Original Research Article

Natural Variation in DNA-Damage Checkpoint Gene Expression and Genotype in Relation to Oxidative Stress Vulnerability in *C. elegans*

Karen Song

Walnut High School, 400 Pierre Rd, Walnut, CA 91789, United States

ABSTRACT

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disorder characterized by motor neuron degeneration and progressive paralysis. Familial ALS, which accounts for 5–10% of cases, remains largely uncharacterized, shown by the large variability in disease progression among patients. Emerging evidence suggests that the disruption of DNA damage response (DDR) pathways may contribute to genome instability, a feature implicated in neurodegeneration. In this study, the model organism *Caenorhabditis elegans* is used to investigate how natural variation in DNA damage checkpoint genes: *mrt-2*, *hus-1*, *hpr-17*, and *atm-1*, all of which interact with or participate in the 9-1-1 DNA damage response pathway, predicts vulnerability to oxidative stress. RNA-seq expression data from 208 genetically diverse *C. elegans* strains were normalized for sequencing depth (counts per million) after filtering out low-count transcripts. Three of the four genes showed significant differences in gene expression between the strains, with *hpr-17* displaying the highest heritability. *Cis*-expression quantitative trait locus (*cis*-eQTL) mapping was used to analyze hard-filtered variant information from *Caenorhabditis elegans* Natural Diversity Resource (CaeNDR), revealing no strong local signal after kinship correction, suggesting *trans*-regulation, rare-variant effects, or epigenetics as influences of expression differences. To evaluate whether checkpoint gene expression predicts oxidative stress vulnerability, paraquat sensitivity data were obtained from the CaeNDR Public Phenotype Database and analyzed. The lower body length after paraquat exposure indicates greater sensitivity. Predictive models, including Ridge regression and XGBoost, were then applied to test whether checkpoint gene expression predicts sensitivity. Results suggest that variation in checkpoint expression does not predict oxidative stress sensitivity; however, an exploratory genetic scan identified a locus near *hpr-17* that was associated with paraquat response, survived kinship correction, and was independently replicated using a larger imputed variant set, warranting further research to understand its direct effect on neuron health.

Keywords: Amyotrophic lateral sclerosis; 9-1-1 complex; DNA damage response; *mrt-2; hus-1; hpr-17; atm-1;* oxidative stress

INTRODUCTION

ALS (Amyotrophic Lateral Sclerosis) is a progressive neurodegenerative disorder characterized by motor neuron degeneration and resulting in progressive paralysis and death, typically within 3 to 5 years after the onset of ALS symptoms (1, 2). The prevalence

Corresponding author: Karen Song, E-mail: karensong9009@gmail.com.

Accepted August 4, 2026
https://doi.org/10.70251/HYJR2348.44772787

of ALS is projected to increase significantly over the coming years, with estimates suggesting a rise of more than 10% in the United States before 2030 (3). Familial ALS accounts for around 5-10% of ALS cases and exhibits substantial variation in the age of onset and progression rate amongst patients, even in patients with the same causal mutation (4,5). This variability suggests that additional genetic modifiers beyond the primary mutations associated with disease onset play a crucial role in directing disease trajectory. However, these modifying factors remain poorly understood, and the mechanisms underlying variability in ALS between patients is still poorly characterized.

Emerging evidence suggests that disrupting the DDR pathway contributes to neurodegeneration (6). Post-mitotic neurons are especially vulnerable to the accumulation of unrepaired DNA damage over an individual's lifetime, and dysregulated genome maintenance has been proposed as a possible contributing factor in ALS and other related neurodegenerative disorders (7). A core component of DDR is the highly conserved 9-1-1 DNA damage checkpoint complex, which is composed of RAD9, RAD1, and HUS1, which are loaded onto damaged DNA by the RAD17-RFC clamp loader (8,9). Together, these proteins play a crucial role in detecting replication stress and double-strand breaks that result in the halting of cell cycle progression, allowing for these issues to be repaired, or triggering apoptosis in cells when the damage is irreparable. The ATM kinase, a central upstream regulator of DDR, responds similarly to double-strand breaks and works to coordinate checkpoint signaling in cells. In humans, the genes that encode these proteins (RAD1, HUS1, RAD17, ATM) and their natural variation have been previously linked to cancer susceptibility, aging, and neurodegeneration (10). However, whether natural variation in checkpoint gene expression influences cellular stress remains largely unexplored.

C. elegans serves as a powerful model organism to address this gap and to study ALS (11). The nematode's 9-1-1 pathway is highly conserved, with *mrt-2, hus-1, and hpr-17, atm-1* all serving as direct orthologs of RAD1, HUS1, RAD17, and ATM respectively (12). Loss-of-function in *mrt-2* and *hus-1* produces a well-studied mortal germline phenotype, where failure of DNA damage checkpoint signaling causes progressive telomere shortening, chromosome end-to-end fusions, and eventual sterility across generations (13). This directly demonstrates the critical role of this pathway in ensuring genome stability in proliferating tissues. In recent years, the CaeNDR project has generated whole-genome sequences and population-scale RNA-seq (RNA sequencing) data for hundreds of genetically diverse wild *C. elegans* strains along with a quickly growing set of quantitative phenotypes (14). These resources provide an opportunity to investigate how naturally occurring, non-pathogenic variation in checkpoint gene expression and genotype relates to organismal stress phenotypes that are independent of the rare, high-impact mutations which are traditionally studied in classical genetics.

Notably, a direct link between *C. elegans* DDR biology and ALS is further supported by work on the *C. elegans* orthologs of the ALS-associated RNA-binding proteins TDP-43 and FUS. *tdp-1* and *fust-1*, which are the orthologs of TDP-43 and FUS in humans, interact with DNA damage and proteostasis pathways that influence stress resistance and genome stability (15). It has been found that the loss of *tdp-1* sensitizes worms to oxidative stress, alters resistance to heat-shock-factor deficiency, and modulates lifespan through the DAF-16/FOXO pathway (16-19). Furthermore, both *tdp-1* and *fust-1* each exhibit temperature-sensitive, genome-stability-linked phenotypes in combination with transcriptional regulators (16, 20, 21). These findings suggest that variation in checkpoint gene activity could influence the same cellular stress response pathways that are often implicated in ALS protein toxicity. However, no previous study has examined whether naturally occurring variation in checkpoint gene expression is directly linked to ALS related stress phenotypes in *C. elegans*.

Prior work using this same CaeNDR wild-isolate collection has shown that natural variation in telomere length, a downstream result of the 9-1-1 pathway and telomerase activity, is itself substantial across the various nematode strains (22). However, telomere length was not associated with the amount of offspring produced or the lifespan of a nematode. Much of this variation was instead explained by a coding variant in the telomere binding gene *pot-2*, independent of checkpoint gene expression. These results suggest that telomere length and checkpoint gene activity are distinct biological processes rather than interchangeable ways to measure DDR function. Consequently, it remains unknown whether natural variation in checkpoint gene expression predicts stress-related phenotypes independently, regardless of downstream telomere length.

In this study, population-scale RNA-seq, whole-genome genotype data, and oxidative-stress phenotypes from the CaeNDR wild-isolate collection were used to investigate four hypotheses regarding the checkpoint

genes *mrt-2*, *hus-1*, *hpr-17*, and *atm-1*. First, it is hypothesized that these genes show significant and heritable inter-strain variation in expression. Second, this expression variation is hypothesized to be explained primarily by common local (*cis*-acting) genetic variants. Third, checkpoint gene expression is hypothesized to correlate with telomere length, a downstream readout of 9-1-1 pathway and telomerase activity. Fourth, checkpoint gene expression is hypothesized to predict nematode vulnerability to oxidative stress, a proteotoxicity-relevant phenotype linked to the protein dysfunction studied in *C. elegans* ALS models. Oxidative stress vulnerability was measured using paraquat response data from the CaeNDR Public Phenotype Database, reported as the control-normalized median *C. elegans* body length after exposure to 250 µM paraquat (Data Sources and Accession). Because paraquat impairs development, lower values indicate shorter body length than expected and therefore greater sensitivity to oxidative stress, whereas higher values indicate greater resistance. Results show partial support for the first hypothesis, with three of the four genes showing significant heritable expression variation, but no support for the second, third, or fourth. It is therefore concluded that checkpoint gene expression level does not predict oxidative stress sensitivity. However, an exploratory genotype-based scan identified a specific genomic region near *hpr-17* with a robust, replicable association with paraquat response, revealing a candidate locus for further investigation.

METHODS AND MATERIALS

Data Sources and Accession

RNA-seq raw count data for 208 genetically diverse wild *C. elegans* isotypes, 609 samples in total, were obtained from the National Center for Biotechnology Information's Gene Expression Omnibus, accession GSE186719 (23). Whole-genome genotype data were obtained from the CaeNDR, specifically the January 2021 release in two formats: a hard-filtered isotype VCF (Variant Call Format) with 29,192 variants and 540 isotypes, and an imputed isotype VCF (2,919,294 variants and 540 isotypes). Telomere length estimates for 152 wild isotypes were obtained from Cook *et al.* (22). Oxidative stress phenotype data were obtained from the CaeNDR Public Phenotype Database (14), which was measured as normalized median nematode body length after exposure to 250 µM paraquat, a compound that generates oxidative damage through redox cycling. This trait is normalized based on each strain's median body length under non-paraquat control conditions, so that the reported value reflects the paraquat-specific effect on development after removing strain-level baseline size differences. Lower, more negative values, indicate a greater reduction in body length relative to control, and therefore greater developmental impairment and more sensitivity to oxidative stress. Higher, more positive values, indicate a smaller reduction in body length relative to control, and therefore greater resistance. This direction is applied consistently throughout. Strain grouping and isotype information were obtained from Cook *et al.* (22).

Gene Identification, Transcript Extraction, and Expression Normalization

Four target genes and their WormBase sequence names were identified by querying the WormBase gene pages from release WS276 (24): *mrt-2* (WBGene00003417; sequence name Y41C4A.14), *hus-1* (WBGene00002042; H26D21.1), *hpr-17* (WBGene00001998; F32A11.2), and *atm-1* (WBGene00000227; Y48G1BL.2). The annotated transcript isoforms corresponding to each gene in the expression matrix were *mrt-2* (Y41C4A.14.1, Y41C4A.14.2), *hus-1* (H26D21.1.1, H26D21.1.2), *hpr-17* (F32A11.2.1, F32A11.2.2), and *atm-1* (Y48G1BL.2a.1, Y48G1BL.2a.2, Y48G1BL.2b.1).

Raw counts were normalized for library size using counts per million (CPM). For each of the 609 samples, the library size was calculated as the sum of raw counts across all 65,330 transcripts, and each transcript count was divided by that sample's library size and multiplied by 1×10^6. Low-expression transcripts were then removed, keeping only those with CPM ≥ 1 in at least 50% of the 609 samples, reducing the transcript set from 65,330 to 17,420. Library sizes were computed from the complete unfiltered count matrix so that the normalization denominator would not be affected by filtering. Gene-level expression for each target gene was then calculated by summing CPM values across that gene's annotated isoforms listed above, and the summed values were $\log_2(x + 1)$-transformed. Biological replicates were then averaged within each strain to generate a single expression value per gene per strain, producing a final 208-strain × 4-gene expression matrix on the $\log_2(\text{CPM} + 1)$ scale. All these analyses were performed in Python 3 with pandas, numpy, scipy, scikit-learn, and matplotlib.

Batch and Condition Metadata

Sequencing batch (pool) was not used as a covariate in the expression analyses. An inspection of the deposited

sample metadata showed that biological replicates from 194 of the 208 strains were distributed across two or three distinct sequencing pools. This was consistent with the original study's experimental design of processing replicates from each strain in separate batches to minimize batch-strain confounding (23). The remaining 14 strains had all replicates sequenced within a single pool and none of these strains were among the strains highlighted in subsequent descriptive or predictive analyses (N2, BRC20263, ED3040, JU1934).

The developmental stage and growth conditions were standardized by the original experimenters: all animals were synchronized to the young-adult stage and harvested at a reproducible developmental point (first embryo-laying event), verified computationally via transcriptome-based age prediction (23). Growth temperature (20°C), food source (*E. coli* OP50), and culture medium (Nematode Growth Medium with Agar) were held constant across all the strains. Since these variables were either uniformly applied or deliberately distributed across strains by design, additional batch correction or condition adjustment was not performed in this reanalysis.

Strain Harmonization and Dataset Intersection

Strain names were matched across the different datasets by first extracting the isotype prefix from each RNA-seq sample column. Then, only the strains that appeared in all the required datasets were kept for each analysis. This was done by computing pairwise and multi-way intersections of strain sets across all datasets. The main analysis set (expression, genotype, paraquat phenotype) contained 99 strains, the telomere-length analysis contained 120 strains (overlap between expression and telomere data), and the full 4-way intersection (all datasets) contained 59 strains. File S01 from Cook *et al*. (22), which covers the older 2016 strain collection, was not used since it was based on an older strain collection and was found not to include many of the newer ECA-series isolates that were found in the 2021 datasets.

Descriptive Expression Analysis and Heritability Estimation

Inter-strain expression variation was found by computing per-gene median, mean, and coefficient of variation (CV) across the 208 strains. Broad-sense heritability (H^2) was estimated for each gene using a one-way Analysis of Variance (ANOVA) variance-components approach:

$$H^2 = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}} \quad \text{(Eq. 1)}$$

where $\sigma^2_{between}$ and σ^2_{within} were estimated from a one-way ANOVA, using the mean squares for differences between strains and differences among the replicates in the same strain. Since the number of replicates differed across the strains, an unbalanced-design approximation using the weighted mean group size was applied. Before estimating the heritability, CPM-normalized expression values were transformed using $\log_2(x + 1)$ in order to reduce the effect of the very large values and make the overall variance more stable (Gene Identification, Transcript Extraction, and Expression Normalization).

Co-Expression Analysis

Spearman's rank correlation was found between each pair of the four genes using CPM-normalized, $\log_2$-transformed average expression values for each strain (Gene Identification, Transcript Extraction, and Expression Normalization). This produces a 4 × 4 correlation matrix describing how strongly the expression levels of the genes were correlated to one another across all 208 strains.

***Cis*-eQTL Mapping**

Cis-eQTL mapping was performed using the hard-filtered isotype VCF file, including 204 strains that contained both gene expression data and genotype data. For each gene, the *cis*-region was defined as the body of the gene plus 1 megabase (Mb) upstream and downstream of the gene using gene coordinates obtained from Wormbase WS276 (24): *mrt-2* (III:11,746,766–11,750,175), *hus-1* (I:3,698,102–3,699,300), *hpr-17* (II:13,153,464–13,157,047), and *atm-1* (I:183,364–208,468).

Genotypes were converted into dosage values, where 0/0 = 0, 0/1 = 1, and 1/1 = 2. Missing genotype values were replaced using the mean allele dosage for each variant. Variants with a minor allele frequency below 5% were then removed. For each gene, linear regression was used to test the association between expression level and each variant within the *cis*-window. To determine the significance, the strain expression labels were shuffled 1,000 times to create a null distribution, and then the observed top-associated single-nucleotide polymorphism (SNP) was compared to this null distribution.

To account for how genetically related the population structure is, a genome-wide genomic relatedness matrix (GRM) was found using 9,964 quality-passing variants from the hard-filtered VCF. Principal component analysis

was then performed, and the top 10 genetic principal components were used to correct both gene expression and the values of the genotype before proceeding to association testing. Principal component 1 explained 18.7% of genetic variation in the 204-strain dataset. All the reported *cis*-eQTL results gathered are from this population-structure-corrected analysis.

Telomere-Independence Test

To test whether the checkpoint gene expression was independent of the length of the telomere,the genomic locations of the four target genes were compared to the published *pot-2* telomere-length QTL confidence interval on chromosome II (12.9 to 15.3 Mb). This was done to check whether any of the target genes physically overlapped with the telomere-length region that was reported by Cook *et al*. (22). Direct regression of each gene's $\log_2$-transformed expression level was then compared with telomere length across the 120 strains with both expression and telomere data. Pearson correlation and simple linear regression were used to test whether higher or lower expression of each of the genes was associated with longer or shorter telomeres.

Oxidative-Stress Vulnerability Modeling

The main analysis used 99 strains that had both checkpoint gene expression and paraquat response data (Data Sources and Accession; lower/more negative values indicate greater oxidative stress sensitivity). Two models were trained to predict paraquat response using checkpoint gene expression through four features: the CPM-normalized, $\log_2$-transformed expression values of *mrt-2, hus-1, hpr-17,* and *atm-1* (Gene Identification, Transcript Extraction, and Expression Normalization). These features were not additionally standardized before model fitting, as all four genes were already on a common log2-CPM scale following this normalization. The first was a Ridge regression model (scikit-learn RidgeCV) with the regularization parameter α selected by internal cross-validation over 50 log-spaced values from 10^{-3} to 10^{3}. The second was a gradient-boosted tree ensemble (XGBRegressor) with 100 estimators, a maximum tree depth of 2, a learning rate of 0.05, a subsample fraction of 0.8, and a fixed random seed of 42. All the remaining parameters were left at package defaults. Both these models were evaluated by leave-one-out cross-validation (LOOCV), and performance was assessed using LOOCV R^2, mean absolute error (MAE), and the Pearson correlation between the observed and LOOCV-predicted values. Model performance was then compared to a null baseline that predicted the cohort mean for all strains. Feature attribution for the fitted XGBoost model was found using Shapley Additive exPlanations (SHAP) values via the SHAP Explainer across all 99 training samples. Since the model did not outperform the null baseline, these values are only reported as a description of fitted-model behavior and are not interpreted biologically.

Genotype-based association (exploratory). For each of the four gene-region *cis*-windows (±1 Mb), all SNPs passing the MAF $\geq$ 5% filter were directly tested for their association with paraquat response using the same vectorized linear regression and 1,000-permutation significance threshold as the *cis*-eQTL analysis. Both significant hits were then re-tested after applying the same 10-principal-component kinship correction that was used in Co-Expression Analysis, with the GRM derived separately for the 99-strain subset (Principal component 1 = 32.3% variance explained). Finally, significant loci hits from the hard-filter VCF analysis were re-tested using the denser imputed VCF (2,919,294 variants) to determine whether the association signal remained stable at higher marker density and a finer genomic resolution.

Variant Annotation

Significant SNPs were annotated using the SnpEff ANN field in the INFO column of the hard-filtered VCF file. This annotation was used to determine whether each of the SNPs was located in a gene, the intron, the exon, or an intergenic region. To identify nearby genes, each hit position was manually inspected in the WormBase JBrowse2 genome browser with a ±10 kb window around the SNP.

Per-Strain Expression Comparison vs. N2

To provide descriptive context, the CPM-normalized, $\log_2$-transformed expression of each checkpoint gene in each of the 207 non-N2 strains were compared to the corresponding expression value in the N2 reference strain, and the direction and magnitude of the expression difference were calculated. Since most of the strains were represented by only 1–3 biological replicates, per-strain variance could not be reliably estimated. Additionally, formal statistical frameworks for differential expression (e.g., DESeq2, edgeR) requires replicate structure and experimental metadata (batch assignment per group, sequencing depth per group) which are not available at the per-strain level in this reanalysis. This comparison is therefore reported as a descriptive ranking of expression differences relative to N2, without a formal statistical

test or significance threshold. Therefore, the differences that are reported should not be interpreted as evidence of statistically significant differential expression.

RESULTS/ANALYSIS

Significant Inter-Strain Expression Variation in Three of Four Checkpoint Genes

After CPM normalization and low-count filtering were performed (Gene Identification, Transcript Extraction, and Expression Normalization), three of the four candidate genes showed significant expression differences between strains based on one-way ANOVA (*mrt-2:* $F = 1.92$, $p = 1.6\times10^{-8}$; *hus-1:* $F = 1.55$, $p = 1.2\times10^{-4}$; *hpr-17:* $F = 2.80$, $p = 5.8\times10^{-19}$). However, *atm-1* did not reach the statistical significance threshold and was then excluded from subsequent gene-level interpretation, but kept in the predictive models for completeness. Broad-sense heritability estimates indicated that *hpr-17* showed the highest heritable component ($H^2 = 0.381$), followed by *mrt-2* ($H^2 = 0.239$), *hus-1* ($H^2 = 0.157$), and *atm-1* ($H^2 = 0.032$). The three significant genes remained positively co-expressed across the strains (Spearman r = 0.31 to 0.81; Table 2), which is consistent with coordinated

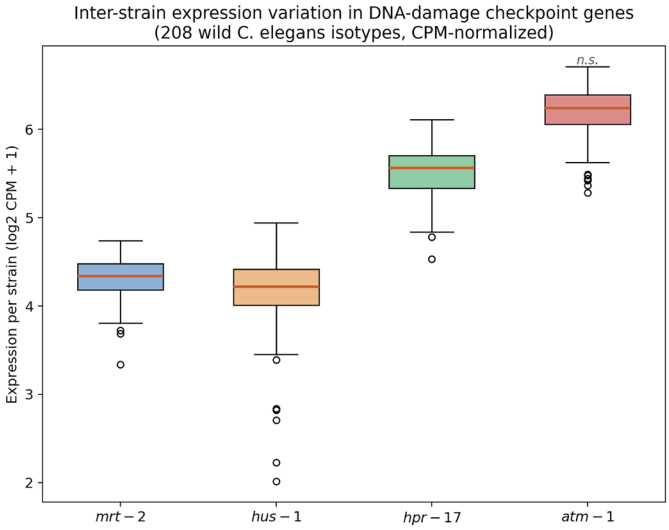

Figure 1. *Inter-strain expression distributions of mrt-2, hus-1, hpr-17, and atm-1 across 208 wild C. elegans isotypes. Boxes represent the interquartile range (IQR). The horizontal line within the box indicates the median. Whiskers extend to 1.5× IQR. The open circles indicate outlier strains. Expression is shown on the $\log_2$(CPM+1) scale after normalization. Although atm-1 ranks highest by median expression, it did not show significant inter-strain variation after normalization (Table 1).*

Table 1. *Descriptive statistics, one-way ANOVA results, and broad-sense heritability estimates for the four DNA damage checkpoint genes across 208 wild C. elegans isotypes after performing library-size normalization (counts per million) and low-count filtering (Gene Identification, Transcript Extraction, and Expression Normalization). Median and CV are reported on the $\log_2$ (CPM+1) scale. atm-1 did not reach the threshold for statistical significance after normalization (n.s. stands for not significant).*

Gene	N strains	Median (log2 CPM)	CV	H^2	ANOVA F	ANOVA p
mrt-2	208	4.341	0.053	0.239	1.92	1.6×10^{-8}
hus-1	208	4.216	0.095	0.157	1.55	1.2×10^{-4}
hpr-17	208	5.564	0.048	0.381	2.80	5.8×10^{-19}
atm-1	208	6.239	0.044	0.032	1.10	0.217 (n.s.)

Table 2. *Pairwise Spearman rank correlation coefficients for checkpoint gene expression across the 208 wild C. elegans isotypes, computed on CPM-normalized, log2-transformed values. All pairwise correlations are positive and statistically significant (Spearman's ρ test, all $p < 0.001$).*

Gene	*mrt-2*	*hus-1*	*hpr-17*	*atm-1*
mrt-2	1.00	0.81	0.79	0.44
hus-1	0.81	1.00	0.70	0.31
hpr-17	0.79	0.70	1.00	0.59
atm-1	0.44	0.31	0.59	1.00

regulation being important components of the same 9-1-1 checkpoint complex. Together, these results support the first hypothesis by demonstrating that three of the four checkpoint genes exhibit significant, heritable inter-strain expression variation.

As a descriptive comparison (Variant Annotation), CPM-normalized expression of each checkpoint gene in each of the 207 non-N2 strains was ranked by the magnitude of its difference from the N2 reference strain. No formal significance testing was performed, given the limited per-strain replicate structure available in this dataset. BRC20263 and ED3040 consistently ranked among the strains with the largest positive differences from N2 across multiple genes, while JU1934 consistently ranked among the strains with the largest negative differences. These patterns provide descriptive context, and are not evidence of statistically significant differential expression.

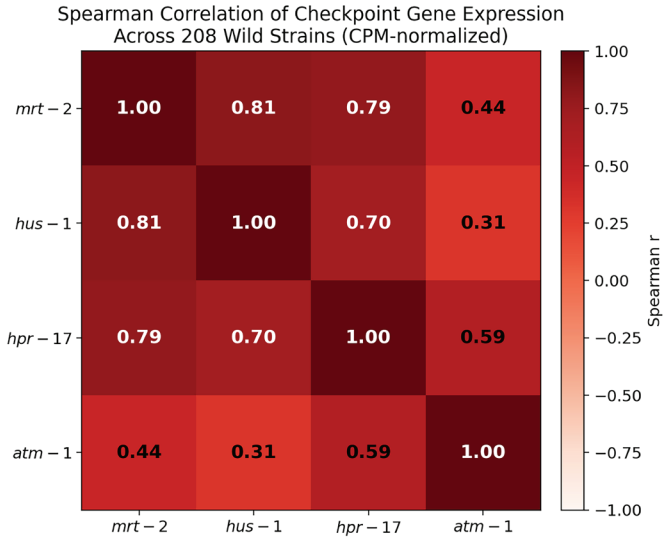

Figure 2. *Spearman rank correlation heatmap of the checkpoint gene expression across 208 wild C. elegans isotypes, computed using CPM-normalized, log_2 -transformed values. All six pairwise correlations are positive (r = 0.31 to 0.81) and statistically significant, which is consistent with the coordinated regulation of genes that are participating in the conserved DDR pathway. However, atm-1 shows a comparatively weaker co-expression with the other three genes. The color intensity indicates the Spearman r value.*

No Significant *Cis*-eQTL After Kinship Correction

Cis-eQTL mapping was performed using windows within ±1 Mb on each of the four genes. This analysis included 204 strains with both expression and genotype data. After filtering variants containing a minor allele frequency below 5%, each gene region contained between 47 to 146 SNPs.

After correcting for population structure using the top 10 principal components, it was found that no gene showed a statistically significant local association between the nearby genetic variants and gene expression. The permutation p-values were 0.901 for *mrt-2,* 0.542 for *hus-1,* 0.132 for *hpr-17,* and 0.622 for *atm-1.* Although *hpr-17* had the lowest p-value and showed the strongest suggestive signal, it did not reach statistical significance after permutation correction.

The top nearby SNP for *hpr-17* had a kinship-corrected nominal p-value of 0.0015, which suggests a possible weak local genetic effect. However, since the corrected permutation p-value was not significant, this result should be interpreted only as a trend. Overall, these findings indicate that the checkpoint gene expression differences are not explained by common nearby *cis*-variants. Instead, the expression differences could be influenced more by distant regulatory variants, rare variants that were not included after filtering, and non-genetic factors. These results do not support the second hypothesis that checkpoint gene expression is primarily regulated by common *cis*-acting variants.

Table 3*. Kinship-corrected cis-eQTL mapping results for the four checkpoint genes. Cis-windows are defined as the gene body ±1 Mb. SNPs tested reflect variants that pass the MAF ≥ 5% filter. Permutation p-values are based on 1,000 label permutations. There is no gene that reaches significance after correction (permutation p < 0.05).*

Gene	SNPs tested	Best SNP position	Kinship-corr. nominal p	Permutation p	Significant?
mrt-2	98	III:12,277,604	0.056	0.901	No
hus-1	47	I:4,247,579	0.039	0.542	No
hpr-17	146	II:13,005,950	0.0015	0.132	No (suggestive)
atm-1	47	I:233,628	0.059	0.622	No

Expression Variation is Independent of Published Telomere-Length QTL

Direct regression of each gene's expression was tested against published TelSeq-derived telomere length estimates from Cook *et al.* (22). The results revealed no significant relationship for any of the genes (*mrt-2*: r = −0.112, p = 0.222; *hus-1*: r = 0.018, p = 0.845; *hpr-17*: r = 0.068, p = 0.461; *atm-1:* r = 0.119, p = 0.195). Notably, *hpr-17* is physically located within the published *pot-2* telomere-length QTL confidence interval (chromosome II, 12.9–15.3 Mb; 22), however, its expression shows no significant correlation with telomere length. This indicates that *hpr-17* expression variation is not the mechanism by which the *pot-2* locus is able to exert its effect on telomere length. The other three genes lie on different chromosomes from the telomere QTL interval and can be considered independent of that locus because of their positions not physically overlapping with the published telomere-length QTL region. These results do not support the third hypothesis as checkpoint gene expression does not correlate with telomere length for any of the four genes.

Checkpoint Gene Expression Does Not Predict Oxidative Stress Sensitivity

To test whether checkpoint gene expression could accurately predict oxidative stress vulnerability, predictive models were trained with the paraquat-induced developmental impairment set as the outcome. The 99 strains that had both gene expression and paraquat sensitivity data were used for the analysis. The four input features were the $\log_2$-transformed expression levels of *mrt-2, hus-1, hpr-17*, and *atm-1.*

Neither the Ridge regression model nor the XGBoost ensemble model showed meaningful predictive capabilities. The Ridge model explained essentially

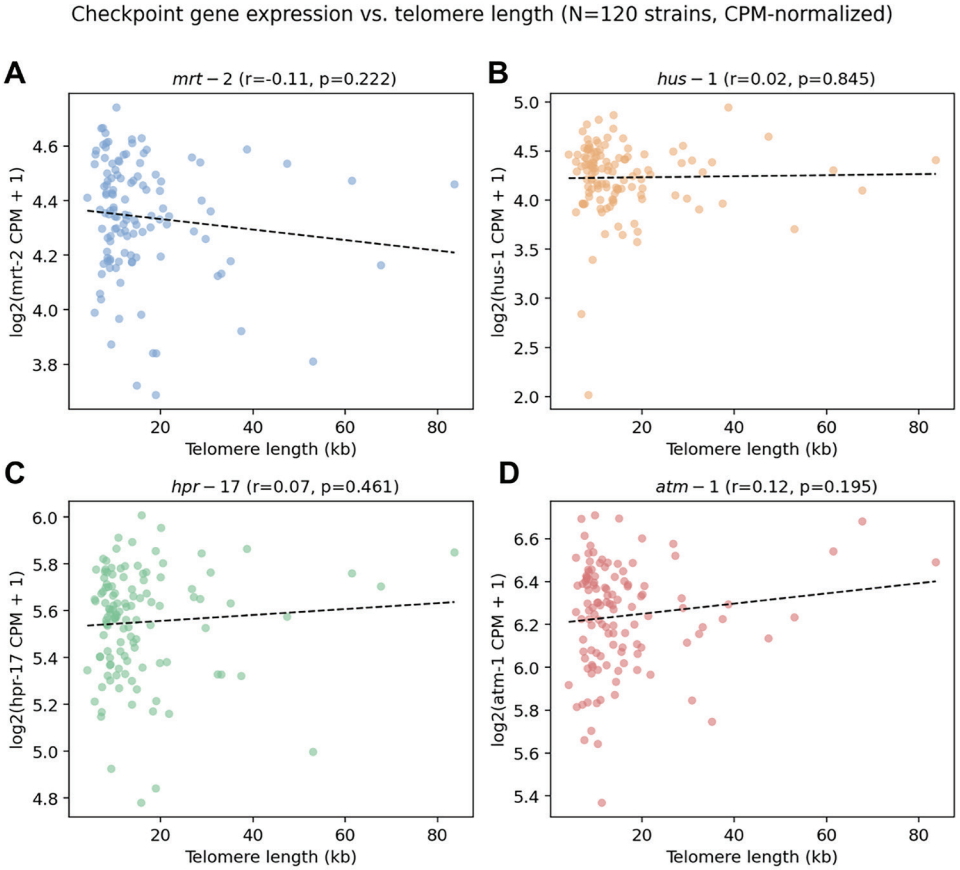

Figure 3. *Checkpoint gene expression versus telomere length across the 120 wild C. elegans isotypes with both expression and telomere data. The dashed line shows a linear regression fit. No gene exhibits a significant relationship (all p > 0.15). The gene hpr-17 is physically located within the published pot-2 telomere-length QTL confidence interval (chrII: 12.9–15.3 Mb), but it demonstrates no expression and has no telomere correlation.*

no variation in paraquat sensitivity (LOOCV R^2 = 0.019, MAE = 21.08, Pearson r = 0.147, p = 0.147). The XGBoost model performed worse than the null baseline (R^2 = −0.037, MAE = 21.81, r = 0.197, p = 0.051). A null baseline model, which predicted the cohort mean for all strains, gave an MAE of 20.98. Overall, these results indicate that checkpoint gene expression by itself is unable to explain the differences in oxidative stress vulnerability across strains. These results do not support the fourth hypothesis, as checkpoint gene expression does not predict oxidative stress sensitivity.

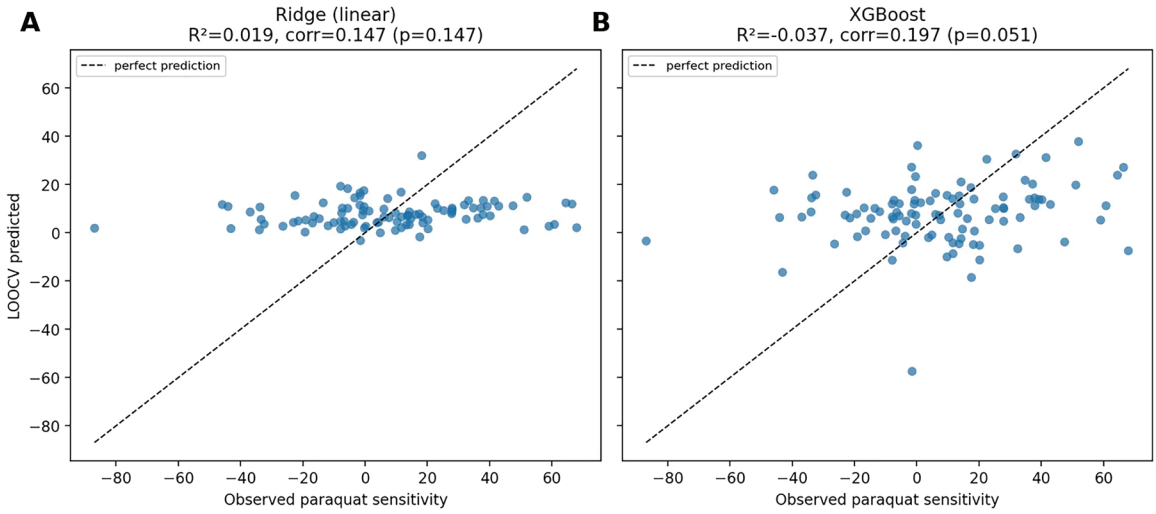

Figure 4. *LOOCV predicted versus observed paraquat sensitivity for 99 wild C. elegans strains using CPM-normalized checkpoint gene expression as the predictor variables. The dashed line represents the perfect prediction (y=x). Neither of the two models exhibited meaningful predictive performance, with Ridge regression explaining little of the variation observed (LOOCV R^2 = 0.019) and XGBoost performing worse than the null model (LOOCV R^2 = −0.037). These results show that the natural variation in checkpoint gene expression is not predictive of paraquat-induced oxidative stress susceptibility across wild C. elegans strains. Higher values on both axes indicate greater retained body length (relative resistance); lower values indicate greater developmental impairment (relative sensitivity).*

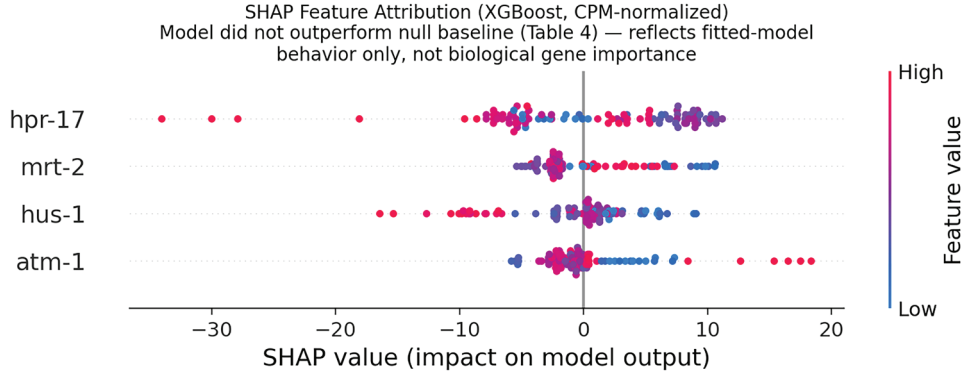

Figure 5. *SHAP feature-attribution values from the fitted XGBoost model across all 99 training strains using CPM-normalized checkpoint gene expression features. Each point represents one strain, the x-axis shows the SHAP value (impact on model output), and the color gradient indicates feature value (pink/red = high expression, blue = low expression). hpr-17 exhibits the highest mean absolute SHAP value. However, because the underlying XGBoost model was unable to outperform the null baseline (Table 4), these attribution values reflect the behavior of the fitted model rather than act as reliable biological feature importance. Therefore, these results should not be interpreted as evidence that hpr-17 is a biologically important predictor of oxidative stress susceptibility.*

Table 4. *LOOCV performance of Ridge regression and XGBoost models predicting paraquat response (N = 99 strains) using CPM-normalized expression features. Neither model is able to explain meaningful variance relative to a null baseline, which predicted the cohort mean for all strains. XGBoost hyperparameters match the original raw-count analysis exactly (100 estimators, max depth of 2, learning rate of 0.05, and seed 42).*

Model	LOOCV R^2	MAE	Pearson r	p-value	Beats null?
Ridge regression	0.019	21.08	0.147	0.147	No
XGBoost	−0.037	21.81	0.197	0.051	No
Null (predict mean)	—	20.98	—	—	Reference

Exploratory Genotype Scan Identifies a Robust Locus Near *hpr-17*

To test whether local genetic variants, rather than the checkpoint gene expression levels, were associated with oxidative stress response, SNPs detected within ±1 Mb of the target gene were tested for their association with paraquat response. This analysis used 99 strains that had both the genotype and paraquat phenotype data.

Two of the four gene-region windows showed significant associations after permutation correction: the *mrt-2* window (strongest SNP at III:11,013,892, β = 10.87, nominal p = 7.8×10^{-5}, permutation p = 0.005) and the *hpr-17* window (strongest SNP at II:13,787,889, β = 11.19, nominal p = 1.2×10^{-4}, permutation p = 0.016). For both loci, strains carrying the alternate allele showed substantially greater resistance to paraquat than non-carriers, retaining more of their expected body length under exposure: mean normalized length residual +21.5 versus −0.2 (*mrt-2* window, 36/99 carriers) and +23.5 versus +1.1 (*hpr-17* window, 29/99 carriers). Since the phenotype is a control-normalized length residual in which more negative values indicate greater developmental impairment (Data Sources and Accession), the positive effect sizes at both loci (β = 10.87 and 11.19) provide evidence that each additional alternate-allele dose is associated with higher resistance rather than increased sensitivity. The *hus-1* and *atm-1* windows showed no significant association (permutation p = 0.237 and 0.991, respectively).

The two significant loci were then re-tested after applying the same kinship correction used for *cis*-eQTL mapping (top 10 Principal Components of an N = 99 GRM; Principal Component-1 explaining 32.3% of genetic variance). Both significant associations remained after the correction: *mrt-2* window (corrected r = 0.287, p = 0.0040); *hpr-17* window (corrected r = 0.356, p = 0.0003). However, the signal was weaker after the correction, suggesting that a part of the association may be influenced by the genetic similarities and relatedness among the strains. Since the signal did not disappear and was also confirmed using the denser imputed VCF dataset, the *hpr-17* region appears as a promising and robust candidate locus that is associated with oxidative stress response. In *hus-1* and the *atm-1* regions, no significant associations were identified.

Table 5. *Genotype-paraquat response association results for all of the four gene-region cis-windows (±1 Mb, N = 99 strains). β is the effect size in normalized paraquat length-residual units per alternate-allele dose; positive values indicate that the alternate allele is associated with greater retained body length, which indicates a greater resistance to paraquat (Data Sources and Accession). Only windows with a significant permutation p-value in the hard-filter analysis are shown with their full statistics. hus-1 and atm-1 windows were not significant (permutation p = 0.237 and 0.991, respectively). Kinship correction applied through the top 10 PCs of the genome-wide GRM. Significant loci were subsequently evaluated in the denser imputed VCF with 2,919,294 variants to evaluate positional consistency. n.s. indicates not significant.*

Gene window	SNPs tested	Best SNP	β	Uncorrected r (p)	Kinship-corr. r (p)	Imputed best position	Imputed perm. p
mrt-2	91	III:11,013,892	10.87	0.386 (0.0001)	0.287 (0.0040)	III:10,759,700	0.031
hus-1	42	—	—	—	—	—	0.237 (n.s.)
hpr-17	143	II:13,787,889	11.19	0.376 (0.0001)	0.356 (0.0003)	II:13,787,889	0.119
atm-1	37	—	—	—	—	—	0.991 (n.s.)

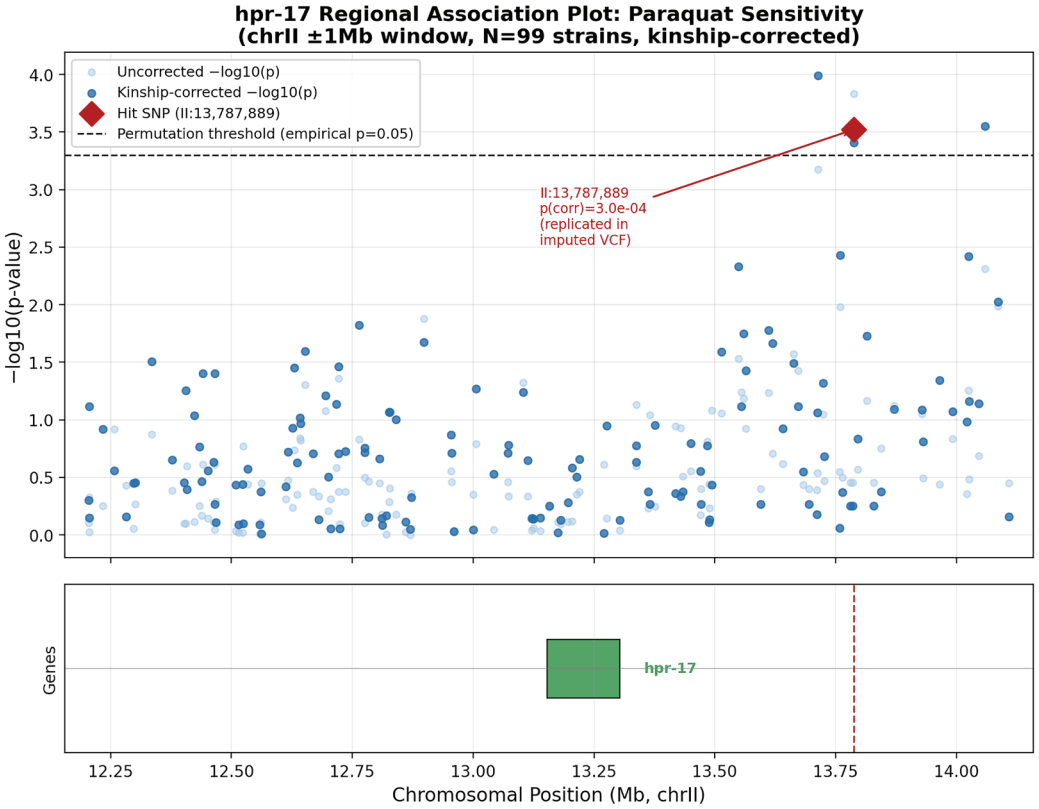

Figure 6. *Regional association plot for the hpr-17 genomic window (chrII ±1 Mb, N = 99 strains). Each point represents one SNP that is tested for association with paraquat response, with $-\log_{10}$(p-value) on the y-axis. The light blue points show uncorrected association and dark blue points show associations after kinship correction using the top 10 principal components of the genome-wide genomic relatedness matrix (GRM). The red diamond marks the hit SNP (II:13,787,889; kinship-corrected $r = 0.356$, $p = 3.0\times10^{-4}$), which is independently replicated at the identical position inside the imputed VCF when the analysis was repeated with the denser imputed variant dataset. The dashed horizontal line indicates the permutation-based significance threshold (empirical p = 0.05, 1,000 label permutations). A small number of additional variants in the window also exceed this threshold, and because these fall within the same associated interval as the lead SNP and have no functional annotation that distinguishes them, only the hit SNP is reported and no independent secondary signal is reported. The gene track below depicts the hpr-17 gene body position, and the red dashed vertical line marks the hit SNP position.*

To assess the robustness and genomic localization of the associations identified, the analysis was repeated with the denser imputed VCF that contained 2,919,294 variants, providing a substantially greater marker density compared to the hard-filtered dataset. The *hpr-17* hit replicated at the identical genomic position (II:13,787,889; permutation p = 0.119 in the imputed analysis), confirming its positional stability across variant sets separated by an approximately 85-fold increase in marker density inside the tested window (143 versus 12,166 variants). The nominal p-value and effect size were also unchanged (β = 11.19). However, the *mrt-2* hit did not replicate at the original position under fine-mapping. The strongest SNP shifted 254 kb away to position III:10,759,700 (permutation p = 0.031), placing it 987 kb from the *mrt-2* gene body, near the outer edge of the tested window. This shift results in the *mrt-2* signal being less convincing as a precisely localized local association.

SnpEff functional annotation of the original hard-filter hit positions showed that the *mrt-2*-window SNP (III:11,013,892) is an intronic variant located within an uncharacterized gene (WBGene00012967; Y48A6B.6). The *hpr-17*-window hit (II:13,787,889) did not have any transcript-level annotation, which is consistent with its location in a gene-sparse intergenic region. No nearby annotated gene containing a known DNA damage response or oxidative stress function was found at either

position.

Overall, the locus near the *hpr-17* gene was the strongest candidate signal discovered in this study because it survived kinship correction, appeared at the same genomic position in the denser, imputed dataset, and lies within the same broader genomic region, making *hpr-17* a compelling candidate for follow-up research. Since *hpr-17* also showed the highest expression heritability and a suggestive *cis*-eQTL trend, this region may be a biologically relevant candidate to investigate in future studies.

DISCUSSION

This study used population-scale genomic and transcriptomic data from 208 wild *C. elegans* isotypes to investigate whether natural variation in four highly conserved 9-1-1 DNA damage checkpoint genes: *mrt-2*, *hus-1*, *hpr-17*, and *atm-1*, predicts oxidative stress vulnerability in relation to neurodegeneration and ALS pathology. The findings support the first hypothesis: checkpoint gene expression varies significantly for three of the four genes and has a modest heritable component. However, the second, third, and fourth hypotheses were not supported, as gene expression was not explained by local genetic variants, did not correlate with telomere length, and did not meaningfully predict inter-strain differences in oxidative stress sensitivity. Instead, an exploratory genotype-level association analysis identified a genomic region near *hpr-17* that remained associated with paraquat response after population-structure correction and exhibited positional stability when tested with a much denser imputed variant dataset. This represents the strongest and most novel finding in this study. Together, these results suggest that the mechanism linking natural checkpoint gene variation to stress vulnerability functions through coding or regulatory variants that affect protein function rather than through bulk transcriptional output, and identifies *hpr-17* as the highest-priority gene for future investigation.

Expression Variation Without *Cis*-Genetic Drivers

Three of the four checkpoint genes (*mrt-2*, *hus-1*, and *hpr-17*) exhibit significant inter-strain expression variation, with ANOVA p-values being well below 0.01, and broad-sense heritability estimates ranging from $H^2 = 0.157$ to 0.381. *atm-1* did not show significant inter-strain variation after CPM normalization ($H^2 = 0.032$), which suggests that its apparent significance under raw-count analysis was driven by sequencing-depth artifacts rather than genuine variation in a biological context; specifically, the failure mode that this reanalysis was designed to detect. This establishes that variation in the remaining three genes is not technical noise, but reflects a detectable genetic component that contributes to expression differences across strains. However, the absence of any significant *cis*-eQTL after kinship correction indicates that common local DNA variants are unable to explain this heritable component. The pattern identified, heritable expression variation without a local genetic driver, is common in population transcriptomics. *Trans*-regulatory variation (distal loci affecting expression of multiple genes simultaneously) frequently accounts for a larger proportion of expression heritability than *cis*-variation in *C. elegans* wild isolates, as consistent with the results in the original Zhang *et al*. study from which the expression dataset used in this study was obtained (23).

The strong positive co-expression among *mrt-2, hus-1, and hpr-17* (Spearman r = 0.70 to 0.81; Table 2) further supports the idea that these genes may be regulated by shared upstream mechanisms or coordinated processes on the pathway level rather than by independent *cis*-regulatory variants at each locus. In contrast, co-expression involving *atm-1* was much weaker (r = 0.31–0.59), consistent with its lack of significant inter-strain variation. Future *trans*-eQTL mapping across the full genome would be required to identify the specific distant loci driving this coordinated expression variation.

The heritability estimates for the three significant genes ($H^2 = 0.157$ to 0.381) also help explain why expression-based prediction of paraquat sensitivity was weak. With only 16–38% of expression variance in these genes attributable to stable genetic differences among strains, the remaining variance likely reflects a combination of environment, technical, and stochastic biological factors that were not genetically determined in this dataset. The poor predictive performance of the expression-based models is therefore more consistent with the limited predictive signal in the expression data available, rather than with deficiencies in the specific modeling methods used.

Expression Independence from Telomere-Length Variation

The finding that checkpoint gene expression does not correlate with telomere length is especially informative for *hpr-17,* which is physically located within the published *pot-2* telomere-length QTL confidence interval (22). This positional overlap could have suggested

that *hpr-17* expression mediates the effect of the *pot-2* locus on telomere length. However, the absence of any association between *hpr-17* expression and telomere length does not support this hypothesis. The *pot-2* locus was previously demonstrated to act through a specific coding variant (F68I in the OB-fold domain of POT-2), not through transcriptional variation (22). The expression analysis conducted is consistent with this mechanism, as the telomere QTL appears to function at the protein level rather than through altered *hpr-17* or nearby-gene expression. This result also reinforces a broader point that was established by Cook *et al*. (22), that telomere-length variation in wild *C. elegans* does not correlate with fitness traits such as offspring production or longevity of the nematodes, and should therefore not be assumed to be a substitute for general genomic instability or organismal stress resistance in the nematodes. This study extends this observation by providing further evidence that checkpoint gene expression, the transcriptional output of the pathway that controls telomere maintenance, is equally uncorrelated with telomere length across strains.

Genotype Rather Than Expression Predicts Oxidative Stress Response

The contrast between the negative expression-based prediction results and the stronger genotype-based association near *hpr-17* is one of the major findings of this study. Baseline checkpoint gene expression alone was unable to meaningfully predict paraquat response, with expression-based models showing little to no predictive capabilities. In contrast, the genotype-based association analysis identified a local variant near *hpr-17* that remained associated with paraquat response after correction for population structure. Notably, the direction of this association provides evidence that the alternate allele could help worms resist paraquat damage, making it a possible natural resistance variant.

These two findings are not contradictory to each other. Expression level reflects the steady-state transcriptional output of a gene that is averaged across tissues and developmental stages in whole-animal RNA-seq, therefore, it could miss detecting functional differences that occur only in specific cell types, at certain times, or after stress exposure. A genetic variant could also affect stress response by altering protein functions, splicing, or local regulation without resulting in a large change in the overall transcript abundance. In that case, the variant would only be detectable in a genotype-based association test but undetectable in an expression-based prediction model.

The *hpr-17* region results are consistent with this interpretation. The association found appears to reflect a local genetic variant or a nearby linked variant, instead of a difference in *hpr-17* transcript abundance itself. This pattern is also consistent with past findings in human genetics, where disease-associated variants are often not aligned with *cis*-eQTL signals for the gene that is closest to it (25). Instead, they may act through other mechanisms such as protein function, alternative splicing, cell-type-specific regulation, or stress-dependent effects that are unable to be captured by bulk RNA-seq. Overall, these results indicate that natural checkpoint gene variation in *C. elegans* can influence oxidative stress response through local genetic effects rather than only through baseline expression level.

The *hpr-17*-Region Locus: Robustness and Limitations

The *hpr-17*-region finding satisfies several criteria corroborated in similar, past experiments, supporting its robustness. The association remained significant after kinship correction for population structure, it independently replicated at the exact same genomic position (II:13,787,889) when retested with an approximately 100-fold denser genome-wide (2,919,294 versus 29,192 variants), and it is consistent with *hpr-17*'s elevated heritability and suggestive *cis*-eQTL signal across independent analyses. The convergence of *hpr-17* across multiple methods of analysis, the highest heritability (Significant Inter-Strain Expression Variation in Three of Four Checkpoint Genes), the strongest *cis*-eQTL suggestive signal (No Significant Cis-eQTL After Kinship Correction), physical location within the telomere QTL interval (Expression Variation is Independent of Published Telomere-Length QTL), and the only robustly replicated genotype-paraquat hit (Exploratory Genotype Scan Identifies a Robust Locus Near *hpr-17*), is unlikely to be a result of coincidence and suggests that the *hpr-17* region is a compelling candidate that should be prioritized for future research.

However, there are several important limitations. First, the hit SNP (II:13,787,889) falls around 634 kb from the *hpr-17* gene body itself, in an intergenic region that is sparse in genes, includes no transcript-level annotation, and no nearby gene with an established DDR or oxidative-stress function. Therefore, the current data are unable to confirm whether the SNP tags a causal variant in or near *hpr-17* through linkage disequilibrium, which can be extensive in *C. elegans* wild isolates, or whether it only represents a distinct nearby locus that does not directly relate to *hpr-17.*

Second, while the association survived the principal-component-based kinship correction applied here, a full linear mixed model (as implemented in cegwas2/rrBLUP) would provide a more rigorous correction and should be applied in follow-up studies. Third, there was no second independent stress phenotype available for replication in this dataset. If there were a second CaeNDR trait that corroborates the same genomic region, the claim would be strengthened greatly. Fourth, the sample size of 99 isolates is modest for a genetic association study, and the effect sizes reported here ($r \approx 0.36$ after correction) should be interpreted as an approximation until the results are reproduced in a larger panel of strains.

In contrast, the *mrt-2*-region association should be evaluated with more caution: its best-associated position shifted 254 kb under fine-mapping with the imputed VCF and now sits 987 kb from the *mrt-2* gene, at the outer boundary of the tested window. This positional instability suggests that the original signal could reflect residual population structure, a distal unrelated locus, or a less precisely localized association rather than an actual local *mrt-2* proximal effect.

Implications for ALS and DDR Biology

These results reframe the relationship between checkpoint gene variation and ALS-relevant stress biology in *C. elegans*. The original hypothesis, that checkpoint gene expression level has the capabilities to predict proteotoxic or oxidative stress sensitivity analogous to what TDP-1/FUST-1 dysfunction causes in previously established ALS models, was not supported by the study. This suggests that the link between DDR checkpoint variation and neuronal vulnerability, if it exists, is not mediated through a simple expression-level mechanism that could be detected in whole-animal bulk RNA-seq. This is corroborated by biological evidence, as motor neurons and germline cells, the tissues most relevant to ALS and the Mrt phenotype, respectively, represent a small fraction of the whole-organism transcriptome, and tissue-specific effects are weakened in this type of data.

The *hpr-17*-region genotype finding provides a more specific and testable hypothesis for future studies: that a coding or regulatory variant near *hpr-17* controls oxidative stress tolerance, potentially through methods such as altered RAD17 clamp-loader function, checkpoint activation threshold, or an entirely separate biological mechanism in an uncharacterized gene nearby. The link between checkpoint gene biology and ALS-protein toxicity is supported by prior work showing that TDP-43/FUS interactomes are altered by DNA damage (26), loss of *tdp-1* increases the sensitivity of *C. elegans* to oxidative stress, and that checkpoint-deficient strains show temperature-sensitive fertility and genome-stability phenotypes that parallel the stress-sensitization seen in *tdp-1/fust-1* mutants (16, 20, 21). This study adds a new population genetics perspective by identifying a reproducible candidate genomic region near *hpr-17* for future investigation.

CONCLUSION

This study provides the first population-scale characterization of natural variation in 9-1-1 DDR gene expression and genotype across 208 wild *C. elegans* isotypes, testing whether this variation predicts oxidative stress vulnerability as an indicator of ALS-relevant neuronal sensitivity. Three of the four genes studied: *mrt-2*, *hus-1*, and *hpr-17*, showed significant, heritable inter-strain expression variation with positive co-expression across the panel (Spearman $r = 0.31–0.81$). The gene *atm-1* did not show significant inter-strain variation after normalization for sequencing depth. No significant *cis*-eQTL was detected after kinship correction, and expression level did not predict paraquat-induced oxidative stress sensitivity in either the Ridge regression or XGBoost models (LOOCV $R^2 \approx 0$). These results indicate that bulk transcriptional variation in checkpoint genes is not the primary process through which natural DDR genetic diversity is able to influence stress outcomes. However, an exploratory genotype-level scan identified a locus near *hpr-17* whose association with paraquat response survived kinship correction ($r = 0.356$, $p = 0.0003$) and independently replicated at the identical genomic position under fine-mapping with a 100-fold denser imputed variant set. Convergent evidence across four independent analyses consistently implicates *hpr-17* as the highest-priority candidate for follow-up studies, and this work establishes a foundation for future fine-mapping, tissue-specific expression profiling, and transgenic validation to determine whether natural checkpoint gene variation directly influences neurodegeneration-relevant stress biology in *C. elegans* and possibly humans.

Several future projects would directly test and potentially strengthen the conclusions of this study. Fine-mapping of the *hpr-17*-region locus: using additional variants, higher-resolution phenotyping, or conditional analysis, would help localize the causal variant and identify the responsible gene. Testing the association

with a second independent stress phenotype available in CaeNDR (e.g., heat-stress or juglone-sensitivity traits) would provide independent replication evidence. Introducing top candidate strains by *hpr-17*-region genotype into established *TDP-43*/*FUS* transgenic backgrounds would directly test whether the genomic region identified here contributes to ALS-protein-driven neurodegeneration in a whole-organism system, which can be potentially aided by CRISPR/Cas9-based allele editing (27). Finally, single-cell RNA-seq across representative strains would allow tissue-specific expression quantification, which has the potential to reveal neuron or germline-specific expression differences that are not observable in bulk whole-animal data. Together, these further studies would extend the current study's statistical findings toward direct relevance to the biology and mechanisms underlying ALS.

FUNDING SOURCES

The author received no financial support for this research, authorship, and/or publication of this article.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. *Eur J Neurol*. 2020; 27 (10): 1918–1929. doi:10.1111/ene.14393.
2. Brotman RG, Moreno-Escobar MC, Joseph J, Munakomi S, Pawar G. Amyotrophic lateral sclerosis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2017. Available from: https://www.ncbi.nlm.nih.gov/books/NBK556151/ (accessed on 2026-04-03).
3. Mehta P, Raymond J, Nair T, Han M, Berry J, Punjani R, *et al*. Amyotrophic lateral sclerosis estimated prevalence cases from 2022 to 2030, data from the National ALS Registry. *Amyotroph Lateral Scler Frontotemporal Degener*. 2025; 26 (3–4): 290–295. doi: 10.1080/21678421.2024.2447919.
4. Berry JD, Blanchard M, Bonar K, Drane E, Murton M, Ploug U, *et al*. Epidemiology and economic burden of amyotrophic lateral sclerosis in the United States: a literature review. *Amyotroph Lateral Scler Frontotemporal Degener*. 2023; 24 (5–6): 436–448. doi:10.1080/21678421.2023.2165947.
5. Witzel S, Wagner M, Zhao C, Kandler K, Graf E, Berutti R, *et al*. Fast versus slow disease progression in amyotrophic lateral sclerosis — clinical and genetic factors at the edges of the survival spectrum. *Neurobiol Aging*. 2022; 119: 117–126. doi:10.1016/j.neurobiolaging.2022.07.005.
6. Madabhushi R, Pan L, Tsai LH. DNA damage and its links to neurodegeneration. *Neuron*. 2014; 83 (2): 266–282. doi:10.1016/j.neuron.2014.06.034.
7. Konopka A, Atkin JD. DNA damage, defective DNA repair, and neurodegeneration in amyotrophic lateral sclerosis. *Front Aging Neurosci*. 2022; 14: 786420. doi:10.3389/fnagi.2022.786420.
8. Parrilla-Castellar ER, Arlander SJH, Karnitz L. Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair (Amst)*. 2004; 3 (8–9):1009–1014. doi:10.1016/j.dnarep.2004.03.032.
9. Zou L, Cortez D, Elledge SJ. Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev*. 2002; 16 (2): 198–208. doi:10.1101/gad.950302.
10. Al Sultan A, Kirby J, Waller R, Heath P. The genetics of amyotrophic lateral sclerosis: current insights. *Degener Neurol Neuromuscul Dis*. 2016; 6: 49–64. doi:10.2147/DNND.S84956.
11. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A C. elegans mutant that lives twice as long as the wild type. *Nature*. 1993; 366 (6454): 461–464. doi:10.1038/366461a0.
12. Lai CH, Chou CY, Ch'ang LY, Liu CS, Lin W. Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. *Genome Res*. 2000; 10 (5): 703–713. doi:10.1101/gr.10.5.703.
13. Ahmed S, Hodgkin J. MRT-2 checkpoint protein is required for germline immortality and telomere replication in C. elegans. *Nature*. 2000; 403 (6766): 159–164. doi:10.1038/35003120.
14. Crombie TA, McKeown R, Moya ND, Evans KS, Widmayer SJ, LaGrassa V, *et al*. CaeNDR, the Caenorhabditis Natural Diversity Resource. *Nucleic Acids Res*. 2024; 52 (D1): D850–D858. doi:10.1093/nar/gkad887. Available from: https://caendr.org (accessed on 2026-04-03).
15. Kumsta C, Chang JT, Lee R, Tan EP, Yang Y, Loureiro R, *et al*. The autophagy receptor p62/SQST-1 promotes proteostasis and longevity in C. elegans by inducing autophagy. *Nat Commun*. 2019; 10 (1): 5648. doi:10.1038/s41467-019-13540-4.
16. Vaccaro A, Tauffenberger A, Ash PEA, Carlomagno Y, Petrucelli L, Parker JA. TDP-1/TDP-43 regulates stress signaling and age-dependent proteotoxicity in Caenorhabditis elegans. *PLoS Genet*. 2012; 8 (7):

e1002806. doi:10.1371/journal.pgen.1002806.
17. Hope IA, Mounsey A, Bauer P, Aslam S. The forkhead gene family of Caenorhabditis elegans. *Gene*. 2003; 304: 43–55. doi:10.1016/S0378-1119(02)01175-7.
18. Kaletsky R, Lakhina V, Arey R, Williams A, Landis J, Ashraf J, *et al*. The C. elegans adult neuronal IIS/FOXO transcriptome reveals adult phenotype and regulators. *Nature*. 2016; 529 (7584): 92–96. doi:10.1038/nature16483.
19. Lee H, Lee SJV. Recent progress in the regulation of aging by insulin/IGF-1 signaling in Caenorhabditis elegans. *Mol Cells*. 2022; 45 (11): 763–770. doi:10.14348/molcells.2022.0097.
20. Zhang T, Hwang HY, Hao H, Talbot C Jr, Wang J. Caenorhabditis elegans RNA-processing protein TDP-1 regulates protein homeostasis and lifespan. *J Biol Chem*. 2012; 287 (11): 8371–8382. doi:10.1074/jbc.M111.311977.
21. Taylor M, Marx O, Norris A. TDP-1 and FUST-1 co-inhibit exon inclusion and control fertility together with transcriptional regulation. *Nucleic Acids Res*. 2023; 51 (18): 9610–9628. doi:10.1093/nar/gkad665.
22. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, *et al*. The genetic basis of natural variation in Caenorhabditis elegans telomere length. *Genetics*. 2016; 204 (1): 371–383. doi:10.1534/genetics.116.191148.
23. Zhang G, Roberto NM, Lee D, Hahnel SR, Andersen EC. The impact of species-wide gene expression variation on Caenorhabditis elegans complex traits. *Nat Commun*. 2022; 13 (1): 3462. doi:10.1038/s41467-022-31208-4. Gene Expression Omnibus accession GSE186719. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186719 (accessed on 2026-04-03).
24. WormBase. Release WS276. Available from: https://wormbase.org (accessed on 2026-04-03).
25. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al*. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016; 48 (5): 481–487. doi:10.1038/ng.3538.
26. Kawaguchi T, Rollins MG, Moinpour M, Morera AA, Ebmeier CC, Old WM, *et al*. Changes to the TDP-43 and FUS interactomes induced by DNA damage. *J Proteome Res*. 2020; 19 (1): 360–370. doi:10.1021/acs.jproteome.9b00575.
27. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014; 346 (6213): 1258096. doi:10.1126/science.1258096.