

Supplemental Material: Long intramolecular duplex structures are associated with essential and highly expressed genes on *C. elegans* autosome distal arms

Daniel P. Reich and Brenda L. Bass

Supplemental Methods

Transposon enrichment

A suite of applications available at <https://github.com/4ureliek/> was used to determine significantly enriched transposons in essential and highly expressed genes (Kapusta et al. 2013). We ran the application TE-analysis_Shuffle_bed.pl with parameters “-s bed -n 1000”, excluding chromosome centers and Chromosome X, to determine which essential and highly expressed genes on autosome arms overlapped RepeatMasker-annotated transposons. This application also determined expected overlaps by randomly shuffling genes around autosome distal arm domains over 1000 iterations.

Analyses of N2 and CB4856 gene expression with/without palindromic transposons

Genomewide microarray data cataloguing gene expression in three biological replicates of N2 and CB4856 whole L4 larvae (Kamkina et al. 2016) was downloaded from the WormQTL database (<http://www.wormqtl.org/molgenis.do>). We matched probe designations to appropriate genes using a table downloaded from GEO platform accession GPL10094. For genes measured by multiple probes, we averaged expression values across all probes for that gene within each biological replicate.

A list of palindromic transposons present in N2 and deleted in CB4856 was curated from Additional File 11 in (Vergara et al. 2014). Likewise, palindromic transposons absent in N2 and present in CB4856, generated from analyses in (Laricchia et al. 2017), were determined from data available from the *C. elegans* Natural Diversity Resource (<https://www.elegansvariation.org/>). To identify genes with transposons in one strain but not the other, we overlapped the coordinates of deleted or inserted

transposons with a bed file of intron-containing genes (see Supplemental File S2) using the bedtools2 applicable intersectBed.

Intron PATC analysis

We downloaded a BigWig file of balanced PATC scores previously calculated for the *C. elegans* genome at 25 nt resolution (Frokjaer-Jensen et al. 2016). Using the UCSC Genome Browser utility bigWigAverageOverBed, we calculated the average PATC score for each unique *C. elegans* intron. We then determined the highest intron PATC score for each intron-containing gene.