Supplementary Material

Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in *C. elegans*

Sharon Kim[1], Stephen R.F. Twigg[2], Victoria A. Scanlon[3], Aditi Chandra[1], Tyler J. Hansen[1], Arwa Alsubait[3], Aimee L. Fenwick[2], Simon J. McGowan[4], Helen Lord[5], Tracy Lester[5], Elizabeth Sweeney[6], Astrid Weber[6], Helen Cox[7], Andrew O.M. Wilkie[2], Andy Golden[1], Ann K. Corsi[3*]

[1]Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892, USA

[2]Clinical Genetics Group, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK

[3]Department of Biology, The Catholic University of America, Washington, DC, 20064, USA

[4]Computational Biology Research Group, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK

[5]Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Foundation Trust, Churchill Hospital, Oxford OX3 7LE, UK

[6]Department of Clinical Genetics, Liverpool Women's NHS Foundation Trust, Liverpool L8 7SS, UK

[7]Clinical Genetics Unit, Birmingham Women's NHS Foundation Trust, Birmingham Women's Hospital, Birmingham B15 2TG, UK

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors and the last three as joint Corresponding Authors.

MATERIALS AND METHODS

Assay to score phenotypes

Defecation assay: Young to mid-adult animals were picked to a fresh lawn of bacteria and allowed to acclimate for 5 minutes before assaying. This was to avoid any changes to the defecation cycle as a result of picking and transferring the animals. Only actively feeding worms were assayed under a dissecting scope. Each worm was closely observed over a period of 5 pBocs (posterior body muscle contraction) and it was noted whether the pBoc was followed by an expulsion of feces (22). Thirty worms were assayed for each Glu29† allele and N2. Severe alleles occasionally displayed "leaking" of feces between pBocs due to intestinal pressure, but only feces following a pBoc were counted as true expulsions as a result of some degree of muscle function.

All other methods can be found in the main paper.

References

22. Thomas, J.H. (1990) Genetic analysis of defecation in *Caenorhabditis elegans*. *Genetics*. **124**, 855-872.

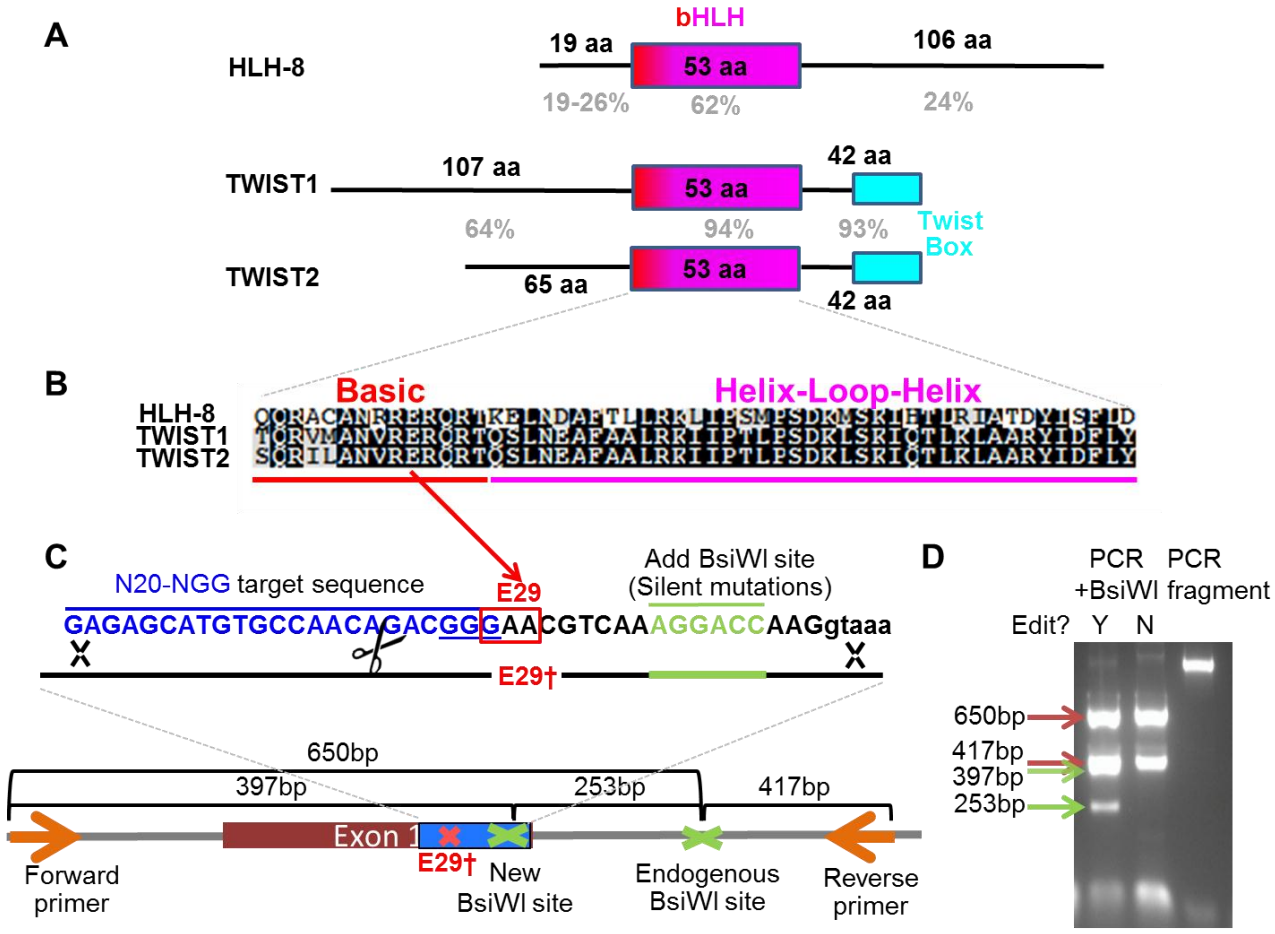

Figure S1. Strategy for modeling human TWIST1/TWIST2 Glu29 equivalent mutations in *hlh-8*. (A) Diagram indicating the amino acid identity between HLH-8, TWIST1, and TWIST2. (B) Alignment of the bHLH domains of the three proteins. (C) CRISPR/Cas9 technology was used to create the desired alleles of *hlh-8*. The guide RNA target sequence (blue line above sequence) was used to direct Cas9 to cut the *hlh-8* locus adjacent to the Glu29 codon at the PAM site (blue line below sequence). A repair oligonucleotide (line below sequence) was used that introduces a silent BsiWI site (indicated in green) to enable screening for successful repair events by PCR amplification (primers indicated in orange) and restriction digestion. Figure is not to scale. (D) Example restriction digest showing a successful heterozygous edit (additional fragments of 253 and 397 bp) in the first lane.

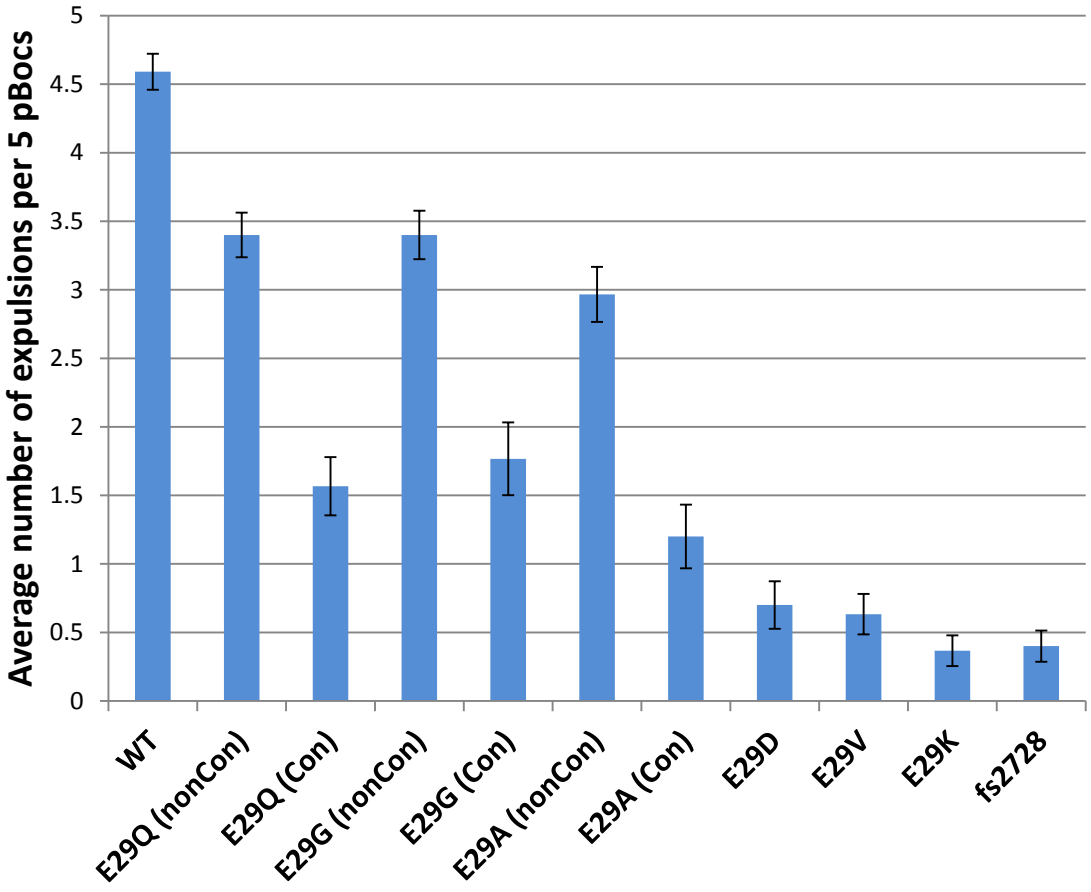

Figure S2. Defecation defects in Glu29† *hlh-8* mutants. Extent of constipation was measured by counting how many expulsion events (enteric muscle dependent) occurred for every five posterior body wall contractions (pBoc; enteric muscle independent). N=30 for each bar. Error bars are standard error of the mean (SEM).

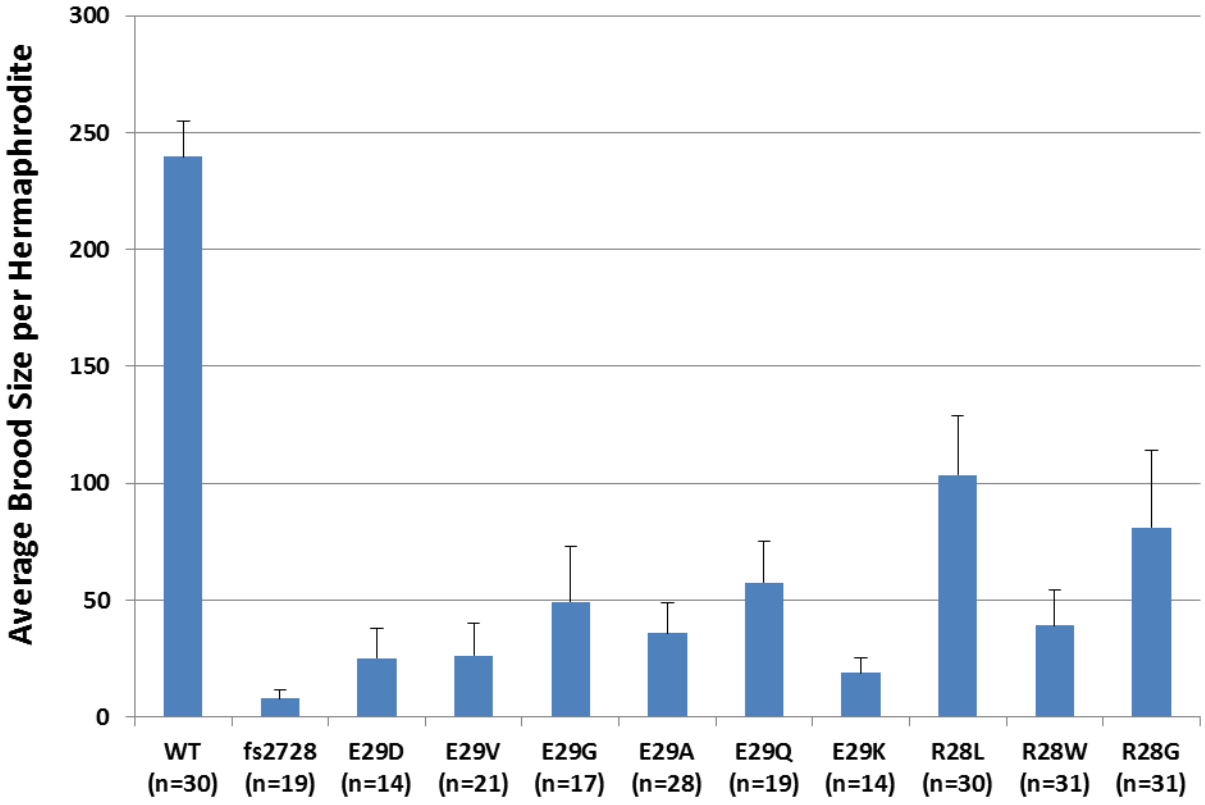

Figure S3. Average brood sizes are lower in Glu29† and Arg28† mutants compared to wild-type (WT) animals. Brood sizes were determined by counting the number of offspring from single hermaphrodites until the animals had stopped laying embryos or had released all of the progeny that hatched internally (over the course of several days). Error bars are standard deviation.

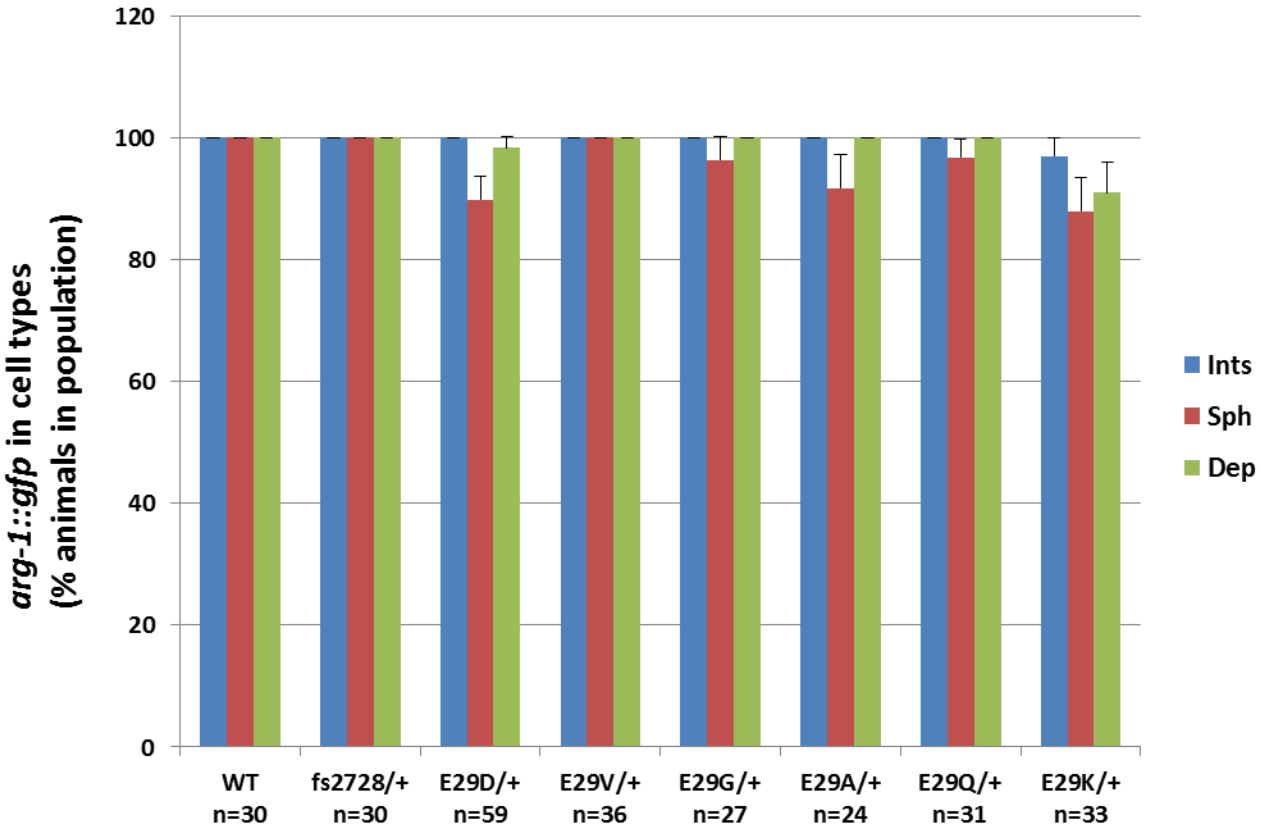

Figure S4. *hlh-8* Glu29†/+ heterozygotes are not defective for *arg-1* expression in enteric muscles. The percentage of animals in a population that were expressing the *arg-1::gfp* in the intestinal muscles (Ints), the anal sphincter (Sph) and the anal depressor (Dep) were determined. Error bars are standard error of the proportion.

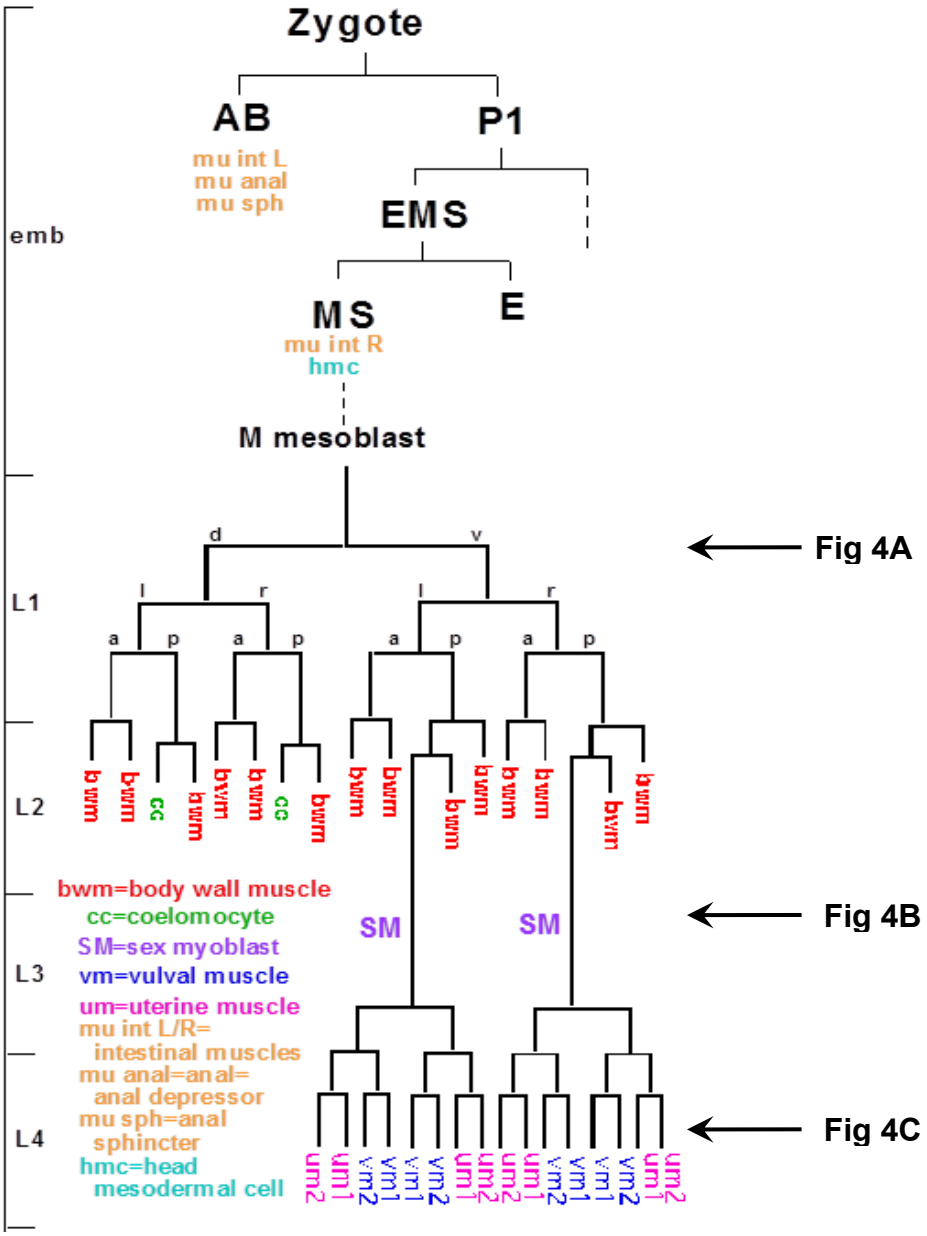

Figure S5. HLH-8-expressing cells in the *C. elegans* embryo and the M lineage. An abbreviated lineage showing some of the cells that arise during development (stage indicated on the left: emb, embryogenesis, L1-L4, 1st, 2nd, 3rd, and 4th larval stages; not drawn to scale). Each vertical line represents a cell and horizontal line a cell division (the plane of division is indicated as d, dorsal, v, ventral, a, anterior, p, posterior, l, left, r, right). The stage at which the M lineage was examined in Figure 4 is indicated on the right (Fig. 4A, two d/v M descendants; Fig. 4B, 2 sex myoblasts (SMs); Fig. 4C, 16 SM descendants (SMd)).

Supplementary Table 1. Rare variants identified in the exome sequence of Subject 1.[a]

Gene	Description	Chr	Position	cDNA change	predicted aa change	Deleterious score	dbSNP	1000G_all	EVS	ExAC_allele_freq	Inheritance
TWIST1	twist family bHLH transcription factor 1	7	19116972	c.350A>T	p.E117V	4	-	-	-	-	de novo
IL12RB1	interleukin 12 receptor, beta 1	19	18069582	c.1153C>G	p.L385V	-	-	-	-	-	de novo
PLCB4	phospholipase C, beta 4	20	9384269	c.922C>T	p.R308C	6	rs78074693	0.000399	0.0004	0.0004	maternal
PLCB4	phospholipase C, beta 4	20	9362947	c.421G>A	p.V141I	5	rs144345083	-	0.0003	0.0002	paternal
HECTD4	HECT domain containing E3 ubiquitin protein ligase 4	12	112231711	c.5600G>A	p.G1867D	5	rs374620778	-	0.000083	1.04E-05	maternal
HECTD4	HECT domain containing E3 ubiquitin protein ligase 4	12	112236983	c.5004G>A	-	-	rs371288014	-	0.0002	5.69E-05	paternal
DNAH1	dynein, axonemal, heavy chain 1	3	52381773	c.7742A>G	p.N2581S	4	rs200839854	0.000399	0.0009	0.0016	maternal
DNAH1	dynein, axonemal, heavy chain 1	3	52390959	c.9646C>G	p.L3216V	3	rs200158571	0.000399	0.0007	0.0005	paternal
DNAH1	dynein, axonemal, heavy chain 1	3	52362371	c.4981-17G>A	-	-	rs370018310	0.0002	0.0008	0.0014	maternal
CFAP47	cilia and flagella associated protein 47	X	35971764	c.2149A>G	p.R717G	-	rs201682692	-	0.000095	7.07E-05	maternal

[a]Variants listed are those compatible with a genetic origin of sporadic disease (*de novo* dominant, autosomal or X-linked recessive mechanisms).

Abbreviations: Chr, chromosome; aa, amino acid; 1000G, 1000 Genomes Project; EVS, Exome Variant Server; ExAC, Exome Aggregation Consortium.