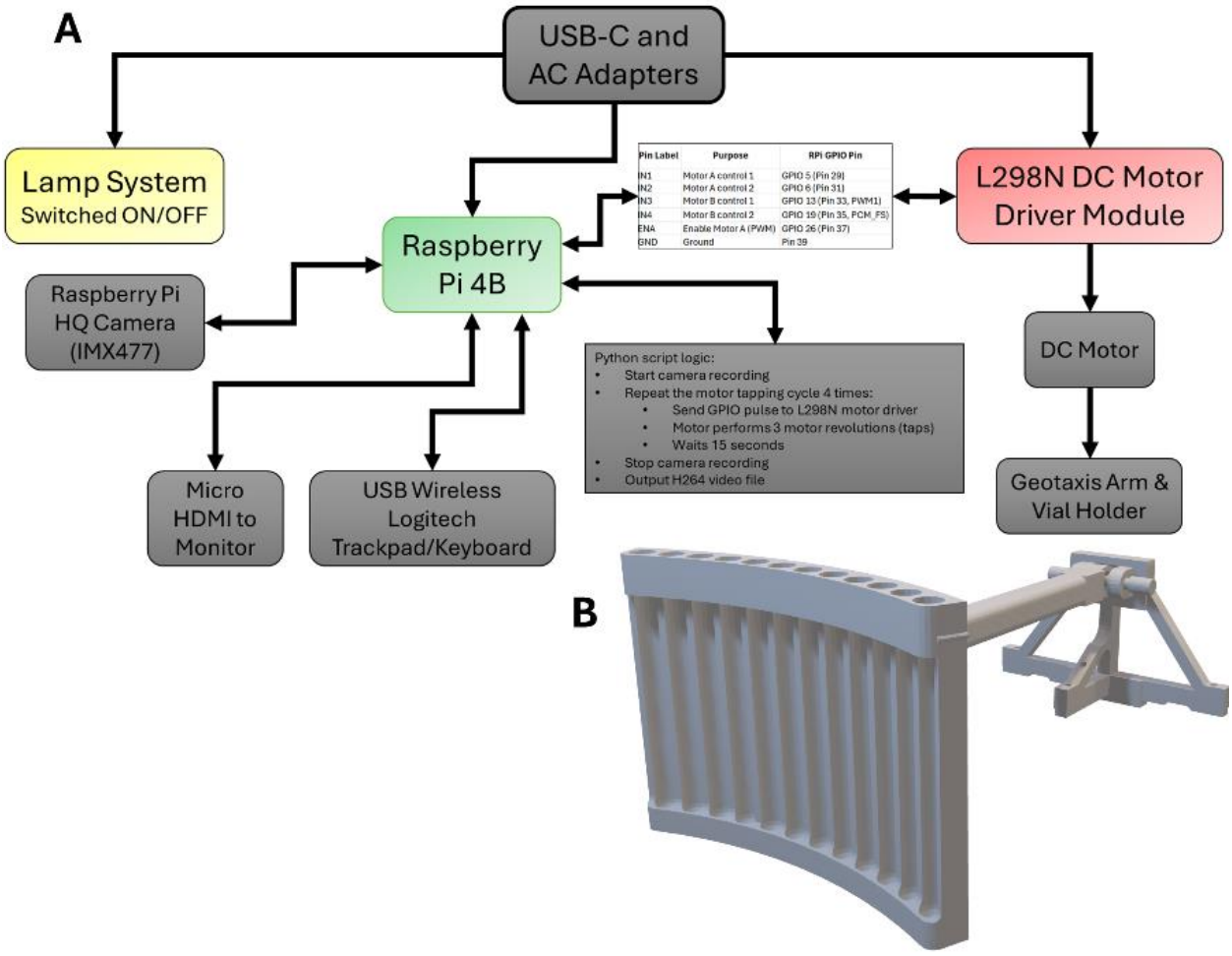
Design and Implementation of an Automated Drosophila Locomotor Assay Using Computer Vision Tracking

Dave Melkani[1], Neelaksh Harnwal[1], Shubhankar Desai[1], Dev Patel[1], Girish Melkani[1, 2*]

[1]Department of Pathology, Division of Molecular and Cellular Pathology, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA. [2] UAB Nathan Shock Center, Birmingham, AL 35294

Supporting information contains.

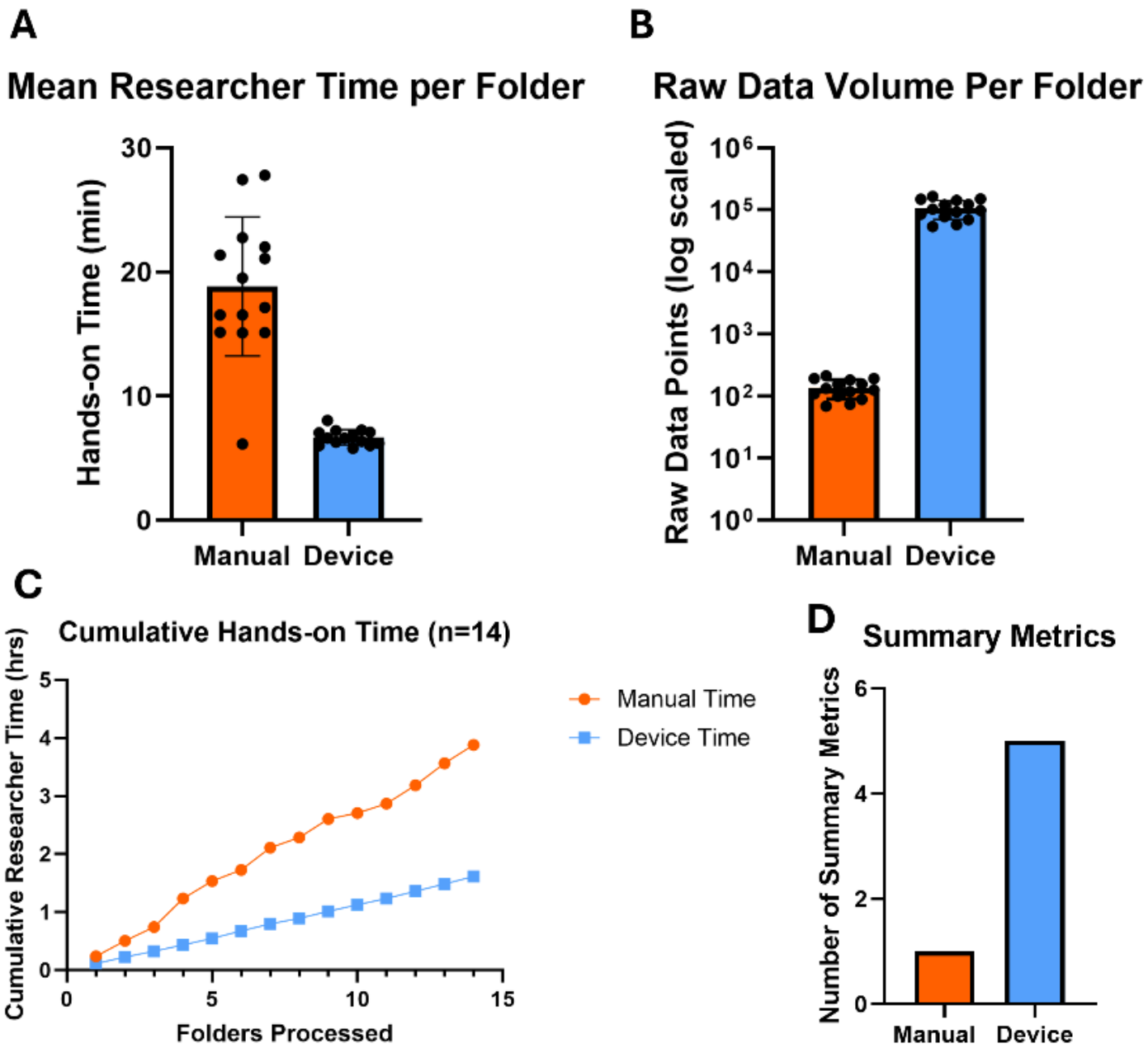


Supplementary Figure 1. Design and Assembly of the Automated Geotaxis Assay Device (a) Control-logic flow chart schematic of power and signal routing: Raspberry Pi GPIO → L298N → DC motor; USB-C/AC power inputs; Python script logic for “tap-pause” cycles and video recording. (b) Screen capture of the CAD rendering of the assembled device from combined STL files. Vial holders and back support illustrate a third-class lever mechanism.

Device and manual assessment

The fully automated geotaxis pipeline substantially outperforms the traditional manual assay in both speed and data richness. Over 14 random video folders, manual scoring required users to pause each of the 4 trim videos for each folder at around 8 seconds, draw a horizontal reference line, and tally the percentage of flies above the reference line for all the vials. The manual scoring averaged 18.8 ± 5.4 min per folder, compared to just 6.8 ± 0.7 min needed to set up and run the automated pipeline showing a 2.8-fold reduction in hands on time as shown in **Supplementary Figure 2a**. These timing values were collected by having researchers use stopwatches to record the duration of each manual

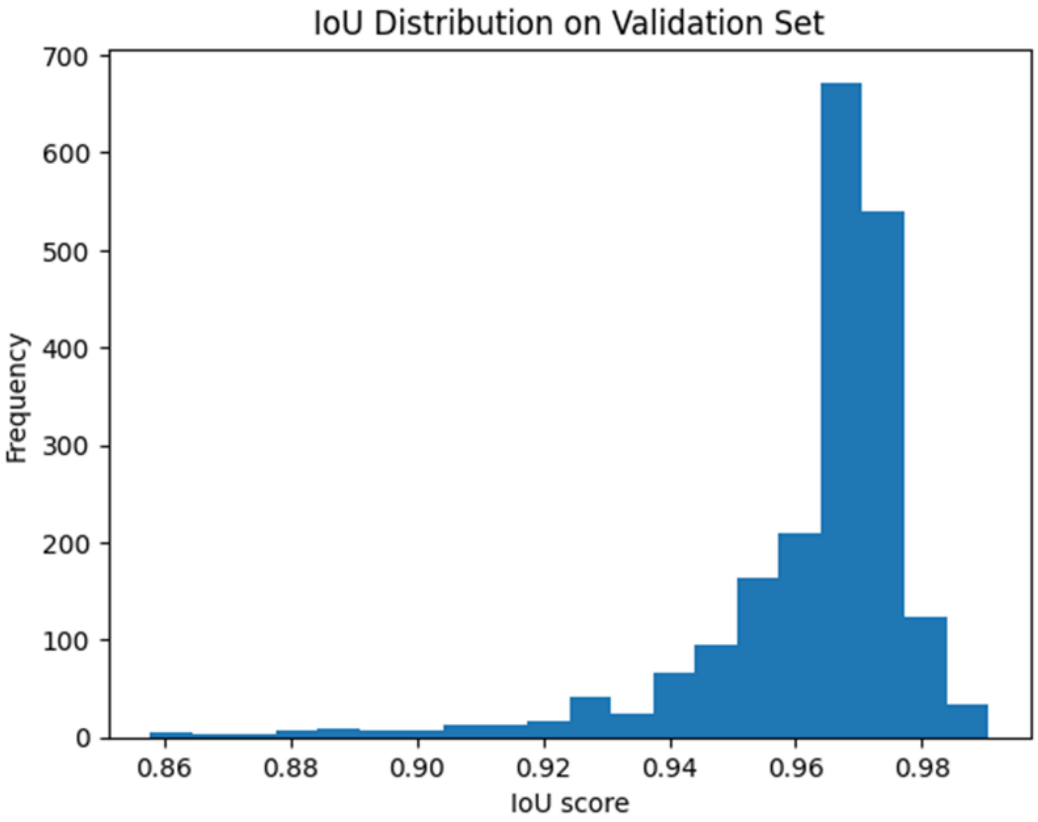
scoring session on each video folders along with launching and initializing the automated pipeline from start to finish side-by-side noting the time after each video folder was processed. Beyond throughput, the automated method captures orders of magnitude more raw observations. **Supplementary Figure 2b** shows how manual scoring yields only $1.36 \times 10^2 \pm 0.44 \times 10^2$ data points per trim video (measures all the flies in a single frame for each vial) averaged over the video folders trims whereas the automated device extracts $1.06 \times 10^5 \pm 0.35 \times 10^5$ points per trim video (measures all the flies in 780 frames for each vial) averaged over the video folders trims. This means the automated method calculates approximately 779 times more data than the manual method. Aggregating the durations across all 14 video folders, **Supplementary Figure 2c** shows the manual method totaling 3.88 hours compared to 1.61 hours for the automated pipeline. These cumulative values were computed by summing each folder's individual timing measurements. This richer dataset also translates into more comprehensive summary outputs as **Supplementary Figure 2d** shows the manual assay produces a single "percent above line" metric per folder, but our pipeline automatically generates five distinct metrics: mean position and velocity time-series plus proportions of low, middle, and high performers time-series via standardized functions in the analysis script. Note that all automated pipeline computations were performed using the University of Alabama at Birmingham's High-Performance Computing (Cheaha) resources, ensuring efficient processing of large-scale video datasets.



Supplementary Figure 2. Device Performance Metrics vs. Manual Assay: (a) Comparison of average processing time per video folder between manual scoring and the automated device, highlighting the reduced hands-on time of the pipeline. (b) Total number of data points extracted per trim video, demonstrating the vastly higher data yield from the automated method. (c) Cumulative duration of the full workflow across multiple folders, showing the overall time savings achieved with automation. (d) Number of output metrics generated per folder, illustrating the richer set of behavioral measures provided by the automated pipeline.

Deep learning performance

We evaluated detection performance both by monitoring training convergence and by measuring IoU overlap on held-out frames. The single-epoch training curve demonstrates swift learning, with loss dropping below 0.02 within 1,140 steps on a ResNet-50 + FPN backbone under minimal augmentation. On inference, the model achieves near-perfect overlap (mean IoU > 0.95) against manual annotations as shown in **Supplementary Figure 3**. By imposing a confidence threshold of 0.85 and applying spatial sorting, we guarantee that only high-certainty vial detections progress to the behavioral-tracking stage. This integration of deep learning not only automates and standardizes vial identification but also underpins precise quantification of climbing metrics (position, speed, and zone occupancy) across genetic and environmental conditions, markedly improving throughput, repeatability, and analytical rigor.



Supplementary Figure 3. IoU Distribution for Deep Learning Performance: Histogram showing the distribution of Intersection-over-Union (IoU) scores between predicted and ground truth vial bounding boxes across validation frames. Scores range from 0 to 1, with higher values indicating better spatial alignment and detection accuracy.

Additional results & discussions

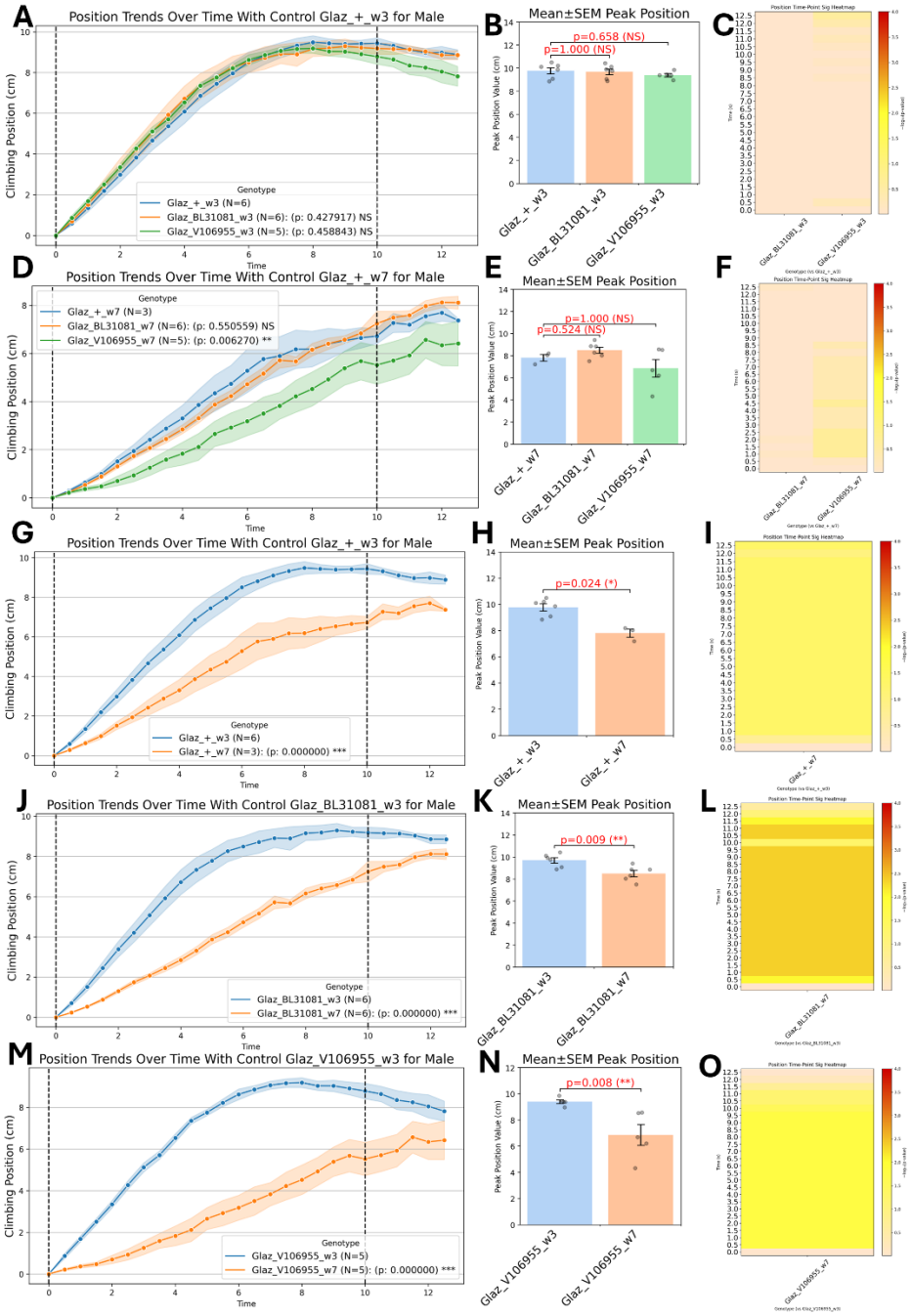
Glaz PolG Lines

Here we analyzed a large dataset from the Glaz-driven PolG knockdown study that includes comparisons for two different ages: 3 weeks and 7 weeks along with two mutant genotypes: $\mathrm{Glaz}>\mathrm{UAS\text{-}PolG}^{RNAi\ (BL31081)}$ (Glaz BL31081) and $\mathrm{Glaz}>\mathrm{UAS\text{-}PolG}^{RNAi\ (\mathrm{V}106955)}$ (Glaz V106955) being compared to the wild-type control $\mathrm{Glaz}/+$ (Glaz +). The data is separated by gender (male and female) and includes multiple quantifying metrics: position, velocity, low performer (LP), middle performer (MP), and high performer (HP). **Supplementary Figure 4** summaries the results of male Glaz lines for just the position metrics. Below is a comprehensive overview of all the Glaz line comparisons for both sexes and all the quantifying metrics with clinical and biological insights.

For males, across all five-quantifying metrics, neither Glaz BL31081 nor Glaz V106955 males differed significantly from Glaz + controls at 3 weeks (all $p > 0.31$, NS). Likewise, peak-level comparisons of maximum metrics showed no statistical differences (all $p \geq 0.165$, NS). At this relatively early adult age, neither disruption allele (BL31081 or V106955) produces a detectable impairment in overall geotaxis performance compared to wild-type Glaz + males. At 7 weeks, for Glaz + versus Glaz BL31081 males, again no significant genotype effects were detected for any of the metrics (all LME $p > 0.33$, NS; peak-level all $p > 0.52$, NS). However, at 7 weeks, for Glaz + vs Glaz V106955, the male mutants climbed significantly lower than Glaz + controls ($p = 0.0063$, **). A higher proportion of V106955 males fell into the low-performer category ($p = 0.0013$, **). The middle-performer rate was modestly but significantly reduced in V106955 ($p = 0.0294$, *). Velocity and high performers showed no significant genotype differences ($p = 0.3282$ and $p = 0.1152$, NS). By 7 weeks, the V106955 allele manifests a clear geotaxis deficit in males as it reduced overall climbing height and a shift toward poorer-performing individuals. The BL31081 allele, however, remains indistinguishable from wild type, indicating that V106955 may be a stronger loss-of-function or dominant-negative allele affecting sustained climbing ability. The selective increase in low performers without a change in maximum velocity suggests that V106955 males often fail to initiate or maintain climbing rather than being inherently slower for those that do climb. For analyzing the aging effects from 3 week to 7 week flies, all three male genotypes exhibited significant age-related declines in geotaxis: Glaz + showed a modest but clear drop in overall climbing trajectory and peak height (all $p < 0.0001$ to $p = 0.024$), with more low performers and fewer high performers but stable velocity and mid-range performance; Glaz _BL31081 aged precipitously across every metric (all $p < 0.0005$), including peak height ($p = 0.009$) and shifts in performer categories; and Glaz _V106955 likewise deteriorated sharply (all LME $p < 0.0001$) with significant losses in peak height ($p = 0.008$) and high-performer rates, though peak velocity and mid-performer changes trended non-significant. Thus, while aging impairs climbing in all males, BL31081 mutants show the most accelerated decline and V106955 combines a baseline impairment with further age-driven losses.

For females, geotaxis performance at 3 weeks showed minimal genotype effects: neither Glaz BL31081 nor Glaz V106955 mutants differed significantly from Glaz + controls across all LME metrics (all $p > 0.236$, NS). Peak-level data were similarly unremarkable, except for a single significant difference

in middle-performer percentage favoring Glaz + over Glaz BL31081 ($p = 0.005$, **), though this isolated effect lacked support from the trajectory-level data. These findings suggest that early adult females maintain normal climbing behavior regardless of genotype, indicating functional compensation or minimal early disruption from either allele. By 7 weeks, however, a more pronounced divergence emerges where both Glaz BL31081 and Glaz V106955 females outperformed Glaz + controls in position metrics, with significantly higher climbing heights from both mutants (Glaz BL31081 $p = 0.0058$, **; Glaz V106955 $p = 0.0022$, **). Glaz BL31081 females exhibited fewer low performers ($p = 0.0421$, *) and more middle performers ($p = 0.0338$, *), while Glaz V106955 females showed increased middle-performer rates ($p = 0.0499$, *). Although velocity and high-performer categories did not differ significantly in the main LME model, peak-level analyses confirmed enhanced maximum heights (Glaz BL31081 $p = 0.022$, *; Glaz V106955 $p = 0.005$, **), reduced peak low performers (Glaz V106955 $p = 0.011$, *), and increased middle and high performers (both $p = 0.011$, * and $p = 0.004$, **, respectively). These findings contrast sharply with the male phenotype, indicating a potential sex-specific benefit of Glaz mutations in aging females. When comparing age trajectories, Glaz + females demonstrated the expected aging-related decline in position, velocity, and performer distribution (all $p < 0.014$), with peak analyses reinforcing substantial drops in height and velocity alongside a shift toward lower performer categories. Both mutant lines are similarly aged with significant reductions across LME and peak metrics (most $p < 0.0001$) yet still retain higher performance levels than Glaz + at 7 weeks. Notably, Glaz V106955 females maintained stable velocity trajectories ($p = 0.1426$, NS) despite declines in climbing height, suggesting preserved instantaneous capacity amid reduced sustained effort. All the p values referenced can be found for males (**Supplementary Table 1a**) and females (**Supplementary Table 1b**). Together, these data reveal that Glaz allele effects in females are characterized by enhanced late-life climbing and altered age-related decline patterns, implicating potential protective or compensatory mechanisms that warrant further investigation into the sex-specific regulation of geotaxis performance.



Supplementary Figure 4. Glaz-driven PolG RNAi lines Geotaxis Behavior and Statistical Analysis Visualizations (a) A comprehensive summary of genotype-specific differences in male geotaxis behavior across four pairwise comparisons: genotype effects: Glaz + wk3 vs Glaz BL31081 wk3 & Glaz + wk3 vs Glaz V106955 wk3 (A-C), Glaz + wk7 vs Glaz BL31081 wk7 & Glaz + wk7 vs Glaz V106955 wk7 (D-F), aging effects: Glaz + wk3 vs Glaz + wk7(G-I), Glaz BL31081 wk3 vs Glaz BL31081 wk7(J-L), Glaz V106955 wk3 vs Glaz V106955 wk7(M-O). Each comparison includes four subpanels: (1) average climbing trajectories over time (±SEM) annotated with significance from a LME model, (2) a bar plot of individual peak climbing metrics with Mann–Whitney U test annotations; and (3) a heatmap visualizing time-resolved $-\log_{10}(p)$ values from Mann–Whitney U tests, indicating the temporal windows during which each genotype diverges significantly from its control.

A

Metrics	Glaz_+_w3 vs Glaz_BL31081_w3	Glaz_+_w3 vs Glaz_V106955_w3	Glaz_+_w7 vs Glaz_BL31081_w7	Glaz_+_w7 vs Glaz_V106955_w7	Glaz_+_w3 vs Glaz_+_w7	Glaz_BL31081_w3 vs Glaz_BL31081_w7	Glaz_V106955_w3 vs Glaz_V106955_w7
Gender	male	male	male	male	male	male	male
position_LME	0.428 (NS)	0.459 (NS)	0.551 (NS)	0.006 (**)	0.000 (***)	0.000 (***)	0.000 (***)
position_Peak	1.000 (NS)	0.658 (NS)	0.524 (NS)	1.000 (NS)	0.024 (*)	0.009 (**)	0.008 (**)
velocity_LME	0.316 (NS)	0.066 (NS)	0.510 (NS)	0.328 (NS)	0.229 (NS)	0.000 (***)	0.000 (***)
velocity_Peak	0.788 (NS)	0.494 (NS)	1.000 (NS)	1.000 (NS)	0.024 (*)	0.093 (NS)	0.056 (NS)
low performer_LME	0.337 (NS)	0.344 (NS)	0.333 (NS)	0.001 (**)	0.000 (***)	0.000 (***)	0.000 (***)
low performer_Peak	0.892 (NS)	0.852 (NS)	0.524 (NS)	0.786 (NS)	0.026 (*)	0.017 (*)	0.011 (*)
middle performer_LME	0.613 (NS)	0.551 (NS)	0.475 (NS)	0.029 (*)	0.400 (NS)	0.000 (***)	0.000 (***)
middle performer_Peak	0.481 (NS)	0.165 (NS)	1.000 (NS)	0.786 (NS)	0.024 (*)	0.004 (**)	0.151 (NS)
high performer_LME	0.453 (NS)	0.487 (NS)	0.569 (NS)	0.115 (NS)	0.000 (***)	0.000 (***)	0.000 (***)
high performer_Peak	0.970 (NS)	0.355 (NS)	0.762 (NS)	1.000 (NS)	0.024 (*)	0.002 (**)	0.008 (**)

B

Metrics	Glaz_+_w3 vs Glaz_BL31081_w3	Glaz_+_w3 vs Glaz_V106955_w3	Glaz_+_w7 vs Glaz_BL31081_w7	Glaz_+_w7 vs Glaz_V106955_w7	Glaz_+_w3 vs Glaz_+_w7	Glaz_BL31081_w3 vs Glaz_BL31081_w7	Glaz_V106955_w3 vs Glaz_V106955_w7
Gender	female	female	female	female	female	female	female
position_LME	0.236 (NS)	0.639 (NS)	0.006 (**)	0.002 (**)	0.000 (***)	0.000 (***)	0.000 (***)
position_Peak	0.732 (NS)	0.590 (NS)	0.022 (*)	0.005 (**)	0.001 (***)	0.001 (**)	0.004 (**)
velocity_LME	0.314 (NS)	0.621 (NS)	0.218 (NS)	0.513 (NS)	0.014 (*)	0.029 (*)	0.143 (NS)
velocity_Peak	1.000 (NS)	0.732 (NS)	0.146 (NS)	0.020 (*)	0.001 (***)	0.534 (NS)	0.009 (**)
low performer_LME	0.267 (NS)	0.427 (NS)	0.042 (*)	0.054 (NS)	0.000 (***)	0.000 (***)	0.000 (***)
low performer_Peak	1.000 (NS)	0.505 (NS)	0.121 (NS)	0.011 (*)	0.002 (**)	0.001 (**)	0.004 (**)
middle performer_LME	0.645 (NS)	0.579 (NS)	0.034 (*)	0.050 (*)	0.000 (***)	0.000 (***)	0.000 (***)
middle performer_Peak	0.005 (**)	0.203 (NS)	0.121 (NS)	0.011 (*)	0.002 (**)	0.002 (**)	0.004 (**)
high performer_LME	0.715 (NS)	0.819 (NS)	0.053 (NS)	0.232 (NS)	0.000 (***)	0.000 (***)	0.000 (***)
high performer_Peak	0.469 (NS)	0.590 (NS)	0.023 (*)	0.004 (**)	0.001 (**)	0.008 (**)	0.004 (**)

Supplementary Table 1. P-values for trajectory and peak metrics comparing Glaz-driven PolG RNAi lines for (a) males and (b) females. Footnote: NS, not significant ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$).

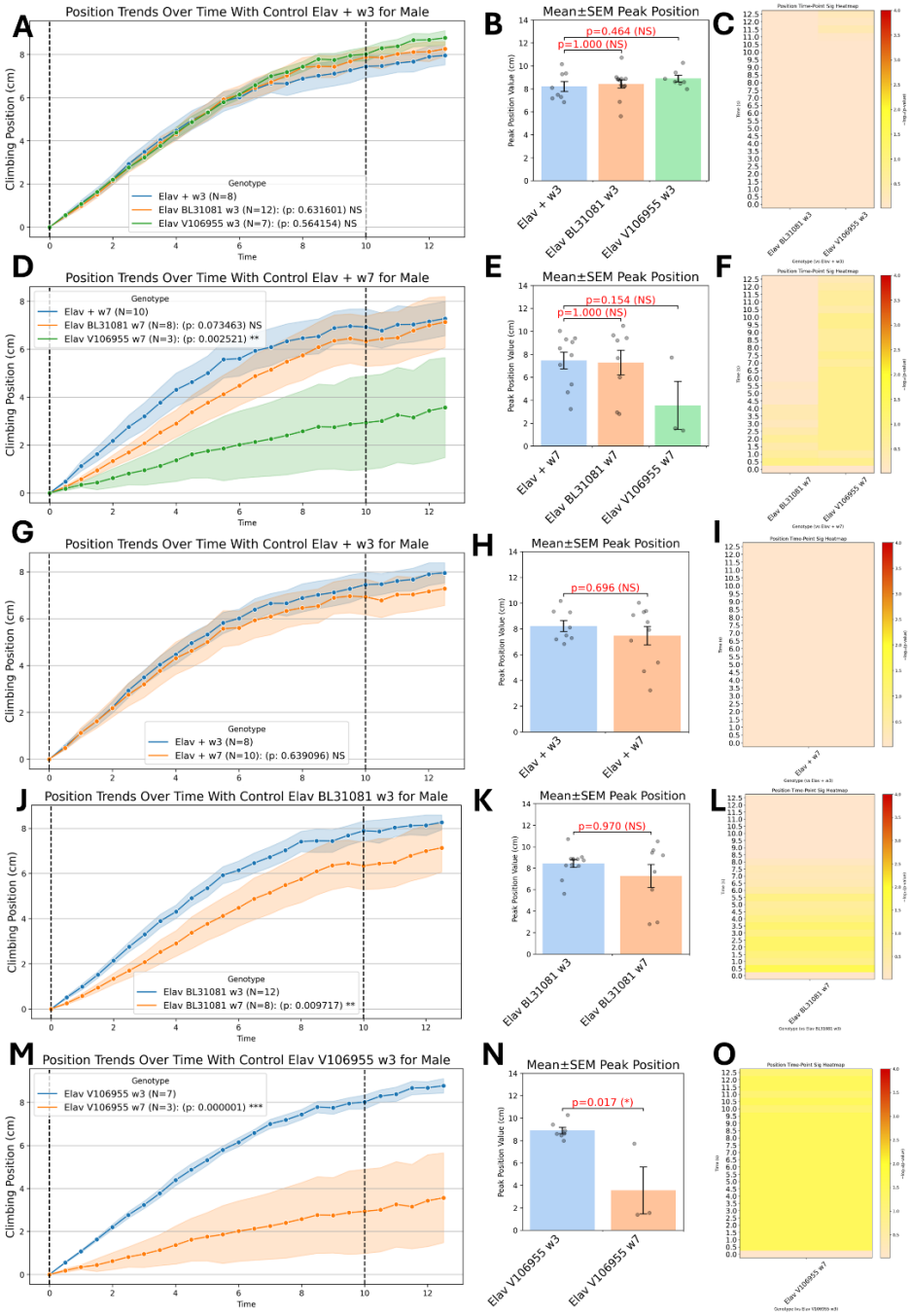
Elav PolG Lines

Here we analyzed geotaxis performance for Elav-driven PolG knockdown where we compared two RNAi alleles: $\mathrm{Elav}>\mathrm{UAS\text{-}PolG}^{RNAi\,(BL31081)}$ (Elav BL31081) and $\mathrm{Elav}>\mathrm{UAS\text{-}PolG}^{RNAi\,(\mathrm{V}106955)}$ (Elav V106955) to the wild-type control $\mathrm{Elav}/+$ (Elav +) at 3 and 7 weeks, and tracked aging from 3 to 7 weeks within each genotype. **Supplementary Figure 5** summaries the results of male Elav lines for just the position metrics. Below is a comprehensive overview of all the Elav line comparisons for both sexes and all the quantifying metrics with clinical and biological insights.

For males, at 3 weeks neither Elav BL31081 nor Elav V106955 differed from Elav + controls in any of the five quantifying metrics (all LME $p > 0.16$, NS; all peak $p \geq 0.363$, NS), indicating no early adult climbing deficits. By 7 weeks, Elav BL31081 males showed only modest impairments with no change in climbing trajectory (LME $p = 0.0735$, NS) but significantly reduced velocity (LME $p = 0.0480$, *), a higher proportion of low performers (LME $p = 0.0166$, *), and stable mid-performer and high-performer rates ($p > 0.10$, NS). Elav V106955 males exhibited a clearer deficit at 7 weeks, with lower overall climbing trajectories (LME $p = 0.0025$, **) and more low performers (LME $p = 0.0038$, **) but unchanged velocity, mid-performers, and high-performer distributions ($p > 0.08$, NS); however, neither allele produced significant differences in any peak-level metrics at this age (all peak $p > 0.15$, NS). In aging comparisons from 3 to 7 weeks, Elav + controls remained stable across all trajectory and peak metrics (all $p > 0.24$, NS), whereas Elav BL31081 males showed significant age-related declines in position and velocity (LME $p < 0.01$, **), increased low performers ($p = 0.0004$, ***), and modest shifts in peak measures (all peak $p > 0.30$, NS). Elav V106955 males experienced profound deteriorations across every trajectory metric (all LME $p < 0.03$, *), with significant losses in peak height ($p = 0.017$, *), velocity ($p = 0.017$, *), low-performer increases ($p = 0.017$, *), middle-performer decreases ($p = 0.033$, *), and high-performer decreases ($p = 0.017$, *). While aging effects is spared for the wild-type Elav + males, both PolG RNAi alleles amplify age-driven geotaxis decline, with Elav V106955 combining baseline deficits and accelerated deterioration.

For females, at 3 weeks both PolG knockdowns outperformed Elav + in climbing trajectories: Elav BL31081 females climbed higher (LME $p = 0.0006$, ***) with fewer low performers ($p = 0.0022$, **) and more middle performers ($p = 0.0033$, **). Elav V106955 females likewise showed elevated trajectories (LME $p = 0.0010$, **), dramatic reductions in low performers ($p = 0.0001$, ***) and increases in middle ($p < 0.0001$, ***) and high performers ($p = 0.0171$, *). Peak-level analyses partially mirrored these trends for V106955 (peak position $p = 0.009$, **; low $p = 0.005$, **; middle $p = 0.001$, **; high $p = 0.016$, *), whereas BL31081 peak differences were limited to mid-performers (peak $p = 0.010$, **). By 7 weeks, Elav BL31081 females showed no significant genotype effects on any trajectory or peak metric (all $p > 0.37$, NS). Elav V106955 females at 7 weeks exhibited only a modest reduction in middle-performer rate (LME $p = 0.0313$, *) with all other metrics unchanged ($p > 0.05$, NS) and no peak-level differences (all $p > 0.20$, NS). In aging effects from 3 to 7 weeks, Elav + females remained stable across trajectory and peak measures (all $p > 0.43$, NS), whereas Elav BL31081 females experienced significant declines in trajectory height, velocity, and performer distributions (LME $p < 0.01$, **), with

peak analyses confirming decreased height ($p = 0.030$, *) and velocity ($p = 0.001$, **) with reduced high-performer rates ($p = 0.025$, *). Elav V106955 females likewise showed pronounced age-related trajectory losses in height, low- and mid-performer shifts (all LME $p < 0.0001$, ***) but stable velocity and high-performer trajectories ($p > 0.11$, NS); peak-level comparisons corroborated significant decreases across all five metrics (all $p = 0.016$, *). All the p values referenced can be found for males (**Supplementary Table 2a**) and females (**Supplementary Table 2b**). Together, these data reveal a sex-specific pattern whereby early adult female PolG knockdowns benefit from enhanced climbing that erodes with age with Elav BL31081 fully normalizing by 7 weeks and Elav V106955 retaining only a middle-performer deficit, while male mutants display delayed deficits that worsen markedly over time.



Supplementary Figure 5. Elaz-driven PolG RNAi lines Geotaxis Behavior and Statistical Analysis Visualizations (a) A comprehensive summary of genotype-specific differences in male geotaxis behavior across four pairwise comparisons: genotype effects: Elav + wk3 vs Elav BL31081 wk3 & Elav + wk3 vs Elav V106955 wk3 (A-C), Elav + wk7 vs Elav BL31081 wk7 & Elav + wk7 vs Elav V106955 wk7 (D-F), aging effects: Elav + wk3 vs Elav + wk7(G-I), Elav BL31081 wk3 vs Elav BL31081 wk7(J-L), Elav V106955 wk3 vs Elav V106955 wk7(M-O). Each comparison includes four subpanels: (1) average climbing trajectories over time (±SEM) annotated with significance from a LME model, (2) a bar plot of individual peak climbing metrics with Mann–Whitney U test annotations; and (3) a heatmap visualizing time-resolved $-\log_{10}(p)$ values from Mann–Whitney U tests, indicating the temporal windows during which each genotype diverges significantly from its control.

A

Metrics	Elav + w3 vs Elav BL31081 w3	Elav + w3 vs Elav V106955 w3	Elav + w7 vs Elav BL31081 w7	Elav + w7 vs Elav V106955 w7	Elav + w3 vs Elav + w7	Elav BL31081 w3 vs Elav BL31081 w7	Elav V106955 w3 vs Elav V106955 w7
Gender	male	male	male	male	male	male	male
position_LME	0.632 (NS)	0.564 (NS)	0.073 (NS)	0.003 (**)	0.639 (NS)	0.010 (**)	0.000 (***)
position_Peak	1.000 (NS)	0.464 (NS)	1.000 (NS)	0.154 (NS)	0.696 (NS)	0.970 (NS)	0.017 (*)
velocity_LME	0.423 (NS)	0.612 (NS)	0.048 (*)	0.306 (NS)	0.632 (NS)	0.002 (**)	0.030 (*)
velocity_Peak	1.000 (NS)	0.379 (NS)	0.474 (NS)	0.098 (NS)	0.762 (NS)	0.069 (NS)	0.017 (*)
low performer_LME	0.172 (NS)	0.162 (NS)	0.017 (*)	0.004 (**)	0.301 (NS)	0.000 (***)	0.000 (***)
low performer_Peak	1.000 (NS)	0.363 (NS)	1.000 (NS)	0.151 (NS)	0.755 (NS)	0.521 (NS)	0.017 (*)
middle performer_LME	0.292 (NS)	0.084 (NS)	0.141 (NS)	0.098 (NS)	0.249 (NS)	0.062 (NS)	0.000 (***)
middle performer_Peak	0.541 (NS)	0.028 (*)	1.000 (NS)	0.409 (NS)	0.450 (NS)	0.305 (NS)	0.033 (*)
high performer_LME	0.625 (NS)	0.405 (NS)	0.107 (NS)	0.081 (NS)	0.728 (NS)	0.084 (NS)	0.000 (***)
high performer_Peak	1.000 (NS)	0.794 (NS)	1.000 (NS)	0.154 (NS)	0.573 (NS)	0.792 (NS)	0.017 (*)

B

Metrics	Elav + w3 vs Elav BL31081 w3	Elav + w3 vs Elav V106955 w3	Elav + w7 vs Elav BL31081 w7	Elav + w7 vs Elav V106955 w7	Elav + w3 vs Elav + w7	Elav BL31081 w3 vs Elav BL31081 w7	Elav V106955 w3 vs Elav V106955 w7
Gender	female	female	female	female	female	female	female
position_LME	0.001 (***)	0.001 (**)	0.582 (NS)	0.077 (NS)	0.565 (NS)	0.000 (***)	0.000 (***)
position_Peak	0.411 (NS)	0.009 (**)	1.000 (NS)	0.203 (NS)	0.227 (NS)	0.030 (*)	0.016 (*)
velocity_LME	0.212 (NS)	0.464 (NS)	0.437 (NS)	0.526 (NS)	0.432 (NS)	0.009 (**)	0.113 (NS)
velocity_Peak	1.000 (NS)	1.000 (NS)	1.000 (NS)	0.261 (NS)	0.129 (NS)	0.001 (**)	0.016 (*)
low performer_LME	0.002 (**)	0.000 (***)	0.374 (NS)	0.057 (NS)	0.648 (NS)	0.000 (***)	0.000 (***)
low performer_Peak	0.473 (NS)	0.005 (**)	1.000 (NS)	0.178 (NS)	0.114 (NS)	0.055 (NS)	0.016 (*)
middle performer_LME	0.003 (**)	0.000 (***)	0.383 (NS)	0.031 (*)	0.646 (NS)	0.000 (***)	0.000 (***)
middle performer_Peak	0.010 (**)	0.001 (**)	1.000 (NS)	0.178 (NS)	0.598 (NS)	0.055 (NS)	0.016 (*)
high performer_LME	0.084 (NS)	0.017 (*)	0.607 (NS)	0.299 (NS)	0.648 (NS)	0.002 (**)	0.195 (NS)
high performer_Peak	0.695 (NS)	0.016 (*)	1.000 (NS)	0.200 (NS)	0.114 (NS)	0.025 (*)	0.016 (*)

Supplementary Table 2. P-values for trajectory and peak metrics comparing Elav-driven PolG RNAi lines for (a) males and (b) females. Footnote: NS, not significant ($p > 0.05$); * ($p \leq 0.05$); ** ($p \leq 0.01$); *** ($p \leq 0.001$).