**Averaged species Ranking Criteria**

- **A’** target gene region** site score
- **B’** target gene region** score
- **C’** target gene region** site count
- **D’** target gene total score
- **E’** target gene total site count
- **F’** target gene region** conservation
- **G’** target gene offset variance

**C. elegans Ranking Criteria**

- **A’** target gene region** site score
- **B’** target gene region** score
- **C’** target gene region** site count
- **D’** target gene total score
- **E’** target gene total site count

**Region-upstream and intron**

Values in the grey area are weights, w, applied to ranking criteria A–K used to derive the cumulative site score* in H and I.

**Standardized MSU comparison test results for individual ranking criteria in TTX-3/CEH-10 known target genes**

<table>
<thead>
<tr>
<th>Plot</th>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>J</td>
<td>A</td>
<td>0</td>
</tr>
</tbody>
</table>

**Default method (J’)** used by TargetOrtho for generation of cumulative site scores

**TTX-3/CEH-10 unknown target genes (minus 6 genes used to generate MSU bias corrected data) from upstream and intronic regions**

- **A’** target gene region** site score
- **B’** target gene region** score
- **C’** target gene region** site count
- **D’** target gene total score
- **E’** target gene total site count

**Region-upstream and intron**

Values in the grey area are weights, w, applied to ranking criteria A–K used to derive the cumulative site score in H and I.

**Cumulative site score = \sum_{i=1}^{N} c_i \times \left( \frac{a_i - b_i}{w_i} \right) \times \left( \frac{c_i}{a_i} \right) \times \left( \frac{b_i}{c_i} \right)**

**Region-upstream and intron**

Values in the grey area are weights, w, applied to ranking criteria A–K used to derive the cumulative site score in H and I.

**Legend**

- A: Averaged species Ranking Criteria
- B: target gene region** site score
- C: target gene region** site count
- D: target gene total site count
- E: target gene total score
- F: target gene region** conservation
- G: target gene offset variance

**TTX-3/CEH-10 unknown target genes**

- **A’** target gene region** site score
- **B’** target gene region** score
- **C’** target gene region** site count
- **D’** target gene total score
- **E’** target gene total site count

**Region-upstream and intron**

Values in the grey area are weights, w, applied to ranking criteria A–K used to derive the cumulative site score in H and I.

**Default method (J’)** used by TargetOrtho for generation of cumulative site scores

**Legend**

- A: Averaged species Ranking Criteria
- B: target gene region** site score
- C: target gene region** site count
- D: target gene total site count
- E: target gene total score
- F: target gene region** conservation
- G: target gene offset variance

**TTX-3/CEH-10 unknown target genes**

- **A’** target gene region** site score
- **B’** target gene region** score
- **C’** target gene region** site count
- **D’** target gene total score
- **E’** target gene total site count

**Region-upstream and intron**

Values in the grey area are weights, w, applied to ranking criteria A–K used to derive the cumulative site score in H and I.

**Default method (J’)** used by TargetOrtho for generation of cumulative site scores

**Legend**

- A: Averaged species Ranking Criteria
- B: target gene region** site score
- C: target gene region** site count
- D: target gene total site count
- E: target gene total score
- F: target gene region** conservation
- G: target gene offset variance
Figure S7  AIY motif analysis with verification bias corrected data. *ceh-10/tx-3* dependent target gene data (blue) compared to random coding gene data (grey). The set of previously characterized *ceh-10/tx-3* dependent genes (except those used to construct the AIY motif) and 1000 random coding genes were submitted to TargetOrtho using the ASE motif as input (Figure S1E). Data distributions for each TargetOrtho ranking criteria were compared between known target genes and random coding genes.

CDF plots of individual ranking criteria (plots A-E and plots A’-G’): CDF plots are shown for individual ranking criteria A-E and A’-G’. TargetOrtho ranking criteria derived from averaged species data (A’-G’) better distinguish previously validated TF target genes from random genes compared to using *C. elegans* (reference genome) data alone (A-E). CDF plots A-E show ranking criteria derived from *C. elegans* genome data only while CDF plots A’-E’ show the corresponding ranking criteria derived from averaged species data. CDF plot F’ and G’ show averaged species data having no reference genome counterpart including the conservation and offset variance data distributions.

CDF plots of cumulative site scores (plots H, I and plots H’, I’, J’): Data distributions for cumulative site scores derived from unique combinations of TargetOrtho ranking criteria are shown in CDF plots H,I,H’,I’,J’. CDF plot H shows the cumulative site score distributions derived from *C. elegans* upstream and intronic data only calculated from A-C. The left panel, plots A’-C’ shows the cumulative site score CDF plots calculated from the corresponding averaged species upstream and intronic data. CDF plot I shows cumulative site scores derived from criteria shown in CDF plots A-E where CDF plots D and E represent total gene ranking criteria in *C. elegans* only (D. *C. elegans* averaged upstream and intronic site scores and E. *C. elegans* averaged site score across all gene regions). CDF plot I’ (left panel) shows the data distribution of cumulative site scores derived from A’-E’ where CDF plots D’ and E’ represent the corresponding total gene ranking criteria averaged across species. CDF plot J’ shows cumulative site scores derived from all averaged species ranking criteria (A’-G’).

K. $-\log_{10}(P \text{ value})$ for each ranking criteria comparison test where transcription factor dependent genes were compared to 1000 random coding genes. Compare *C. elegans* data A-E to average species data A’-E’ plus F’ and G’.

L. $-\log_{10}(P \text{ values})$ for each comparison test where cumulative site scores in transcription factor dependent genes are compared to scores in random coding genes. Compare *C. elegans* derived cumulative site score (H and I) to averaged species derived cumulative sites scores (H’, I’, and J’).