later-staged tri-fold embryos within the hermaphrodite uterus, even though the animal continues to lay eggs. In contrast, wild-type hermaphrodites typically have a single row of embryos in the uterus. In addition, neither of these 2 mutant alleles have a noticeably enlarged intestinal lumen or Con phenotype. Therefore, the TWIST-Box domain appears to play an important role in proper HLH-8 function in the VMs that are important for egg-laying but not the ENTs that are needed to eliminate waste.

### Alanine substitutions reveal that Phe99 is required for some HLH-8 function leading to a Ret phenotype

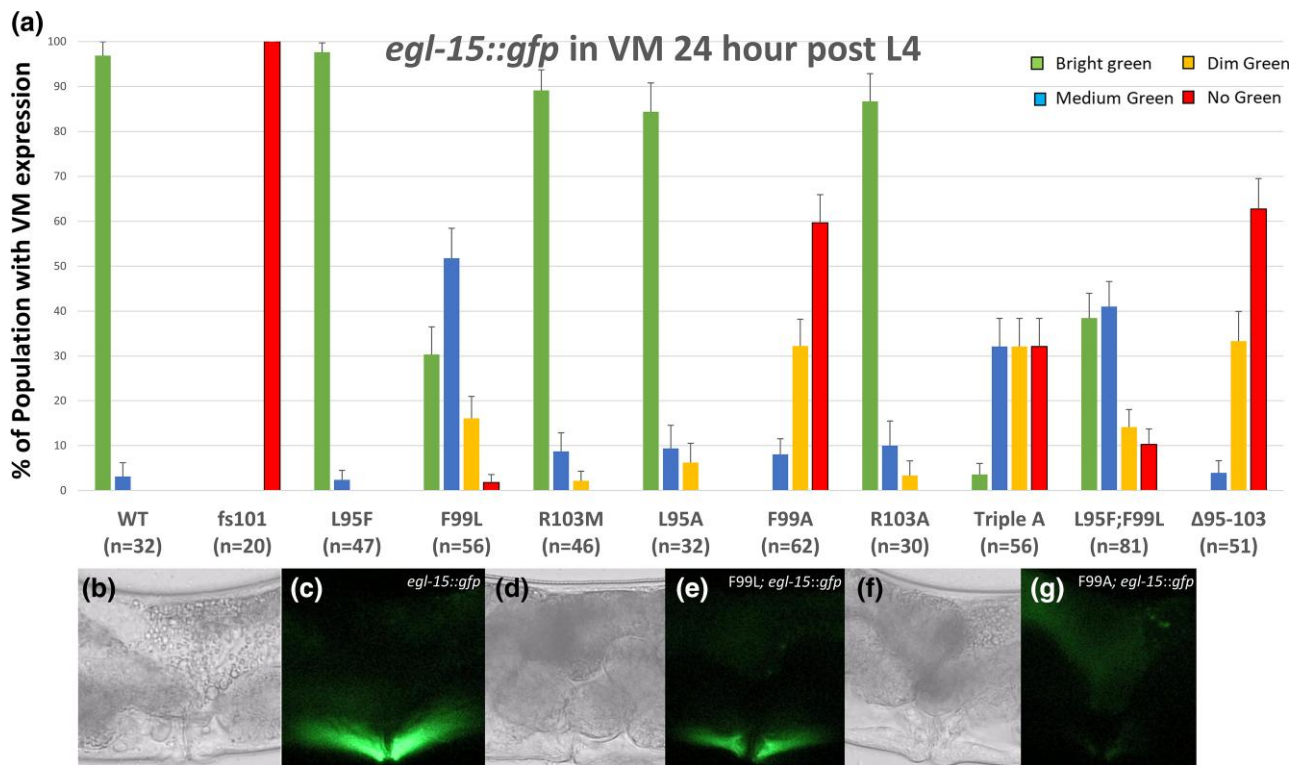
Since the HLH-8 TWIST-Box domain is playing an important role in VM development and the conservation and equidistant location of the 3 amino acids suggest they may be important for the domain function, individual alanine substitutions were used to probe their function in HLH-8 (Fig. 3a). We generated these alanine missense mutations in the genome utilizing CRISPR-Cas9 and HDR (Supplementary Fig. 1) to make L95A, F99A, and R103A strains (Table 1). In addition, a triple alanine mutant (L95A F99A R103A), or Triple A, was generated to test whether disrupting the 3 amino acids would phenocopy the  $\Delta 95-103$  animals. The L95A and R103A alleles appeared WT and did not display a Ret phenotype. In contrast, the F99A and the Triple A alleles were Ret, similar to the  $\Delta 95-103$  animals (Fig. 3d). Therefore, Phe99 is likely to play a more pivotal role than the Leu95 or the Arg103 in HLH-8 for the proper formation and function of the VMs.

### Saethre–Chotzen syndrome TWIST-Box mutations have milder phenotypes compared to alanine mutations at the analogous positions in HLH-8

In previous clinical studies, 2 patients were identified with mutations in 2 of the 3 conserved amino acids in the TWIST-Box: Phe187Leu (Kress et al. 2006) and Arg191Met (Peña et al. 2010). These patients displayed typical SCS phenotypes, including coronal suture synostosis. Interestingly, neither of the patients had syndactyly or webbing of digits/toes found in other SCS patients. We predicted that making the analogous mutations in HLH-8 would result in similar phenotypes to the alanine substitutions if they were similarly deleterious or would have milder phenotypes if they retained more function. The same CRISPR/Cas9 strategy was used to make F99L and R103M patient mutations in HLH-8. In addition, 2 additional mutants were generated. The first was an L95F substitution. We predicted that having 2 phenylalanines in close proximity on the same face of an alpha helix may be disruptive for important interactions with adjacent amino acids. Finally, to test whether having all 3 amino acids (Leu, Phe, and Arg) is sufficient for proper HLH-8 function regardless of the position in the domain, we constructed a double mutant L95F; F99L to change the domain from LXXXFXXXR to FXXXLXXXR. We established all 4 mutant strains, confirmed them by sequencing, and found no visible defective phenotypes in any of these strains.

### Average brood size of the TWIST-Box mutants is similar to WT animals

Since the Ret phenotype of the TWIST-Box domain mutants might be due to subtle VM defects, we measured brood size in all of our



**Fig. 5.** The fs101, F99A, Triple A, and  $\Delta$ 95–103 mutants are defective for *egl-15::gfp* expression. a) The expression of the *hlh-8* target gene reporter, *egl-15::gfp*, was observed in TWIST-Box mutant populations that were categorized for their GFP intensity in the VMs: bright green [green bars in a), micrograph in c)], medium green [blue bars in a) micrograph in e)], dim green [yellow bars in a) micrograph in g)], or no green [red bars in a) micrograph not shown] in the VMs. b, d, f) Nomarski and c, e, g) GFP images of the vulval region in b, c) WT, d, e) F99L and f, g) F99A that have an *egl-15::gfp* multi-copy integrated reporter in each strain. All animals were scored 24 h post-L4 stage and the GFP images were all taken with the same exposure time (39 ms). Error bars are the standard error of the proportion.

mutants (Gruss and Corsi 2022). We expected a decrease in total brood size in any of the animals that displayed a Ret phenotype (fs101,  $\Delta$ 95–103, F99A, Triple A). However, only the fs101 animals had a significant decrease in brood size (approximately 70–80 total progeny) compared to the WT animals that laid an average of 250 embryos (Supplementary Fig. 2). The fs101 animals were capable of laying embryos at a reduced rate early on, but as the worms aged they eventually become completely Egl leading to fewer offspring. Therefore, although the VMs appear to be disrupted enough to cause a Ret phenotype in the  $\Delta$ 95–103, F99A, and Triple A mutant alleles, there seems to be enough remaining VM function to expel embryos at a reasonable rate to still have a WT brood size.

### The fs101, F99A, Triple A, and $\Delta$ 95–103 alleles are the only TWIST-Box alleles that retain embryos

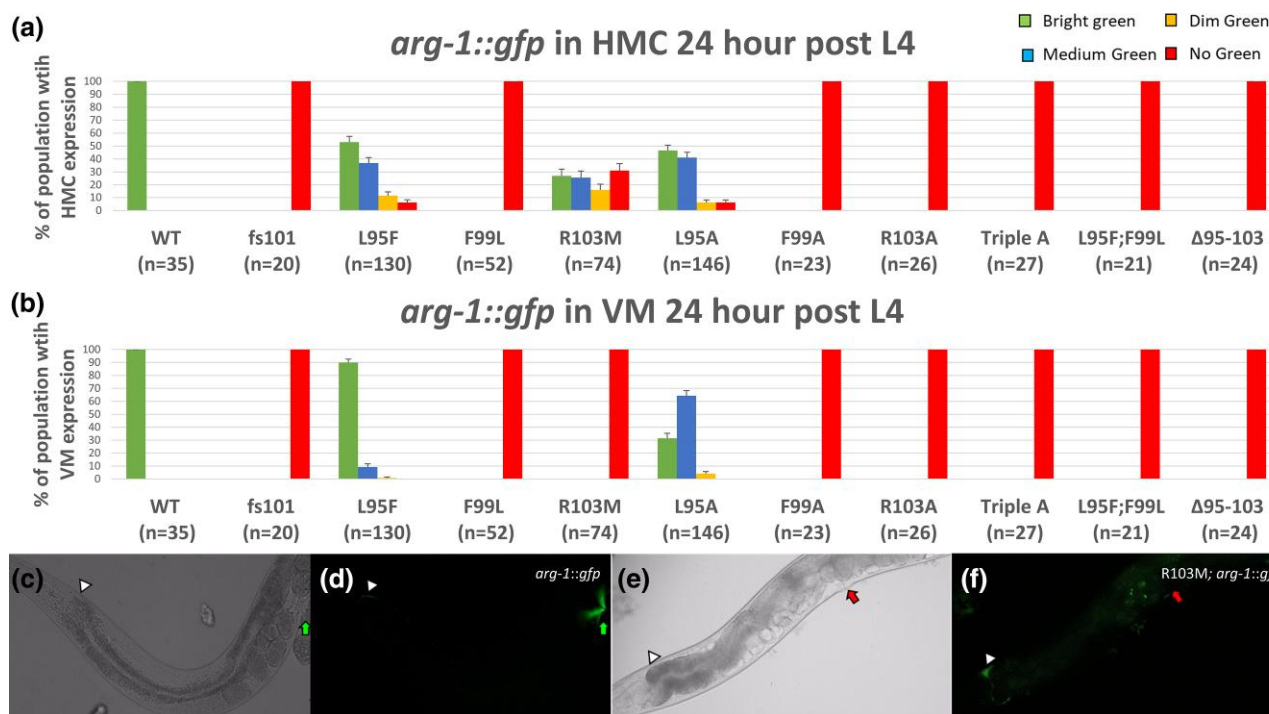
Since the brood size did not differ significantly in any of the TWIST-Box mutants compared to WT animals, an alternative approach was taken to determine the degree of penetrance of the Ret phenotype. We used a bright pharyngeal marker (*myo-2::gfp*) that begins to express in later embryogenesis, at which point the embryo should already have been expelled from the parent, as an indicator that embryos were being retained (Gruss and Corsi 2022). The fs101 strain had the most severe phenotype with 100% of the animals containing at least 4 or more of the late-stage embryos in the hermaphrodite uterus (Fig. 4). Further, the F99A, Triple A, and  $\Delta$ 95–103 animals had a less severe Ret phenotype with 30–50% of the hermaphrodites having 4 or more late-stage embryos and 10–20% of them retaining at least 1 late-stage embryo (Fig. 4). Finally, as expected based on our initial observations, none of the other

HLH-8 TWIST-Box mutants had any appreciable retention of embryos in this assay (Fig. 4). Therefore, the fs101 strain is likely to be more defective in VM development compared to the F99A, Triple A, and  $\Delta$ 95–103 alleles, which likely have similar partial defects in VM development since they all have similar phenotypes in this assay.

### Characterization of postembryonic M-cell lineage indicates defects in animals that retain embryos

To further assess the role of the HLH-8 TWIST-Box mutant alleles in the development of the VMs, we examined the cellular lineage from which these muscles are derived. The VMs are descendants of the M mesoblast, which is born in the embryo and divides during larval development (Sulston and Horvitz 1977; Sulston and White 1980). In order to follow the M cell divisions, we used an *hlh-8::gfp* transcriptional reporter. In wild-type animals, M divides dorso-ventrally at the L1 stage (Supplementary Fig. 3, a and b); at the L3 stage, the GFP reporter is expressed in 2 sex myoblast (SM) cells (Supplementary Fig. 3, e and f) that migrate and proliferate in the region of the developing vulva to form the 8 vulval and 8 uterine muscles from the 16 SM descendants (SMDs) (Supplementary Fig. 3, i and j). We looked for alterations of the expected division planes or number or size of descendant. These defects were seen in the Ret alleles at low penetrance, F99A, Triple A, and  $\Delta$ 95–103 had defects in approximately  $\leq$ 5% of the population, and the fs101 in closer to 20% of the population with defects including more than 2 SMs or 16 SMD that were abnormally located on the dorsal side of the animal (Supplementary Fig. 3, g and h, k-n). Additionally, the L95F;F99L double mutant allele displayed defects at the L1 (Supplementary Fig. 3 c and d), L3, and L4 stages





**Fig. 6.** HLH-8 TWIST-Box is required for *arg-1::gfp* expression in the VMs and HMC with F99 and R103 playing a critical role in both tissues. a, b) The *hlh-8 arg-1::gfp* target gene reporter was examined in each tissue of the TWIST-Box mutants. The animals were categorized according to the intensity of GFP expression: bright green (green bars), medium green (blue bars), dim green (yellow bars), or no green (red bars) for each cell type: a) head mesodermal cell (HMC) or b) vulval muscles (VMs). c, d) WT hermaphrodites express bright green *arg-1::gfp* in the HMC (arrowhead) and VMs (arrow). e, f) A tissue-specific defect in the R103M animals is observed with bright green HMC expression (arrowhead) and no expression in the VMs (arrow). All animals were scored 24 h post-L4 stage and GFP images were all taken with the same exposure time (39 ms). Error bars are the standard error of the proportion.

similar to the Ret alleles even though these worms are not Ret and neither the L95F nor the F99L single alleles displayed any defect in the M lineage. Finally, in the L95A mutants, although we did not find any M lineage defects, the animals had bright background GFP, resembling auto-fluorescent lipofuscin granules, at levels none of the other alleles, WT animals, nor the L95A allele without any GFP reporters express (Supplementary Fig. 3, o and p); we did not explore this phenotype any further. Overall, the low penetrance of the Ret phenotype found in the F99A, Triple A, and the Δ95-103 TWIST-Box domain alleles seems to be consistent with the low penetrance of M cell descendant defects, whereas the increased penetrance of the M lineage defects in the fs101 allele correlates with the more defective Ret phenotype.

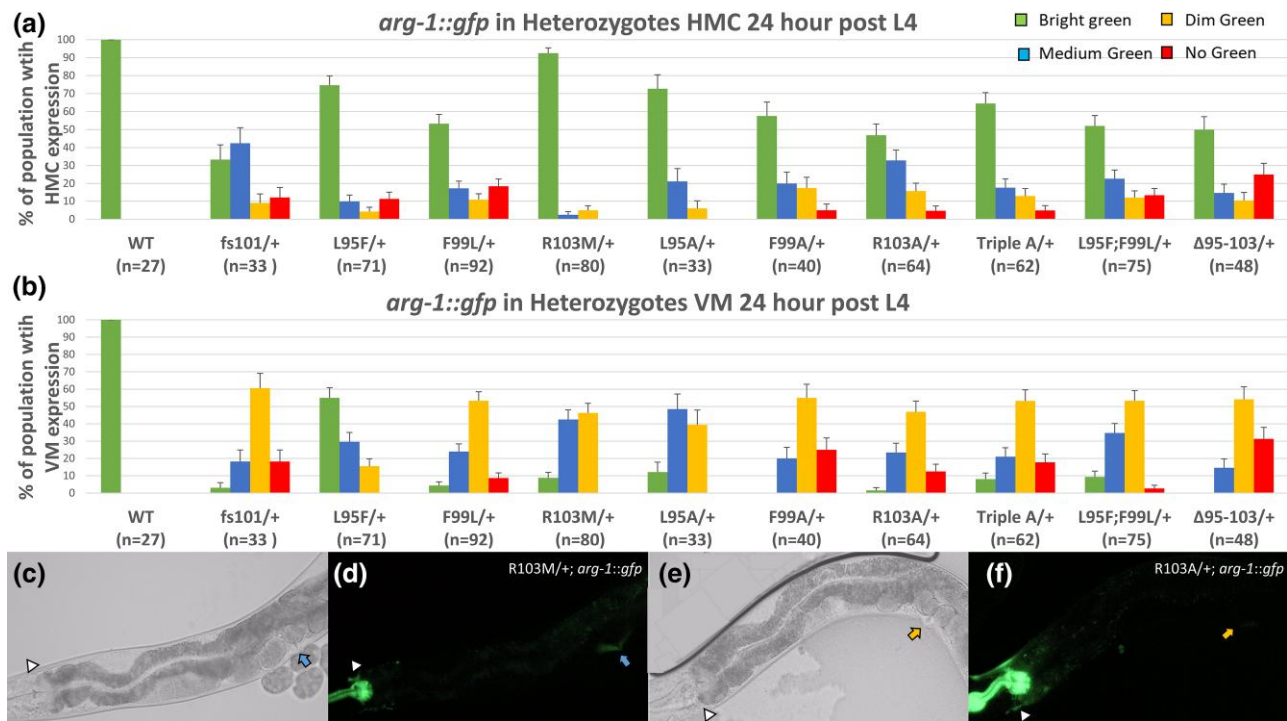
### The F99 TWIST-Box residue plays a critical role in the transcriptional activation of an HLH-8 target gene, *egl-15*

In order to test whether the TWIST-Box domain plays a role in transcriptional activation of target genes by HLH-8, several GFP transgenic reporters were crossed into each of the mutants to observe levels of expression. We first examined the multi-copy, integrated transgene *egl-15::gfp*. The *egl-15* gene encodes an FGF receptor and the null mutant results in worms that are Egl due to the improper migration of the SMs to the developing vulva (Stern and Horvitz 1991). In WT animals, *egl-15::gfp* expresses in the 4 VM1 cells (Fig. 5b, c) (Harfe et al. 1998). Because the *egl-15::gfp* transgene is integrated into the genome in multiple copies, it has very bright expression in all 4 VM1 in all animals in a wild-type population (Harfe et al. 1998). When observing the TWIST-Box mutants containing this reporter, we considered

the overall expression in all the cells and categorized each animal as either having bright (WT), medium (clearly distinct from WT), dim (barely see the outline of the VM cells), or no expression (refer to Fig. 8 for a detailed summary of all the mutant phenotypes). The most severe phenotype was found in fs101 animals, which did not express *egl-15::gfp* in any of the animals, which is a phenotype similar to what was observed previously in *hlh-8(A)* animals (Corsi et al. 2000) (Fig. 5). In addition, it was observed that the strains with the Ret phenotype (F99A, Triple A, and Δ95-103) had defects similar to each other in *egl-15::gfp* expression where 60-90% of them displayed dim or undetected levels of GFP expression in the VMs (Fig. 5a, f, g). We also found that the F99L SCS patient allele displayed a milder defect in *egl-15::gfp* expression compared to the F99A mutant. Approximately 80% of the F99L animals expressed the GFP at bright or medium levels compared to less than 10% of the F99A animals that expressed the GFP at similar levels (Fig. 5d, e) suggesting that leucine retained more function than alanine at that position in the protein. All other mutations, including the R103M SCS patient mutation, had *egl-15::gfp* expression levels that were comparable to the bright green expression in the VMs of WT animals. Therefore, the L95 and R103 residues seem to have a smaller role in HLH-8's ability to regulate *egl-15* expression whereas the F99 position seems to play a more critical role.

### F99 is also required for HLH-8 target gene *arg-1* expression

To further characterize the HLH-8 potential TWIST-box domain, we examined a second target gene transcriptional reporter, *arg-1::gfp* (Zhao et al. 2007). The *arg-1* gene encodes a ligand of the Delta/Serrate/LAG-2 (DSL) family that functions in the



**Fig. 7.** TWIST-Box mutants that were defective for *arg-1* expression as homozygotes are semi-dominant as heterozygotes mainly in the VMs. a, b) The HLH-8 *arg-1::gfp* target gene reporter was examined in each tissue of heterozygous TWIST-Box mutants. The animals were categorized according to the intensity of GFP expression (see Fig. 6) in the a) head mesodermal cell (HMC) or b) vulval muscles (VMs). c, d) R103M/+ animals express bright green *arg-1::gfp* in the HMC (arrowhead) and medium green VMs (arrow). e, f) R103A/+ animals express bright green HMC expression (arrowhead) yet only have dim green expression in the VMs (arrow). The animals are also expressing *myo-2::gfp* in the anterior (left) that was used as a marker for heterozygotes. All animals were scored 24 h post-L4 stage and GFP images were all taken with the same exposure time (39 ms). Error bars are the standard error of the proportion.

Delta/Notch signaling pathway, which is important for cell-fate specification (Mello et al. 1994; Fitzgerald and Greenwald 1995; Chen et al. 2004). A null mutation in *arg-1* does not lead to a visible phenotype; therefore, *arg-1* serves a redundant role in the Delta/Notch signaling pathway (McGovern et al. 2009). The *arg-1::gfp* is expressed in several tissues: the head mesodermal cell (HMC), the VMs, and the enteric muscles (ENTs) (Fig. 6d; Supplementary Fig. 5B and 6B). GFP expression is lost in all these tissues in the *hlh-8* ( $\Delta$ ) strain since HLH-8 is required for *arg-1* expression (Corsi et al. 2000). In homozygous strains of each mutant TWIST-box allele containing the integrated, multi-copy *arg-1::gfp* reporter, tissues were scored using the same methodology as the *egl-15::gfp* reporter (Fig. 5). When the TWIST-Box domain is removed ( $\Delta$ 95–103), the level of GFP expression is completely lost in all 3 tissues (VMs and HMC, Fig. 6, a and b; and ENTs, Supplementary Fig. 5). This severe phenotype was also seen in the fs101, F99A, and the Triple A mutant worms. In contrast to the reasonably robust *egl-15::gfp* expression in the F99L patient mutant as well as the L95F; F99L double mutant, *arg-1::gfp* expression was undetectable in all 3 tissues even though these mutants are not Ret (Fig. 6, a and b; Supplementary Fig. 5). The common factor in all of the strains that lost *arg-1::gfp* expression is that none of them had the wild-type F99 residue present (with the exception of the frameshift allele fs101) suggesting that this critically conserved amino acid is required for *arg-1* expression in all tissues.

### The R103 TWIST-Box residue plays a critical role in *arg-1::gfp* expression in the VMs and SPH muscle

To test whether R103 was playing a role in *arg-1* expression, the GFP reporter was examined in mutant animals. We found that the R103A

substitution led to a lack of *arg-1::gfp* expression in all tissues (Fig. 6, Supplementary Fig. 5), whereas the analogous SCS patient mutation, R103M, led to a lack of expression in the VMs and the anal sphincter (SPH) but had more mosaic expression in the HMC, the anal depressor (DEP), and the intestinal muscles (INTs) (Fig. 6, Supplementary Fig. 5). This residue's defects were tissue-specific, depending on the amino acid substitution leading to the conclusion that R103 is required for proper expression of *arg-1* in the VMs and SPH, but this amino acid has a less important function in the HMC, INTs, and DEP.

### The L95 TWIST-Box residue plays a minor role in *arg-1::gfp* expression in the VMs, INTs, and DEP

The final conserved TWIST-Box domain residue, L95, also was examined for *arg-1* expression. Unlike the F99 or R103 mutants, *arg-1* expression in the VMs was minimally effected in the L95 mutants (Fig. 6). The L95F allele resembled the WT HLH-8 and the L95A allele only displayed minor VM *arg-1* expression defects (Fig. 6). In examining the enteric muscles in L95A and L95F animals, we found that SPH expression was largely WT expressing well in 80–90% of animals whereas the INTs and DEP expression was partially defective expressing well in about 50% of the animals (Supplementary Figs. 5 and 6). Therefore, the L95 plays a more minor role compared to F99 and R103 in HLH-8 transcriptional activation of *arg-1* in the VM, INTs, and DEP cells.

### HLH-8 TWIST-Box domain heterozygous mutant animals show minor defects in *arg-1* expression

In order to determine whether the TWIST-Box mutant alleles had any dominant defects, heterozygous strains were examined for

STRAIN	egl-15::gfp		arg-1::gfp				Retentive (Ret)	M cell lineage defect?
	VM	VM	HMC	INT	SPH	DEP		
WT	+++	+++	+++	+++	+++	+++	N	N
L95F	+++	+++	+	+	+++	+	N	N
F99L	+/-	-	-	-	-	-	N	N
R103M	+++	-	+/-	+	-	+/-	N	N
L95A	+++	+/-	+	+/-	+++	+/-	N	Ectopic Lipofuscins*
F99A	-	-	-	-	-	-	Y	L1, L3, L4
R103A	+++	-	-	-	-	-	N	N
Triple A	-	-	-	-	-	-	Y	L1, L3, L4
L95F;F99L	++	-	-	-	-	-	N	L1, L3, L4
Δ95-103	-	-	-	-	-	-	Y	L1, L3, L4
fs101	-	-	-	-	-	-	Y	L1, L3, L4
L95F/+	ND	+	++	+/-	+++	+	N	N/A
F99L/+	+++	-	+	-	++	-	N	N/A
R103M/+	ND	-	+++	++	+++	+	N	N/A
L95A/+	ND	-	++	+/-	+++	+	N	N/A
F99A/+	+++	-	+	+/-	+++	+/-	N	N/A
R103A/+	ND	-	+	-	+++	+/-	N	N/A
TripleA/+	+++	-	++	-	++	+/-	N	N/A
L95F;F99L/+	+++	-	+	-	+++	+/-	N	N/A
Δ95-103/+	+++	-	+	-	++	+/-	N	N/A
fs101/+	+++	-	+/-	-	+++	-	N	N/A

**+ and - indicates % of the population with the WT phenotype (bright GFP signal)**  
 +++, [80–100%]; ++ [60–80%]; +, [40–60%]; +/-, [20–40%]; -, [0–20%]

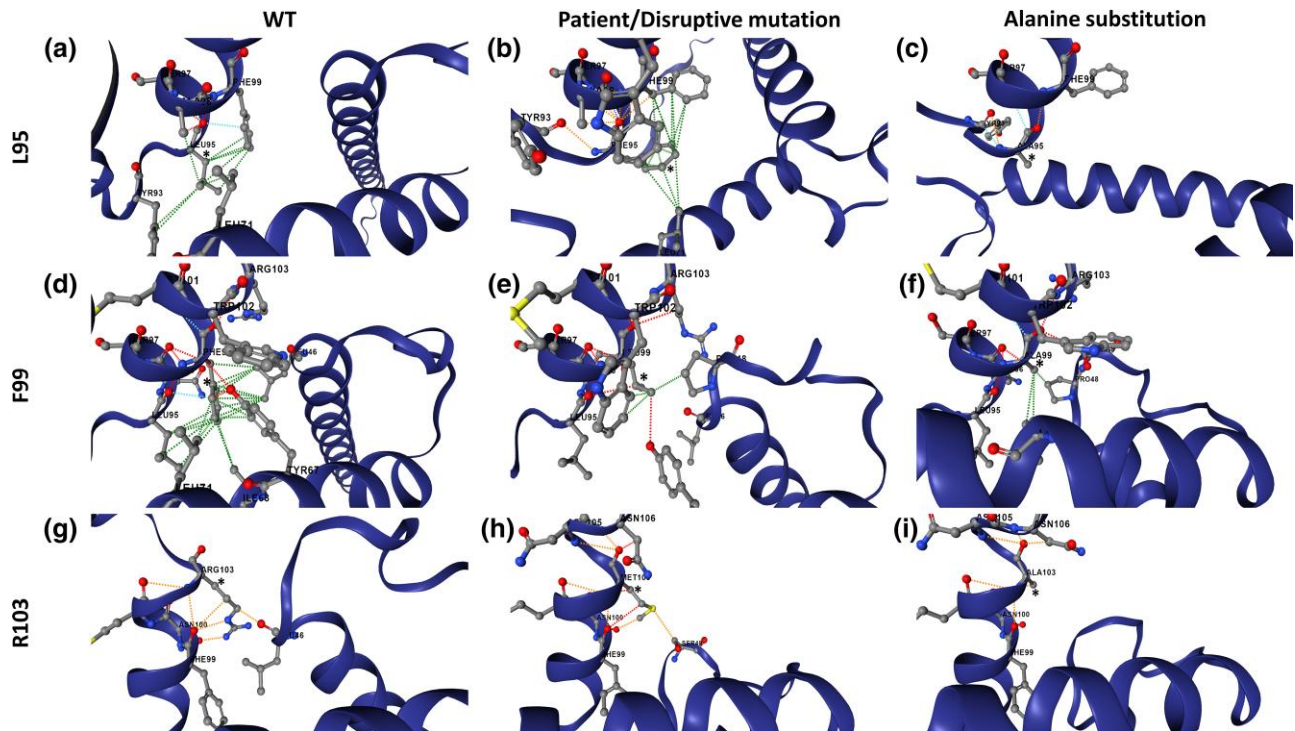
**Fig. 8.** Summary of phenotypes of *C. elegans* *hlh-8* TWIST-Box alleles. The phenotypes examined for the TWIST-Box mutant alleles in comparison to the control N2 (WT) strain are indicated in each row. The first column has the strain names: either TWIST-Box homozygous or heterozygous mutants. The subsequent columns indicate either the GFP expression pattern or Retentive or M lineage defects. The GFP expression is reported according to the legend below the table. The Ret phenotype is either present (Y) or absent (N) in the mutants. The M lineage defects are absent (N) or present at a specific stage (L1, L3, or L4) or not analyzed (N/A). For the *egl-15::gfp* heterozygotes, strains without defects as homozygotes were not analyzed as heterozygotes (ND).

embryo retention (Ret) and *egl-15* or *arg-1* GFP expression. In all alleles that were Ret as homozygotes, they did not retain embryos as heterozygotes (data not shown). In addition, any strain that was defective in the expression of the *egl-15::gfp* reporter as a homozygote (F99A, F99L, L95F; F99L, Triple A, Δ95–103, and fs101) expressed a WT level of *egl-15::gfp* reporter as a heterozygote (Supplementary Fig. 4). On the other hand, the TWIST-Box/+; *arg-1::gfp* mutant alleles appear to have variable expressivity with tissue-specific semi-dominance (Fig. 7; Supplementary Fig. 7). For all TWIST-Box/+; *arg-1::gfp* alleles, *arg-1* expression in the HMC (Fig. 7a) had some minor defects with <30% of the population having dim or no green expression and in the SPH (Supplementary Fig. 7F), the GFP was comparable to WT levels. Whereas in the VMs (Fig. 7b), INTs (Supplementary Fig. 7E), and the DEP (Supplementary Fig. 7G), the GFP expression is decreased or absent in more than 50% of the population in most of the alleles. Further, many of the alleles vary from animal to animal in the penetrance of the heterozygous phenotype. For example, whereas the R103A homozygotes fail to express *arg-1::gfp* (Fig. 6, a and b, Supplementary Fig. 6), R103A/+ heterozygotes vary from animal to animal with dim VMs and bright INTs in 1 animal and brighter VMs and SPH but lack of expression in the DEP and INTs with no clear pattern among the animals (compare Supplementary Fig. 7C and 7D). Therefore, the TWIST-Box mutant alleles are not recessive, but behave more like semi-dominant, allele-specific, and tissue-specific alleles with regard to HLH-8 transcriptional activation of *arg-1*.

## Discussion

Understanding how a sequence of amino acids can form a structural domain with specific functional roles has been aided by advancing strategies of computational protein modeling. However,

protein modeling does not reveal how a domain's structure contributes to protein function in vivo. This study provides an example of predicting a domain's structure and mutational consequences through protein modeling and then engineering those changes in a model organism to probe the structure/function of the domain. This strategy was used to investigate a potential domain contributing to the transcriptional activity of the *C. elegans* Twist homolog, HLH-8 (see Fig. 8 for a summary of all the mutants and phenotypes observed). Deletion of the evolutionarily conserved LXXXFXXXR motif (Δ95–103 allele) demonstrated the domain is important, but not absolutely necessary, for HLH-8 function since the worms were Ret and failed to fully transcriptionally activate downstream target genes in contrast to HLH-8 null animals that are 100% Egl and do not activate downstream target genes at all (Corsi et al. 2000). Adding to previous TWIST-Box studies in other organisms, our *C. elegans* experiments revealed (1) the evolutionarily conserved LXXXFXXXR residues of the TWIST-Box display a hierarchy of functional importance with F99 > R103 > L95; (2) the position of the first and second conserved amino acids is significant, i.e. the F and L cannot be swapped without phenotypic consequences; (3) the TWIST-Box and the LXXXFXXXR residues exhibit tissue and target gene requirements for proper spatio-temporal function of HLH-8; and (4) the phenotypic differences between HLH-8 TWIST-Box alanine and analogous patient mutations provide important structure/function implications for TWIST1 and SCS patients (Fig. 9). In contrast to recessive *hlh-8(Δ)* alleles (Corsi et al. 2000), the phenotypes of the HLH-8 TWIST-Box mutants are semi-dominant, although they are distinct compared with E29 DNA binding basic domain alleles that were previously studied (Corsi et al. 2002; Kim et al. 2017). The milder phenotypes of the HLH-8 TWIST-Box mutants compared to HLH-8 null animals indicate that without the TWIST-Box,



**Fig. 9.** HLH-8 TWIST-Box missense mutations could alter interactions between neighboring residues impacting the overall and tissue-specific structure/function of HLH-8. a–i) Dynamut2-generated models of interactions with neighboring residues of a–c) L95, d–f) F99, and g–i) R103 in HLH-8. The residue of interest is marked with an asterisk (\*) and all models were generated utilizing a Robetta, allele-specific template (e.g. L95F utilized a L95F RoseTTAFold template) then uploaded into Dynamut2 software (see Methods). Each column represents the HLH-8 template type, WT (identical template for a, d, and g), patient (F99L in e and R103M in h) or predicted disruptive allele (L95F in b), and the single alanine mutations (in c, f, and i). Each line between amino acids represents a type of molecular interaction based on the color (red-hydrogen bonds; green-hydrophobic interactions; light blue-VDW; and orange-polar interactions).

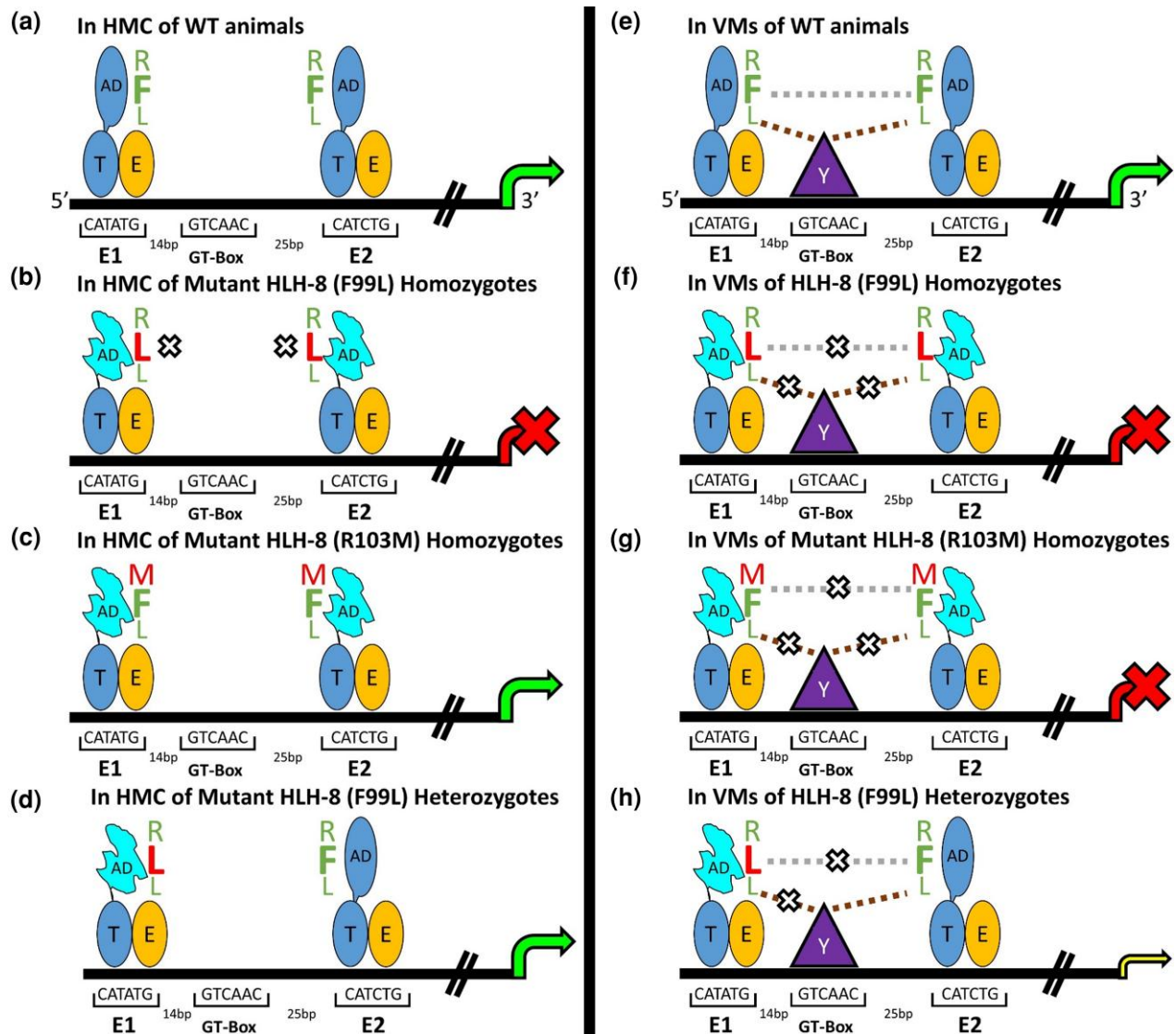
HLH-8 can still bind to DNA and activate some target genes (*egl-15*) but not others (*arg-1*) suggesting it could be a region important for recruiting distinct gene-dependent co-factors.

### HLH-8 has a TWIST-Box with evolutionarily-conserved LXXFXXXR residues that have a hierarchical role in the domain function

In order to determine whether HLH-8 had a functional TWIST-Box, CRISPR/Cas9 was utilized to create a 9 amino acid deletion ( $\Delta 95-103$ ) of the LXXFXXXR evolutionarily conserved motif (Laursen et al. 2007). Deletion of this region resulted in the Ret phenotype, which is distinct from the *hlh-8(d)* Egl phenotype. The Ret phenotype is likely due to partially defective VMs that function just enough to lay eggs perhaps via pressure built up in the uterus, which is a weaker phenotype than Egl where the VMs are not functioning at all. Therefore, the TWIST-Box deletion allele is not a null but rather behaves as a hypomorphic allele. To determine the functional importance of the individual residues in this motif, missense mutations were made. As expected, the Triple A mutant was Ret, similar to the  $\Delta 95-103$  mutant, indicating that the 3 equidistant, conserved amino acids were critical for the domain function. Surprisingly, the F99A amino acid substitution alone could cause the Ret phenotype. This finding, along with the variation among the HLH-8 TWIST-Box mutant allele defects with respect to *egl-15::gfp* and/or *arg-1::gfp* expression, led us to hypothesize that the conserved amino acids were not contributing equally to the domain function.

F99 (F187 in humans and F191 in mice) was the most critical of the 3 conserved TWIST-Box amino acids with the strongest

phenotypes when mutated. The analogous amino acid is required for TWIST1-induced EMT, invasion, anoikis resistance, and metastasis in cancer (Gajula et al. 2013) and was mutated in an ENU-induced screen in mice (F191S) leading to a dominant-negative protein causing hind limb polydactyly (Chen et al. 2020). Laursen and colleagues (2007) also generated an F191A substitution in mouse TWIST1, which they argued would not disrupt helix formation, yet they found decreased transcriptional activity at similar levels as the TWIST1 domain deletion. This outcome is consistent with our results where the HLH-8 TWIST-Box helix is still intact in F99A RoseTTAFold modeling (data not shown) yet the F99A mutant failed to transcriptionally activate HLH-8 target genes, which could suggest that the  $\alpha$ -helix itself is not sufficient for HLH-8 function. The F99 residue, at the center of the predicted  $\alpha$ -helix, may be required to interact with other amino acids through its hydrophobic and larger surface area properties. The Dynamut2 software provided a readout of the various types of intermolecular bonds formed from each modeled HLH-8 TWIST-Box allele (Fig. 9; Supplementary Fig. 8). For example, the WT F99 residue resulted in a large number (24) of hydrophobic interactions with nearby residues, including 5 of those with the conserved L95 (Fig. 9d green lines), as well as many interactions with Helix I (F99 WT with 10 hydrophobic L46 interactions) and Helix II (Y67, I68, L71). Interestingly, these interactions are almost completely lost in the F99L and F99A alleles (Fig. 9, e and f; Supplementary Fig. 8). This result could suggest F99 may have a critical function in intermolecular dynamic stabilization needed for full expression of the *arg-1::gfp* reporter. The less severe defects in *egl-15::gfp* expression in the F99L patient mutation



**Fig. 10.** A model for the HLH-8 TWIST-Box tissue-specific regulation of *arg-1*. a–h) The 385 bp minimal promoter of *arg-1* (solid horizontal line) contains 3 E-boxes [E1, E2, and E3 (not shown)] and an additional inverse E-box (GT-box), which play a role in tissue-specific expression (Zhao et al. 2007). HLH-8 (T ovals) is shown with either a WT TWIST-Box domain (AD ovals) or a mutated one (AD ruffled shapes). HLH-8 forms heterodimers with HLH-2 (E ovals) that are predicted to bind to E1 and E2. E3 is not shown since it plays a role in the ENTs, which are not discussed here. a) In a WT animal, strong expression of *arg-1::gfp* (curved arrow at 3' end) in the HMC is controlled by HLH-8/HLH-2 binding to both E1 and E2. b) In the SCS F99L patient allele homozygous mutants, the critical F99 (bold F next to the ADs in WT) of the HLH-8 TWIST-Box has been substituted with a leucine (bold L) thereby disrupting the ability of HLH-8 to promote transcription (curved X arrow at 3' end). c) However, in the R103M patient allele (M next to AD) or d) in the F99L/+ heterozygotes (1 L and 1 F next to AD), at least 1 of the critical F99 residues is unaltered and a HLH-8/HLH-2 heterodimer can bind and activate at E2. e–h) In the VMs, *arg-1* regulation has been proposed to require an ancillary factor (Y triangles) that binds to the GT-box ensuring maximum expression levels of *arg-1::gfp* (Zhao et al. 2007). e) The HLH-8 TWIST-Box in WT animals might associate with factor “Y” (dashed lines from AD to Y) or could interact with the HLH-8 TWIST-Box of another HLH-8/HLH-2 dimer if factor “Y” is a DNA architectural protein bringing the E1 and E2 E boxes into closer proximity to facilitate HLH-8 TWIST-Box driven (dashed lines between ADs) tetramer formation to drive *arg-1::gfp* expression in the VMs (based on Chang et al. 2015). f, g) In either the F99L or R103M SCS patient alleles, the VM expression of *arg-1::gfp* was completely lost (curved X arrow at 3' end) therefore the interaction with “Y” or the formation of tetramers is disrupted (Xs on dashed lines). h) In the HLH-8 TWIST-Box/+ heterozygous animals, expression of *arg-1::gfp* in the VMs (smaller arrow at 3' end) was more defective compared to the HMCs d) or ENTs (not shown). This expression difference could be because a single WT HLH-8 TWIST-Box may not be sufficient for proper protein-protein interaction with “Y” or for tetramer formation with either the HLH-8 (F99L) or HLH-8 (R103M) TWIST-Box SCS patient allele. In all models, the small font of L95 indicates the less critical role in the domain. The size of each amino acid indicates the functional hierarchy (F99 > R103 > L95) in HLH-8 TWIST-Box transcriptional activity.

compared to F99A suggest that the differential effects of the individual amino acid substitutions are important for this reporter gene.

R103 (R191 in humans, R195 in mice) was found to contribute a modest degree of function in HLH-8. In particular, the R103 mutant alleles displayed the most obvious case for the HLH-8 TWIST-Box playing an important role in tissue-specific and target

gene-specific expression due to differences in *arg-1::gfp* and *egl-15::gfp* activity. We found R103 mutant alleles affected *arg-1* expression primarily in the VMs. The VMs in both the R103A and R103M SCS patient allele appear morphologically normal by *egl-15::gfp* expression (Fig. 5) and function well enough to allow a normal brood size (Supplementary Fig. 2) and egg-laying rate (data not shown). We also observed that the R103 may be

contributing to tissue-specificity since R103A led to a lack of *arg-1::gfp* expression in all tissues, whereas the R103M still expressed the reporter in the HMC and the ENT muscles. In mammals, the R103 equivalent amino acid is required for TWIST-Box mediated inhibition of p53 in sarcomas (Piccinin et al. 2012). However, an R195G mouse mutant protein can still interact via the TWIST-Box with the master chondrogenic regulator Sox9 in fibroblasts whereas L187 and F191 cannot (Gu et al. 2012). The distinct activities of the 2 reporters in the R103 mutants suggest that the residue is contributing to distinct mechanisms of transcriptional regulation rather than tissue-specific degradation or nuclear localization, which should lead to a lack of either reporter being expressed in the VMs. From the Dynamut2 modeling results, the R103A allele is predicted to lack interactions with Helix I whereas the R103M SCS patient allele maintains at least some interaction with Helix 1, which may account for the milder phenotype in these animals and, perhaps, in patients as well (Fig. 9, g–i, Supplementary Fig. 8).

The HLH-8 L95 residue appears to be the least important for transcriptional function since the mutants have the mildest phenotypes. There is some influence on the *arg-1* promoter, specifically in the VMs and the INTs and DEP. As no SCS patient mutation has been reported at this position, an L95F allele was constructed since we predicted 2 bulky phenylalanines on the same face of an  $\alpha$ -helix may disrupt the domain function. The L95F allele most resembled WT for all observed phenotypes, which did not support our hypothesis. In contrast, the L95A allele had a greater loss in expression of *arg-1::gfp* in the VMs compared to L95F perhaps due to the loss of intermolecular hydrophobic interactions (Fig. 9, Supplementary Fig. 8).

Finally, to test whether the position of the conserved amino acids matter, we engineered the L95F; F99L mutant. The double mutant phenocopies the F99L SCS patient mutant with the exception of a low penetrance, *hlh-8::gfp* M mesoblast division phenotype that is absent in the individual mutants (Supplementary Fig. 3). These results are generally consistent with the F99 position being more important for function. In the future, other double mutants may be informative, such as L95A; R103A to see if the presence of the F99 would alleviate some of the defects seen in the Triple A. Maybe more intriguing, would be a L95R; R103L allele to ask the question whether the HLH-8 TWIST-Box is flexible as this allele would contain the 3 conserved amino acids but in the reverse order, RXXXFXXXL.

### The HLH-8 TWIST-Box must not be the only domain contributing to transcriptional activity

We predicted that mutating a domain important for transcriptional activation in HLH-8 would phenocopy the recessive *hlh-8 (Δ)* allele leading to Egl and Con phenotypes. However, the HLH-8 TWIST-Box mutants were neither Egl nor Con and had semi-dominant defects in tissue-specific regulation of *arg-1* in the VMs. Therefore, another part of the protein or the dimer partner must be contributing to the transcriptional activity in the mutants. Since the *egl-15::gfp* reporter was expressed fairly well in the TWIST-Box mutants (Fig. 5), the mutant proteins are not likely to be aberrantly degraded in the VMs. Although we have not been able to determine whether the mutants are expressed at the same level in each tissue where HLH-8 is found, mouse TWIST-Box mutants are expressed at levels equal to WT and also bind to DNA in EMSA experiments (Laursen et al. 2007; Chen et al. 2020). Further, a functional TWIST-Box is needed for protein degradation via polyubiquitination in *Xenopus* through a E3 ubiquitin ligase, or Ppa (Lander et al. 2011). The observation that the HLH-8 TWIST-Box mutants still can activate certain

target gene reporters leads to speculation that they may evade proper degradation and could disrupt the dimer pool in *C. elegans* leading to target gene specific rather than broader transcriptional defects. Further experimentation is needed to determine how HLH-8 is degraded and whether the HLH-8 TWIST-Box is ubiquitinated by a Ppa ortholog similar to *Xenopus*.

### hlh-8 heterozygotes reveal target gene-specific properties

The HLH-8 TWIST-Box mutants have distinct effects on *arg-1::gfp* and *egl-15::gfp* expression. Interestingly, the reporter gene expression is also distinct compared to *hlh-8 (Δ)* animals (Corsi et al. 2000) and amino acid substitutions in E29 in the DNA binding domain, particularly a dominant negative E29K allele (Kim et al. 2017). As homozygotes neither the *hlh-8 (Δ)* nor E29K can activate either reporter gene, whereas the HLH-8 TWIST-Box alleles are able to, albeit in a target and tissue-specific manner. As heterozygotes, the *hlh-8 (Δ)/+* animals have WT *arg-1::gfp* and *egl-15::gfp* expression. In contrast, E29K/+; *egl-15::gfp* heterozygous animals do not express the reporter in the VMs, whereas the E29K/+; *arg-1::gfp* animals had nearly WT expression in the HMC, ENTs, and the VMs (Kim et al. 2017). The HLH-8 TWIST-Box/+ mutants have the opposite effect on these *hlh-8* downstream reporter genes (Fig. 7, Supplementary Fig. 4 and 7). In HLH-8 animals that were defective as homozygotes, e.g. F99L; *arg-1::gfp*, the *arg-1::gfp* reporter expression had tissue-specific defects in the VMs in the TWIST-Box/+ heterozygotes. Altogether, the differences in expression of *arg-1* and *egl-15*, and likely other *hlh-8* downstream target genes, indicate that each promoter is subject to distinct regulation and that, in particular, *arg-1* in the VMs requires a fully functional HLH-8 TWIST-Box for expression (Fig. 10).

### The HLH-8 TWIST-Box may function distinctly in different tissues

The TWIST-Box does not appear to be a traditional AD required for all *hlh-8* activity, but only at certain promoters in distinct tissues with VMs being the most affected tissue. There are various possibilities for how the TWIST-Box could contribute to HLH-8 function. First, the TWIST-Box could influence the dimer preference of HLH-8 with itself or with its only known partner, HLH-2. This idea is consistent with *arg-1* regulation, which has 3 E boxes in the promoter that play distinct roles in gene expression with HLH-8 homodimers functioning in the ENTs and heterodimers in VMs and the HMC (Fig. 10; Zhao et al. 2007; Philogene et al. 2012). Second, the TWIST-Box could influence the overall structure/function of HLH-8 through intermolecular interactions that are distinctly stabilized in different tissue-specific environments, i.e. H-bonds, with other HLH-8 domains (Helix I and II) in order to play a role in internal stabilization. The Dynamut2 modeling is consistent with this idea (Fig. 9). Third, the TWIST-Box could participate in tetramer formation bringing together HLH-8-containing dimers that are in close proximity on the DNA. TWIST-Box driven tetramer formation has been discovered in humans and *Drosophila* (Chang et al. 2015). However, in these cases, the 2 E-boxes are separated by only 5 bps (Chang et al. 2015). Although the 2 E boxes (E1, E2) in the *arg-1* promoter that are responsible for VM expression are further apart (45 bp), they do have an intervening sequence (an inverted E box sequence called a GT box) that is needed for VM expression (Fig. 10; Zhao et al. 2007) and has been proposed to bind to a unknown factor (Y in Fig. 10), which could participate in TWIST-Box driven transcription in the VMs (Fig. 10) potentially by bending the DNA to bring the HLH-8 dimers closer together. Alternatively, Factor Y or the TWIST-Box

could bind to proteins at other unknown enhancers or to other proteins such as a kinase, acetyltransferase, or cofactors/scaffolding proteins (Wu et al. 2021) that all are known to influence transcriptional activation. A final, less likely, function for the HLH-8 TWIST-Box is that it could be interacting with general transcription factors (GTFs) such as TFIID or RNA polymerase II as suggested by Chen and colleagues with the AD3 of E-proteins (Chen et al. 2013). It is harder to imagine how this scenario would be tissue or target-gene specific. The promoter of *arg-1* is especially useful for considering these various possibilities since it has distinct tissue-specific E boxes and future work identifying Factor Y and HLH-8-interacting factors is expected to shed more light on bHLH TF function (Fig. 10).

## Conclusions

Altogether our work supports the existence in vivo of a TWIST-Box domain that is important for tissue-specific transcriptional activity in HLH-8. More importantly, through a combination of alanine screening and SCS patient-specific mutation modeling of the equidistant, evolutionarily conserved LXXXFXXXL motif of the TWIST-Box, we uncovered an F > R > L hierarchy of functional importance to this predicted  $\alpha$ -helical domain. Here, we identified a distinct Ret phenotype in  $\Delta 95$ –103 allele that also was found in the Triple A mutant and the F99A allele where this single amino acid substitution was sufficient to cause the phenotype. Most of the HLH-8 TWIST-Box mutants had a semi-dominant, *arg-1* target-gene, tissue-specific phenotype that is dependent on the conserved amino acid and the particular substitution. Further work to discover HLH-8 TWIST-Box-interacting proteins or co-factors influenced by the domain will reveal more about the function of this TWIST1 homolog, which could lead to a better understanding of how the domain could be defective in SCS patients.

## Data availability

Strains and other materials are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. [Supplemental material](#) available at GENETICS online.

## Acknowledgements

The authors would like to thank the following: Ben Nebenfuhr, Xiaofei Bai, and Andy Golden for advice in CRISPR/Cas9 experimental design and assistance with worm injections; Andy Golden and John Golin for manuscript comments, Reagan Sharkey for technical help, Rabby Khoshnade and William Boudhraa for fruitful discussions, and the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), for providing some of the strains.

## Funding

Research reported in this publication was supported by the National Institute of Dental & Craniofacial Research of the National Institutes of Health under Award Number R15DE018519. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of interest

The author(s) declare no conflict of interest.

## Literature cited

- Arribere JA, Bell RT, Fu BX, Artiles KL, Hartman PS, Fire AZ. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics*. 2014;198(3):837–846. doi:10.1534/genetics.114.169730.
- Barnes RM, Firulli AB. A twist of insight—the role of Twist-family bHLH factors in development. *Int J Dev Biol*. 2009;53(7):909–924. doi:10.1387/ijdb.082747rb.
- Bergstrom DA, Tapscott SJ. Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. *Mol Cell Biol*. 2001;21(7):2404–2412. doi:10.1128/MCB.21.7.2404-2412.2001.
- Bialek P, Kern B, Yang X, Schrock M, Susic D, Hong N, Wu H, Yu K, Ornitz DM, Olson EN, et al. A twist code determines the onset of osteoblast differentiation. *Dev Cell*. 2004;6(3):423–435. doi:10.1016/S1534-5807(04)00058-9.
- Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980;32(3):314–331.
- Castanon I, Von Stetina S, Kass J, Baylies MK. Dimerization partners determine the activity of the Twist bHLH protein during *Drosophila* mesoderm development. *Development*. 2001;128(16):3145–3159. doi:10.1242/dev.128.16.3145.
- Chang AT, Liu Y, Ayyanathan K, Benner C, Jiang Y, Prokop JW, Paz H, Wang D, Li HR, Fu XD, et al. An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev*. 2015;29(6):603–616. doi:10.1101/gad.242842.114.
- Chen J, Li X, Greenwald I. *sel-7*, a positive regulator of *lin-12* activity, encodes a novel nuclear protein in *Caenorhabditis elegans*. *Genetics*. 2004;166(1):151–160. doi:10.1534/genetics.166.1.151.
- Chen RZ, Cheng X, Tan Y, Chang TC, Lv H, Jia Y. An ENU-induced mutation in Twist1 transactivation domain causes hindlimb polydactyly with complete penetrance and dominant-negatively impairs E2A-dependent transcription. *Sci Rep*. 2020;10(1):2501. doi:10.1038/s41598-020-59455-9.
- Chen WY, Zhang J, Geng H, Du Z, Nakadai T, Roeder RG. A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. *Genes Dev*. 2013;27(14):1596–1609. doi:10.1101/gad.216192.113.
- Corsi AK, Brodigan TM, Jorgensen EM, Krause M. Characterization of a dominant negative *C. elegans* Twist mutant protein with implications for human Saethre-Chotzen syndrome. *Development*. 2002;129(11):2761–2772. doi:10.1242/dev.129.11.2761.
- Corsi AK, Kostas SA, Fire A, Krause M. *Caenorhabditis elegans* twist plays an essential role in non-striated muscle development. *Development*. 2000;127(10):2041–2051. doi:10.1242/dev.127.10.2041.
- de Martin X, Sodaei R, Santpere G. Mechanisms of binding specificity among bHLH transcription factors. *Int J Mol Sci*. 2021;22(17):9150. doi:10.3390/ijms22179150.
- Denis CM, Langelaan DN, Kirilin AC, Chitayat S, Munro K, Spencer HL, LeBrun DP, Smith SP. Functional redundancy between the transcriptional activation domains of E2A is mediated by binding to the KIX domain of CBP/p300. *Nucleic Acids Res*. 2014;42(11):7370–7382. doi:10.1093/nar/gku206.
- Fay DS. Classical genetic methods (December 30, 2013), WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.165.1, <http://www.wormbook.org> 2013.

- Fitzgerald K, Greenwald I. Interchangeability of *Caenorhabditis elegans* DSL proteins and intrinsic signalling activity of their extracellular domains in vivo. *Development*. 1995;121(12):4275–4282. doi:10.1242/dev.121.12.4275.
- Franco HL, Casanovas J, Rodríguez-Medina JR, Cadilla CL. Redundant or separate entities? –roles of Twist1 and Twist2 as molecular switches during gene transcription. *Nucleic Acids Res*. 2011;39(4):1177–1186. doi:10.1093/nar/gkq890.
- Gajula RP, Chettiar ST, Williams RD, Thiyagarajan S, Kato Y, Aziz K, Wang R, Gandhi N, Wild AT, Vesuna F, et al. The twist box domain is required for Twist1-induced prostate cancer metastasis. *Mol Cancer Res*. 2013; 11(11):1387–1400. doi:10.1158/1541-7786.MCR-13-0218-T.
- Gitelman I. Twist protein in mouse embryogenesis. *Dev Biol*. 1997; 189(2):205–214. doi:10.1006/dbio.1997.8614.
- Gruss M, Corsi AK. Using *Caenorhabditis elegans* as a model for mechanistic insights of craniofacial development. *Methods Mol Biol*. 2022;2403:1–18. doi:10.1007/978-1-0716-1847-9\_1.
- Gu S, Boyer TG, Naski MC. Basic helix-loop-helix transcription factor Twist1 inhibits transactivator function of master chondrogenic regulator sox9. *J Biol Chem*. 2012;287(25):21082–21092. doi:10.1074/jbc.M111.328567.
- Harfe BD, Vaz Gomes A, Kenyon C, Liu J, Krause M, Fire A. Analysis of a *Caenorhabditis elegans* Twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes Dev*. 1998;12(16):2623–2635. doi:10.1101/gad.12.16.2623.
- Horvitz HR, Sulston JE. Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics*. 1980;96(2):435–454. doi:10.1093/genetics/96.2.435.
- Kaustov L, Yi GS, Ayed A, Bochkareva E, Bochkarev A, Arrowsmith CH. p53 transcriptional activation domain: a molecular chameleon? *Cell Cycle*. 2006;5(5):489–494. doi:10.4161/cc.5.5.2489.
- Kim S, Twigg SRF, Scanlon VA, Chandra A, Hansen TJ, Alsubait A, Fenwick AL, McGowan SJ, Lord H, Lester T, et al. Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in *Caenorhabditis elegans*. *Hum Mol Genet*. 2017;26(11):2118–2132. doi:10.1093/hmg/ddx107.
- Kress W, Schropp C, Lieb G, Petersen B, Büsse-Ratzka M, Kunz J, Reinhart E, Schäfer WD, Sold J, Hoppe F, et al. Saethre-Chotzen syndrome caused by TWIST 1 gene mutations: functional differentiation from Muenke coronal synostosis syndrome. *Eur J Hum Genet*. 2006;14(1):39–48. doi:10.1038/sj.ejhg.5201507.
- Lander R, Nasr T, Ochoa SD, Nordin K, Prasad MS, Labonne C. Interactions between Twist and other core epithelial-mesenchymal transition factors are controlled by GSK3-mediated phosphorylation. *Nat Commun*. 2013;4(1):1542. doi:10.1038/ncomms2543.
- Lander R, Nordin K, LaBonne C. The F-box protein Ppa is a common regulator of core EMT factors Twist, Snail, Slug, and Sip1. *J Cell Biol*. 2011;194(1):17–25. doi:10.1083/jcb.201012085.
- Laursen KB, Mielke E, Iannaccone P, Füchtbauer EM. Mechanism of transcriptional activation by the proto-oncogene twist1. *J Biol Chem*. 2007;282(48):34623–34633. doi:10.1074/jbc.M707085200.
- Li S, Kendall SE, Raices R, Finlay J, Covarrubias M, Liu Z, Lowe G, Lin YH, Teh YH, Leigh V, et al. TWIST1 associates with NF- $\kappa$ B subunit RELA via carboxyl-terminal WR domain to promote cell autonomous invasion through IL8 production. *BMC Biol*. 2012;10(1):73. doi:10.1186/1741-7007-10-73.
- Massari ME, Jennings PA, Murre C. The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both *Saccharomyces cerevisiae* and mammalian cells. *Mol Cell Biol*. 1996;16(1):121–129. doi:10.1128/MCB.16.1.121.
- McAndrew PC, Svaren J, Martin SR, Hörz W, Goding CR. Requirements for chromatin modulation and transcription activation by the Pho4 acidic activation domain. *Mol Cell Biol*. 1998;18(10):5818–5827. doi:10.1128/MCB.18.10.5818.
- McGovern M, Voutev R, Maciejowski J, Corsi AK, Hubbard EJ. A “latent niche” mechanism for tumor initiation. *Proc Natl Acad Sci U S A*. 2009;106(28):11617–11622. doi:10.1073/pnas.0903768106.
- Mello CC, Draper BW, Priess JR. The maternal genes *apx-1* and *glp-1* and establishment of dorsal-ventral polarity in the early *C. elegans* embryo. *Cell*. 1994;77(1):95–106. doi:10.1016/0092-8674(94)90238-0.
- Paix A, Folkmann A, Rasololoson D, Seydoux G. High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-cas9 ribonucleoprotein complexes. *Genetics*. 2015;201(1): 47–54. doi:10.1534/genetics.115.179382.
- Peña WA, Slavotinek A, Oberoi S. Saethre-Chotzen syndrome: a case report. *Cleft Palate Craniofac J*. 2010;47(3):318–321. doi:10.1597/07-202.1.
- Philogene MC, Small SG, Wang P, Corsi AK. Distinct *Caenorhabditis elegans* HLH-8/twist-containing dimers function in the mesoderm. *Dev Dyn*. 2012;241(3):481–492. doi:10.1002/dvdy.23734.
- Piccinin S, Tonin E, Sessa S, Demontis S, Rossi S, Pecciarini L, Zanatta L, Pivetta F, Grizzo A, Sonogo M, et al. A “twist box” code of p53 inactivation: twist box: p53 interaction promotes p53 degradation. *Cancer Cell*. 2012;22(3):404–415. doi:10.1016/j.ccr.2012.08.003.
- Qin Q, Xu Y, He T, Qin C, Xu J. Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell Res*. 2012;22(1):90–106. doi:10.1038/cr.2011.144.
- Quong MW, Massari ME, Zwart R, Murre C. A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. *Mol Cell Biol*. 1993;13(2):792–800. doi:10.1128/mcb.13.2.792-800.1993.
- Rodríguez Y, Gonzalez-Mendez RR, Cadilla CL. Evolution of the twist subfamily vertebrate proteins: discovery of a signature motif and origin of the Twist1 Glycine-rich motifs in the amino-terminus disordered domain. *PLoS One*. 2016;11(8):e0161029. doi: 10.1371/journal.pone.0161029.
- Schwarz JJ, Chakraborty T, Martin J, Zhou JM, Olson EN. The basic region of myogenin cooperates with two transcription activation domains to induce muscle-specific transcription. *Mol Cell Biol*. 1992;12(1):266–275. doi:10.1128/mcb.12.1.266-275.1992.
- Seto ML, Hing AV, Chang J, Hu M, Kapp-Simon KA, Patel PK, Burton BK, Kane AA, Smyth MD, Hopper R, et al. Isolated sagittal and coronal craniosynostosis associated with TWIST box mutations. *Am J Med Genet A*. 2007;143A(7):678–686. doi:10.1002/ajmg.a.31630.
- Singh K, Dilworth FJ. Differential modulation of cell cycle progression distinguishes members of the myogenic regulatory factor family of transcription factors. *FEBS J*. 2013;280(17):3991–4003. doi:10.1111/febs.12188.
- Stern MJ, Horvitz HR. A normally attractive cell interaction is repulsive in two *C. elegans* mesodermal cell migration mutants. *Development*. 1991;113(3):797–803. doi:10.1242/dev.113.3.797.
- Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol*. 1977;56(1):110–156. doi:10.1016/0012-1606(77)90158-0.
- Sulston JE, White JG. Regulation and cell autonomy during postembryonic development of *Caenorhabditis elegans*. *Dev Biol*. 1980; 78(2):577–597. doi:10.1016/0012-1606(80)90353-x.
- Thisse B, Stoetzel C, Gorostiza-Thisse C, Perrin-Schmitt F. Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. *EMBO J*. 1988;7(7):2175–2183. doi:10.1002/j.1460-2075.1988.tb03056.x.
- Twigg SR, Wilkie AO. A genetic-pathophysiological framework for craniosynostosis. *Am J Hum Genet*. 2015;97(3):359–377. doi:10.1016/j.ajhg.2015.07.006.
- Weintraub H, Dwarki VJ, Verma I, Davis R, Hollenberg S, Snider L, Lassar A, Tapscott SJ. Muscle-specific transcriptional activation



- by MyoD. *Genes Dev.* 1991;5(8):1377–1386. doi:10.1101/gad.5.8.1377.
- Wilkie AO. Craniosynostosis: genes and mechanisms. *Hum Mol Genet.* 1997;6(10):1647–1656. doi:10.1093/hmg/6.10.1647.
- Wu Z, Zou X, Xu Y, Zhou F, Kuai R, Li J, Yang D, Chu Y, Peng H. Ajuba transactivates N-cadherin expression in colorectal cancer cells through interaction with Twist. *J Cell Mol Med.* 2021;25(16):8006–8014. doi:10.1111/jcmm.16731.
- Zammit PS. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin Cell Dev Biol.* 2017;72:19–32. doi:10.1016/j.semcdb.2017.11.011.
- Zhao J, Wang P, Corsi AK. The *C. elegans* Twist target gene, *arg-1*, is regulated by distinct E box promoter elements. *Mech Dev.* 2007;124(5):377–389. doi:10.1016/j.mod.2007.01.005.

Editor: T. Schedl