DANCE: An open-source analysis pipeline and low-cost hardware to quantify aggression and courtship in *Drosophila*

R. Sai Prathap Yadav[#1], Paulami Dey[#1], Faizah Ansari[1], Tanvi Kottat[1], P. Pallavi Prabhu[1], Manohar Vasam[1], Shrinivas Ayyangar[1], Swathi Bhaskar S.[1], Krishnananda Prabhu[2], Monalisa Ghosh[1,3], Pavan Agrawal[1]*

1. Centre for Molecular Neurosciences, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi 576104, Karnataka, India.

2. Department of Biochemistry, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Udupi 576104, Karnataka, India.

3. Present address: Gottfried Schatz Research Center, Medical University of Graz, Neue Stiftingtalstr. 6, 8010 Graz, Austria.

#These authors contributed equally to this study.

*Correspondence to: pavan.agrawal@manipal.edu

ORCID: 0000-0002-4494-7859

Abstract

Quantifying animal behaviors is pivotal for identifying the underlying neuronal and genetic mechanisms. Computational approaches have enabled automated analysis of complex behaviors such as aggression and courtship in *Drosophila*. However, existing approaches rely on rigid, rule-based algorithms and expensive hardware, limiting sensitivity to behavioral variations and accessibility. Here, we describe the DANCE (*Drosophila* Aggression and Courtship Evaluator), a low-cost, open-source platform combining machine learning-based classifiers and inexpensive hardware to quantify aggression and courtship. DANCE consists of six novel behavioral classifiers trained using a supervised machine learning algorithm. DANCE classifiers outperform existing rule-based algorithms by capturing dynamic behavioral variations. DANCE hardware is constructed using repurposed medicine blister packs and acrylic sheets, with recordings performed using smartphones, making it affordable and accessible. Benchmarking demonstrated that DANCE hardware performs comparably to sophisticated, high-cost setups. We validated DANCE in diverse contexts, including social isolation versus enrichment that modulate aggression and courtship, RNAi-mediated down-regulation of neuropeptide Dsk, and optogenetic silencing of dopaminergic neurons which promoted aggression. DANCE provides a cost-effective and portable solution for studying *Drosophila* behaviors in resource-limited settings or closer to natural habitats. Its accessibility and robust performance democratizes behavioral neuroscience, enabling rapid screening of genes and neuronal circuits underlying complex social behaviors.

Keywords: *Drosophila*, Aggression, Courtship, Machine learning, Computer vision

Introduction

Detailed and accurate annotation and analysis of complex behaviors is necessary for understanding the underlying neural and molecular mechanisms. Fruit fly *Drosophila melanogaster* is one of the most accessible and well-studied model organisms for identifying neuronal and molecular underpinnings of behavior. Multiple large-scale screens have been conducted using *Drosophila*, to study complex social behaviors such as aggression and courtship (Asahina, 2017; Greenspan and Ferveur, 2000; Hall, 2002; Kravitz and Fernandez, 2015) to identify underlying neural circuitry (Agrawal et al., 2020; Asahina et al., 2014; Davis et al., 2018; Hoopfer et al., 2015; Yadav et al., 2024) and genes involved (Agrawal et al., 2020; Benzer, 1967; Gill, 1963; Hall, 1978; Ishii et al., 2022; Wang et al., 2008). These behaviors exhibit stereotyped patterns. For instance, aggression involves chasing, fencing (Jacobs, 1960), wing threats, boxing (Dow and Schilcher, 1975), lunging, and tussling (Hoffmann, 1987a, 1987b). Similarly, courtship consists of multiple stereotyped behaviors exhibited by the male fly, such as orienting, circling, and following the female (Cook and Cook, 1975; Markow, 1987; O'Dell, 2003). To stimulate the female to be more receptive, the male produces a species-specific song by vibrating and extending its wing (Bennet-Clark and Ewing, 1969; Swain and von Philipsborn, 2021). The male then attempts copulation by curling its abdomen and finally mounts the female for copulation (Bastock and Manning, 1955; Spieth, 1974).

Manual analysis by trained observers is considered the gold standard in behavioral analysis, but it is a time-consuming process unsuitable for large-scale screens (Gomez-Marin et al., 2014; Robie et al., 2017a). 'Computational ethology' (Anderson and Perona, 2014; Datta et al., 2019) helps address this challenge by automating behavioral annotation by leveraging advances in computer vision and machine learning (Robie et al., 2017b). This enables high-throughput behavioral screening to identify responsible genes and circuits. A typical computational ethology workflow involves recording animal behaviors and tracking

their positions along with body movements. This is followed by the analysis and classification of the observed behaviors from hundreds to thousands of video frames capturing behavioral instances. Several software such as Ctrax, Caltech FlyTracker, Deep Lab Cut etc. (Branson et al., 2009; Eyjolfsdottir et al., 2014; Mathis et al., 2018) are used widely for tracking behaviors in *Drosophila*. Each comes with their strengths and weaknesses.

For instance, Ctrax (Branson et al., 2009) helps track fly body position and movements but leads to frequent identity switches among tracked flies, which is not the case while using FlyTracker (Eyjolfsdottir et al., 2014). Effectiveness of various machine learning pipelines is eventually measured by comparing their output to human annotation, called— 'ground-truthing'. Rule-based algorithms such as CADABRA (Dankert et al., 2009) is used to quantify aggression, but it can lead to mis-scoring and identity switches, as revealed by ground-truthing (Simon and Heberlein, 2020), which needs to be corrected in a semi-automated manner (Kim et al., 2018). MateBook (Ribeiro et al., 2018) is another rule-based algorithm to quantify courtship; however, similar to CADABRA, it tends to miss true-positive events, leading to significant mis-scoring of behaviors.

Janelia Automatic Animal Behavior Annotator (JAABA) (Kabra et al., 2013) addresses the challenges of rigid rule-based approaches by employing a supervised learning approach. In the JAABA pipeline, user-labeled data is utilized for training to encompass the dynamic variations in behaviors, allowing it to predict behaviors based on learning from input data.

Recently, a few JAABA-based behavioral classifiers have been developed for measuring aggression (Chowdhury et al., 2021; Leng et al., 2020) and courtship (Gil-Martí et al., 2023; Pantalia et al., 2023). However, these studies either do not make these classifiers publicly available (Gil-Martí et al., 2023; Leng et al., 2020; Pantalia et al., 2023) or require

specialized hardware such as 3D printed parts (Chowdhury et al., 2021; Gil-Martí et al., 2023) and expensive machine vision cameras (Chowdhury et al., 2021; Leng et al., 2020), making them inaccessible for resource-limited settings and limiting their wider adoption.

Here we describe DANCE (*Drosophila* Aggression and Courtship Evaluator), an open-source, user-friendly analysis and hardware pipeline to simplify and automate the process of robustly quantifying aggression and courtship behaviors. DANCE has two components: 1. A set of robust, machine vision-based behavioral classifiers developed using JAABA to quantify aggression and courtship; and 2. An inexpensive hardware utilizing off-the-shelf components to construct behavioral arenas and smartphone cameras to record behaviors. DANCE classifiers outperform previous methods (Dankert et al., 2009; Ribeiro et al., 2018), improving accuracy and reliability, and DANCE hardware eliminates the need for complex, specialized behavioral arenas and expensive machine vision cameras to assess fly behaviors. Therefore, the DANCE assay can be useful for wider adoption, including in resource-strapped settings. DANCE provides a platform for rapid behavioral screening using the *Drosophila* model for discovering mechanisms underlying complex social behaviors and neurological disorders.

Results

DANCE assay analysis pipeline

To overcome the challenge of time-consuming manual behavioral annotation or resource-intensive, complex hardware, we developed an automated, high-throughput quantification pipeline—DANCE and trained new behavioral classifiers using an existing machine learning algorithm- JAABA (Kabra et al., 2013), to robustly quantify aggression and courtship in *Drosophila* (Figure 1). We built a novel hardware platform consisting of repurposed transparent medicine tablet foils/blister packs, acrylic sheets, and paper tape to easily

record aggression and courtship behaviors. To record these behaviors we used Android smartphone cameras and an electronic tablet or smartphone served as a backlight illumination source (Figure 1A, Materials and Methods). We quantify these behaviors using our 'DANCE classifiers'. Unlike existing set-ups that cost approximately USD 3500, 'DANCE hardware' is fabricated using off-the-shelf components and costs less than USD 30 cents (Supplementary file 1). To benchmark the performance of the DANCE classifiers, we used pre-existing setups and rule-based methods for quantifying courtship and aggression and compared their performance with DANCE classifiers (Figure 1B-C).

To train DANCE classifiers using JAABA (Kabra et al., 2013), for aggressive lunges, we used a pre-existing setup described in (Dankert et al., 2009) modified from (Dierick, 2007) (Figure 1A; Figure 1— figure supplement 1); and for courtship behaviors, we used a pre-existing set up described in (Koemans et al., 2017) (Figure 1A; Figure 1— figure supplement 2). We tracked position, motion, and interactions of pairs of flies across video frames using the Caltech FlyTracker (Eyjolfsdottir et al., 2014). To avoid data leak, we randomly divided acquired videos into two categories: "training videos" and "test videos" to train and evaluate DANCE classifiers. These test videos were also manually 'ground-truthed' frame by frame, which is considered the gold standard for behavioral annotation (Figure 1B, D-E). We benchmarked the performance of DANCE classifiers using existing rule-based algorithms, CADABRA (Dankert et al., 2009) and MateBook (Ribeiro et al., 2018); and an existing JAABA aggression classifier (Chowdhury et al., 2021). We then compared their output against manual ground-truth. Figure 1 D-E show examples of courtship and aggression video annotation, suggesting that DANCE classifiers outperform existing methods and are comparable with manual ground-truth. Subsequent sections describe the quantitative analysis of individual DANCE classifiers and benchmarking of DANCE hardware.

138
139 **DANCE lunge classifier to quantify aggressive behavior**

140 Aggression is an innate, complex behavior, and *Drosophila* males exhibit several
141 stereotyped behavioral patterns during aggressive encounters, with lunging used widely as
142 a measure of overall aggression in males (Agrawal et al., 2020; Asahina et al., 2014; Chiu
143 et al., 2021; Chowdhury et al., 2021; Davis et al., 2014; Dierick, 2007; Hoffmann, 1987a;
144 Hoopfer et al., 2015; Hoyer et al., 2008; Jung et al., 2020; Nilsen et al., 2004; Watanabe et
145 al., 2017; Yadav et al., 2024). A lunge is defined as a male fly raising its front legs and
146 hitting down on the other fly. We developed a new classifier using JAABA (Kabra et al.,
147 2013) to robustly quantify aggressive lunges in *Drosophila*, hereafter referred to as the
148 'DANCE lunge classifier.' We quantified lunges using our classifier from 20-minute-long
149 videos and compared the output with manual ground-truthing and existing methods—
150 CADABRA (Dankert et al., 2009) and 'Divider assay classifier' (Chowdhury et al., 2021).
151 CADABRA tends to miss several true positive lunges, likely due to its rigid, rule-based
152 framework, which cannot adapt to the dynamic variations in behavior. To address this,
153 'Divider assay classifier' was developed based on JAABA, but the setup requires use of 3-
154 D printed rectangular arena and dividers as well as expensive machine vision cameras
155 (Chowdhury et al., 2021).

156 Figure 2A shows lunge scores from 40 different videos using ground-truth, DANCE lunge
157 classifier, CADABRA, and Divider assay classifier. While the ground-truth and DANCE
158 classifier's output are comparable, CADABRA and the Divider assay classifier underscored
159 lunges across videos. We ground truthed the DANCE lunge classifier against 40 'test
160 videos' (Figure 2A, Materials and Methods). To address potential observer bias during
161 ground-truthing, we compared the annotations between two separate evaluators for the
162 DANCE lunge classifier and found no significant differences (Figure 2— figure supplement

163 1A). Since aggressive lunges have a large dynamic range, we benchmarked our classifier
164 across a range of aggressive behaviors, and subdivided the ground truthed videos into four
165 categories: 1. low aggressive, 0–70 lunges (Figure 2B); 2. moderately aggressive, 71–160
166 lunges (Figure 2C); 3. highly aggressive, 161–300 lunges (Figure 2D); and 4. hyper-
167 aggressive, >300 lunges (Figure 2E). We found that DANCE lunge classifier and manual
168 ground-truth performance was comparable across all four categories (Figure 2 B-E). In
169 contrast, the CADABRA and Divider assay classifier showed significant underscoring from
170 the manual ground-truth (Figure 2B-E), however, a strong correlation was observed
171 between our DANCE lunge classifier and ground-truth (Figure 2F, Supplementary file 2),
172 whereas, CADABRA and the Divider assay classifier showed only a weaker correlation with
173 the ground truth (Figure 2G-H, Supplementary file 2). We reasoned that CADABRA's lower
174 performance is most likely due to the rigid rules used to define a lunge (Dankert et al., 2009)
175 (Materials and Methods); and the Divider assay classifier's trained in a rectangular arena
176 did not perform as well in a circular arena since JAABA classifiers utilize features related to
177 arena geometry.

178 To further evaluate the performance of our DANCE lunge classifier, we used three
179 metrics— precision, recall, and F1 score and compared the output of the classifier and
180 manual ground-truth across 40 test videos. We calculated the counts of true positives, false
181 positives, and false negatives to derive precision, which depicts the ratio of true positives
182 among overall predictions from the classifier; recall, which depicts the ratio of true positives
183 among the ground-truth scores; and F1 score, which is a harmonic mean of precision and
184 recall (Figure 2I) (Materials and Methods). Our DANCE lunge classifier had a precision of
185 76.80%, a recall rate of 71.53% and an overall F1 score of 73.60% (Figure 2I), which is
186 higher than the CADABRA and the Divider assay suggesting that the DANCE lunge
187 classifier outperforms these existing methods. Together, our analysis suggests that the

188 DANCE lunge classifier performs with high precision and quantifies lunge numbers robustly
189 over a broad range of fighting intensities.

190
191 **DANCE classifiers to quantify courtship behaviors in *Drosophila***

192 The first report of Drosophila courtship behavior, identified stereotypic behaviors such as
193 wing 'scissor-like' movements, males 'swaying around the female,' licking, tapping, and
194 mounting (Sturtevant, 1915). By the 2000s, studies uncovered genes and neural circuits
195 involved in courtship (Dickson, 2008; Pavlou and Goodwin, 2013). Automated analysis
196 techniques for courtship exist but face barriers to adoption due to expensive hardware,
197 reliance on 3D printing, or lack of publicly available analysis code or classifiers (Gil-Martí et
198 al., 2023; Koemans et al., 2017; Leng et al., 2020; Reza et al., 2013).

199 MateBook is a recent rule-based algorithm that was developed to automate quantifying
200 various courtship behaviors in *Drosophila* males (Ribeiro et al., 2018). It relies on predefined
201 mathematical rules based on CADABRA (Dankert et al., 2009). To resolve ambiguities in
202 the two overlapping flies during estimating their trajectories from the video recordings, it
203 assigns the identities by relying on the distinct body sizes of the male and female fly, as the
204 females are larger than males. However, decapitated virgin females are of similar body size
205 as males, and are often used in courtship assays to assess the male attractiveness to other
206 factors such as pheromones, independent of the female behavior (Cook and Cook, 1975;
207 Spieth, 1966). In such cases, rigid rule-based approach can introduce both false positives
208 and false negatives since the condition about differential body size is not met.

209 To address the inherent inflexibility of rule-based algorithms, we trained and validated five
210 new courtship classifiers using JAABA (Kabra et al., 2013) to provide a detailed and robust
211 analysis of multiple steps of courtship ritual (Sokolowski, 2001) including— wing extension,

following, circling, attempted copulation, and copulation. We evaluated the performance of our DANCE classifiers by comparing the output to ground-truth and MateBook (Ribeiro et al., 2018). Criteria used to develop various classifiers are described in Materials and Methods. To compare the output, we calculated a behavior index, since courtship behaviors, unlike aggressive lunges, are of varied duration. To address potential observer bias during ground-truthing, we compared the annotations between two separate evaluators for the DANCE courtship classifiers and found no significant differences, suggesting robustness of our classifiers (Figure 2— figure supplement 1B-F).

Wing Extension

During unilateral wing extension, the male vibrates its wing at a specific frequency to produce a species-specific courtship song to attract the female (Shorey, 1962; Spieth, 1952). Figure 3A shows wing extension index from the ground-truth (grey) and the 'DANCE wing extension classifier' (orange), which are comparable across 15 videos from males courting decapitated virgin females, however, MateBook's output (purple) seems to underscore these events in most videos (Figure 3A, Friedman's ANOVA with Dunn's test; ground-truth *vs.* MateBook, $p=0.0020$, $n=15$; ground-truth *vs.* DANCE classifier, $p>0.9999$**)** (Figure 3A). We also evaluated the overall wing extension index from all 15 videos and found ground-truth scores and the DANCE wing extension classifier matched well, but MateBook significantly underscored wing extension (Figure 3B). We found a strong correlation between the DANCE wing extension classifier index to ground-truth (Figure 3C) but a weaker correlation with MateBook (Figure 3D). To further evaluate our classifier's performance, we calculated precision, recall, and F1 scores, as described earlier. The DANCE wing extension classifier had a precision of 92.19%, a recall rate of 98.09%, and an overall F1 score of 95.05% when compared to the ground-truth which is higher than the MateBook (Figure 3E).

238 Quantification of the videos containing mated females identified a robust correlation
239 between our classifier's output and the ground-truth, whereas a general trend of
240 underscoring was observed for MateBook (Figure 3— figure supplement 1). Together, this
241 suggests that the DANCE wing extension classifier is robust for both decapitated and mated
242 females and performs with high precision.
243
244 **Attempted copulation and copulation**
245 Copulation in male flies lasts for about 15-25 minutes, and its duration is primarily
246 determined by the male (MacBean and Parsons, 1967). Interrupted mating experiments
247 show that the sperms are transferred several minutes after copulation begins (Fowler,
248 1973; Tompkins et al., 1980). Thus this differentiates the mounting behavior into two—
249 successful copulation and unsuccessful attempted-copulation. We developed the 'DANCE
250 attempted copulation' with a threshold of 0.33 to <45 seconds and the 'DANCE copulation'
251 classifiers with a threshold of >45 seconds, based on existing definitions (Ribeiro et al.,
252 2018), but with modifications, since MateBook only quantifies 'copulation' and not
253 'attempted-copulation' (Materials and Methods). For ground-truthing, we combined both
254 the dataset of mated and decapitated virgin females, because we observed that MateBook
255 consistently overscored in both datasets. (Figure 4A-B). MateBook's rule-based approach
256 led to significant overscoring with reduced thresholds for attempted copulations, especially
257 with decapitated virgins, due to both increased false positives and identity switches. As
258 expected, the DANCE classifier matched well with the ground-truth scores for each
259 individual video (Figure 4A— Friedman's ANOVA with Dunn's test; $p>0.9999$, $n=32$), as
260 well as the overall attempted copulation index (Figure 4B). We found a strong correlation
261 between ground-truth and the DANCE attempted copulation classifier index (Figure 4C) but
262 a weaker correlation with MateBook (Figure 4D). The DANCE classifier performed robustly
263 with a precision rate of 82.55%, a recall rate of 89.24%, and a F1 score of 85.77% as

compared to MateBook (Figure 4E). Together, this suggests that the DANCE attempted copulation outperforms existing rule based framework and is able to capture dynamic variations in behaviors robustly.

Since the male readily copulates with the virgin females for several minutes, we used videos where males were paired with virgin females to train the 'DANCE copulation classifier'. Our classifier matched well with the ground truth in the number of bouts, with a 100% precision, recall and F1 score (Figure 4— figure supplement 1). The rule-based approach of MateBook also works well with copulation, which MateBook defines as 'occlusion' persisting for a long time (Materials and Methods). However sometimes, misaligned videos can cause 'arena detection' errors, causing missed events (Figure 4— figure supplement 1A, video#18) or false positives (Figure 4— figure supplement 1, video#13) leading to lower precision, recall and F1 scores.

Circling

The male circles around the female, which is a unique male behavior during courtship. Studies have shown that a refractory female gets restimulated by the circling behavior of the males (Kessler, 1962). In addition, differences in the frequency of circling behavior and other courtship rituals is used to maintain reproductive isolation among different species (Brown, 1965).

To evaluate the 'DANCE circling classifier', we used both the decapitated virgin female (Figure 5) and mated female dataset (Figure 5— Figure Supplement 1). The circling index from the ground-truth and the DANCE classifier output are comparable; however, MateBook underscored these events (Figure 5A), most likely due to identity switches. DANCE circling classifier matched well from the ground-truth for individual videos (Figure 5A, Friedman's ANOVA with Dunn's test; $p<0.2049$, $n=12$). Overall circling index from all 12 videos from the ground-truth scores and the DANCE circling classifier were comparable

290 (Figure 5B), whereas MateBook significantly underscores circling in individual videos
291 (Figure 5A, Friedman's ANOVA with Dunn's test; ****$p<0.0001$, n=12). We found a strong
292 correlation between the DANCE circling classifier index to ground-truth (Figure 5C) but a
293 weaker correlation with MateBook (Figure 5D). The DANCE circling classifier has a
294 precision rate and a recall rate of 98.04% and 92.09%, with a F1 score of 94.97% (Figure
295 5E), suggesting robustness of the classifier.
296
297 **Following**
298 During 'following', the male follows around the female if it is moving in order to get closer to
299 it and initiates the other courtship behaviors such as wing extension, licking, etc. (Spieth,
300 1968). Since following is a relatively simpler behavior, both MateBook and the 'DANCE
301 following classifier' performed well when compared to ground-truth (Figure 5— figure
302 supplement 2A-D). However, the DANCE classifier had better precision, recall, and F1
303 scores (91.16%, 91.07%, and 91.12%) as compared to MateBook (65.77%, 83.39%, and
304 73.53%), indicating lower number of false positives and false negatives (Figure 5— figure
305 supplement 2E).
306
307 **DANCE hardware**
308 To measure aggression and courtship in flies, existing setups (Figure 1A, Figure 1— figure
309 supplement 1-2) (Dankert et al., 2009; Koemans et al., 2017) have several bottlenecks,
310 making it challenging to adopt these assays by the broader neuroscience community. For
311 instance, it requires complex and specialized hardware (Dankert et al., 2009; Koemans et
312 al., 2017), necessitating access to sophisticated machine shops; or 3D printers to fabricate
313 behavioral arenas (Chowdhury et al., 2021; Gil-Martí et al., 2023). Setting up aggression
314 assays (Dankert et al., 2009; Dierick, 2007) requires coating chambers with fluon to prevent

315 flies from walking on the walls, which is tedious. Recording the videos require expensive
316 machine vision cameras and machine vision backlights and technical expertise for data
317 acquisition and processing. To overcome these challenges, we devised the DANCE
318 hardware (Figure 6), which provides an inexpensive, easy, scalable, and robust alternative
319 to existing methods of recording *Drosophila* aggression and courtship behaviors.

320 The DANCE hardware consists of easy to find off-the-shelf components including
321 transparent medicine tablet foils/blister packs of different diameters, which are repurposed
322 to serve as arenas for recording behaviors; acrylic sheets; and paper tape (Figure 6A-D,
323 Supplementary video 1-2). To substitute expensive machine vision cameras, we used
324 widely available Android smartphone cameras for recording and expensive machine vision
325 backlights were replaced with electronic tablets or smartphones running white light
326 application, to provide uniform backlight/illumination source (Figure 6E-G). Aggression and
327 courtship behaviors were recorded at 30 fps with 1080 pixel resolution. In the DANCE
328 aggression hardware (Figure 6A-B, E) medicinal foil is slid over a base plate containing an
329 apple juice agar food layer over which flies fight (Supplementary Video 1-2). For DANCE
330 courtship chambers (Figure 6C-D, F-G), the medicine tablet foil is slit from the middle using
331 a razor, through which a thin 'separator comb' is inserted, which is made using x-ray film.
332 This 'separator comb' keeps the males and females separate until the beginning of the
333 recording, when it is manually removed (Supplementary video 3-4).

334 Since the heat generated from the tablet/smartphone screen during recording can affect
335 the aggression and courtship behaviors, we placed a transparent acrylic sheet with an air-
336 gap of 4 mm on top of the backlight screen to facilitate heat dissipation (Figure 6E-G;
337 Supplementary Video 2, 4), which proved to be crucial to obtain robust data (Figure 7).

338

339 **Benchmarking DANCE hardware**

340 We used our DANCE classifiers to quantify aggression and courtship behaviors and found
341 that wild type flies display these behaviors at similar levels in DANCE hardware as in the
342 existing set ups (Figure7, Supplementary video 5). We investigated the effects of social
343 isolation *vs.* enrichment on aggression and courtship behaviors, since single-housing (SH)
344 when compared to group housing (GH), increases courtship attempts (Dankert et al., 2009;
345 Kim and Ehrman, 1998; Pan and Baker, 2014), and promotes aggression (Agrawal et al.,
346 2020; Wang et al., 2008; Yadav et al., 2024). We found comparable performance between
347 pre-existing courtship setup (Koemans et al., 2017) (Figure 7A) and the DANCE hardware
348 (Figure 7B), both of which captured significant differences in multiple courtship behaviors
349 in SH *vs.* GH males (Figure 7C-J).

350 Comparison of pre-existing set up for aggression (Dankert et al., 2009) (Figure 7K) with the
351 DANCE hardware (Figure 7L), showed that both the set ups faithfully captured even low
352 levels of aggressive lunges in the wild type SH flies (Figure 7M-N). Availability of food
353 affects *Drosophila* aggression (Lim et al., 2014) and yeast in fly food was shown to interact
354 with gut microbiome to affect male aggression (Jia et al., 2021). We therefore investigated
355 if the type of yeast used in food would affect aggression. Substituting yeast granules in fly
356 food with yeast extract powder reduced the baseline aggression when flies were reared and
357 differentially housed on food with yeast extract powder (Figure 7O). To investigate the effect
358 of genetic knockdown using RNAi, which is a widely used tool in *Drosophila* genetics, we
359 downregulated neuropeptide *Drosulfakinin* (Dsk) in insulin-producing neurons using *dilp2-*
360 *GAL4*. Consistent with a previous finding (Agrawal et al., 2020), aggressive lunges were
361 significantly increased in SH males (Figure 7P-Q). Behavioral screens often utilize
362 optogenetics to identify neuronal circuits responsible for aggression (Hoopfer et al., 2015;
363 Wohl et al., 2023; Yadav et al., 2024). It was earlier shown that while specific dopaminergic
364 neurons may play a role in aggression, constitutive inactivation of a larger set of

365 dopaminergic neurons using *TH*-GAL4 led to unhealthy flies with defects in locomotion due
366 to which they did not fight (Alekseyenko et al., 2013). We reasoned that modern optogenetic
367 tools may overcome the limitations of constitutive silencing and sought to investigate the
368 role of dopaminergic neurons using our DANCE set up. We carried out optogenetic
369 silencing during 20 minute interactions, by driving green light sensitive anion
370 channelrhodopsin- *GtACR1* (Govorunova et al., 2015; Mohammad et al., 2017) using *TH*-
371 GAL4, and found significant increase in aggressive lunges in SH flies (Figure 7R). We also
372 observed higher wing flicks as well as high intensity aggressive behaviors such as boxing
373 and tussling (Supplementary Video 6). We did not find significant changes in locomotor
374 activity between GH and SH flies despite 12 hours of continuous silencing (Figure 7— figure
375 supplement 1). Together this suggests that the DANCE set up can be used with most of the
376 common neurogenetic tools to screen for underlying molecular and neuronal circuitry
377 governing aggression and courtship behaviors.

378

379 **Discussion**

380 Here, we present the DANCE assay, an easy to use and robust analysis pipeline and
381 inexpensive hardware to record and quantify aggression and courtship behaviors. We
382 developed six novel behavioral classifiers using supervised learning to accurately quantify
383 aggression and courtship behaviors of *Drosophila* males. We developed a versatile, easy
384 to fabricate and inexpensive hardware setup costing <$0.30, about 10,000 fold
385 improvement over existing set ups, using easily available, off-the-shelf components. The
386 performance of DANCE set up is comparable with more specialized and expensive existing
387 setups. Therefore, DANCE enables rapid behavioral screening and wider adoption by the
388 neuroscience community including in resource limited settings such as undergraduate
389 education.

390 Various components of the DANCE assay including the behavioral classifiers, hardware
391 design and analysis codes are publicly available and can be used independently as well.
392 This offers flexibility to the neuroscience community to further customize the DANCE
393 classifiers for specific needs or incorporate additional data without having to develop a
394 classifier from scratch. Our study can also serve as a template for development of future
395 behavioral classifiers for *Drosophila* aggression such as fencing, wing flick, box, tussle,
396 chase, female headbutt etc. (Chen et al., 2002; Nilsen et al., 2004) and courtship behaviors,
397 such as- tapping by male, licking the posterior abdomen of the female, female rejection etc.
398 (Sokolowski, 2001).
399 Such high resolution analysis of complex social interactions could give us insights into
400 mating dynamics, sexual selection and the influence of genetics and evolution. Further, the
401 behavioral dynamics can help uncover sequencing of dynamic components of behaviors
402 (Nilsen et al., 2004; Seeds et al., 2014; Simon and Heberlein, 2020; Zhang et al., 2020). It
403 can also serve as a template for studying complex behaviors of various species of
404 *Drosophila* and other insects. The ease of adaptability and portability of the DANCE assay
405 can also be useful for ethologists studying insect behaviors closer to their natural habitat.
406

407 **Materials and Methods**

408 The details of all the custom codes, analysis pipeline, sample files to run the analysis and
409 DANCE classifiers are available in our GitHub repository at-
410 https://github.com/agrawallab/DANCE.

411 **Fly Husbandry**

412 Flies were reared on standard food at 25°C and 65% relative humidity with a 12hr:12hr
413 light-dark cycle. All assays were performed at 25°C with 65% relative humidity, unless
414 mentioned otherwise. For aggression and courtship experiments, Canton-S male flies were

415 collected within 24 hours of eclosion and housed in a group (20 male flies per vial, 90mm
416 length and 25mm diameter) or isolated (1 male fly per vial, 70mm length and 10mm
417 diameter) for 6 days.

418 The following fly lines were acquired from the Bloomington Drosophila Stock Center
419 (BDSC), USA: *TH*-GAL4, *Dilp2*-GAL4 (RRID:BDSC_37516), *Dsk*-RNAi
420 (RRID:BDSC_25869) and attP2 empty vector control (RRID:BDSC_36303). The *UAS-*
421 *GtACR1* flies were a gift from Gaurav Das, NCCS, Pune, India and Canton-S (CS) flies
422 were obtained from Ulrike Heberlein, HHMI, Janelia Research Campus, USA.

423 **Aggression assay**

424 The traditional aggression assay was performed as described previously (Dankert et al.,
425 2009; Dierick, 2007). In brief, the behavioral chamber is made up of 12 well-aggressive
426 arenas (10 mm in height and 16 mm in diameter/arena). These arenas were covered by a
427 sliding lid with 2mm loading holes to facilitate the introduction of flies. A pair of male flies
428 that were housed either as a group (GH) or single fly (SH) were introduced into the arena
429 wells by gentle mouth aspiration through the loading holes. After the flies were loaded, the
430 sliding lid was tightened using screws. Fluon (Insect-a-slip, Bioquip: cat #2871B) was
431 coated on the arena walls and was allowed to dry overnight to create a slippery surface,
432 thereby preventing the flies from climbing the walls. Sigmacote (Sigma-Aldrich: SL2) was
433 used to coat the sliding lid of the arena to prevent flies from walking on the arena ceiling.
434 The chamber was placed on the food plate containing commercial apple juice (without
435 added sugars), 2.5% w/v of sucrose, and 2.25% w/v of agarose. For experiments to test
436 effects of fly food nutrients, either 2.4% yeast extract powder (HiMedia: RM0271) or 2.4%
437 Yeast granules (Prime Instant Dry yeast, AB Maury, India) were mixed in fly food. For
438 optogenetic silencing experiments, 520-540 nm wavelength of LEDS (Lumileds:2835) were
439 used at 0.0004 μW intensity, measured using a Power meter (Newport:843-R). For all

440 aggression assays, flies were allowed to acclimatize in the arena for 5 minutes, and then
441 the activity was filmed for 20 minutes. The assays were performed during ZT0 - ZT2.5 i.e.
442 first 2.5 hours of the morning activity peak.

443 **Courtship assay**

444 Traditional courtship assays were performed as described by (Koemans et al., 2017;
445 Ribeiro et al., 2018). We introduced single pairs of male and female flies using an aspirator
446 in an 18-well courtship chamber, consisting of individual wells/arenas with a diameter of 10
447 mm. Male and female flies were introduced into one half of the chamber using the sliding
448 entry holes with a removable separator that divided the chamber into two halves. The flies
449 were allowed to acclimatize to the arena for 5 minutes, after which the separator was
450 removed, and the courtship behavior was recorded for 15 minutes at 30 fps using a white
451 backlight. The assay was performed from ZT0 to ZT3 or ZT9 to ZT12 (during peak activity
452 windows). For generating mated females, we housed 20 females with 10 males together in
453 a single vial for 4-6 days. For decapitated virgins, 2-4 days old virgin females were collected
454 in a vial, and the decapitation was performed after anesthetizing on a CO2 fly-pad, just
455 before the assay. The assays were performed during ZT0-ZT2.5 i.e., first 2.5 hours/morning
456 activity peak or ZT9-ZT11.5 i.e. last 2.5 hours/evening activity peak.

457

458 **MateBook**

459 To quantify courtship data, an automated software called MateBook was used
460 (https://github.com/Dicksonlab/MateBook) (Ribeiro et al., 2018). MateBook tracks the fly's
461 location and position for each video frame using the machine-vision approach. The software
462 has built-in classifiers for each of the behaviors of a male fly during courtship, such as
463 following, wing extension, orientation, copulation, and circling. The quantified data is
464 generated in the form of a .tsv file, which contains bouts and other parameters related to

465 courtship behavior. The software also provides an ethogram, which is a color-coded
466 timeline of all the behaviors.

467 **DANCE hardware**

468 Circular, transparent medicinal tablet foils/blister packs were used to serve as aggression
469 or courtship arenas, and acrylic sheets were used as base plates, and the DANCE
470 hardware setup is held together using paper tape. The DANCE aggression arena was 13
471 mm in diameter and 4 mm in height. A thin acrylic base plate (2 mm thickness) with a
472 loading hole (2 mm diameter) and a base plate with food and a side spacer were held
473 together using paper tape (Supplementary Video 1). The tip of the food plate was also
474 covered using paper tape for easy sliding of the medicinal tablet foil without damaging the
475 food layer .

476 Sigmacote (Sigma-Aldrich: SL2) was used to coat the walls and roof of the medicine blister
477 foil using cotton swab (Solimo, Amazon India), to prevent flies from climbing the walls of
478 the arena. A pair of male files were introduced into the medicinal foil wells by gentle
479 aspiration through the loading holes. After loading flies into the medicinal foil wells, the foil
480 was slid onto the food bed with the help of side spacers and tape on the tip of the food bed.
481 After sliding it, the assembly was taped from ends to avoid gaps and prevent flies from
482 escaping the arenas (Supplementary Video 2).

483 The DANCE courtship arena used transparent medicine blister packs with a diameter of
484 11mm and a height of 4.5 mm. The top surface across all arenas (5 in a row) had thin slits
485 cut using a sharp razor to fit in the separator comb which was built by cutting used x-ray
486 film. The foil and base plate (2 mm thickness) with loading hole (2 mm diameter) was held
487 together by a paper tape. The base plate for courtship had two holes spaced apart in such
488 a way that, male and female flies can be loaded on either side of the separator comb. The
489 assembly was taped from ends to avoid gaps and prevent flies from escaping the arenas.

490 Once the setup was ready to record, the separator comb was gently lifted up, while ensuring
491 the slit edges stay intact to avoid distortions during recordings (Supplementary videos 3-4).

492 **Video acquisition**

493 For the traditional set ups, the interaction of the flies was recorded using machine-vision
494 cameras (DMK 33UX252 USB 3.0 monochrome camera). A white backlight (TMS, BHS4-
495 00-100-X-W-24V) provided the light source for both the courtship and aggression
496 experiments. Videos were recorded at 30 frames per second (fps) for 15 minutes for
497 courtship or 20 minutes for aggression in H.264 (.mp4) format with 1440x1080 resolution.
498 These videos were used for training, testing, and validating the DANCE classifiers.

499 Alternatively, for DANCE hardware testing, we used various Android smartphone cameras
500 (Huawei Y9 2019; OnePlus Nord CE 2 Lite 5G, model: CPH2381 or Redmi Note 11 Pro+
501 5G, model: 221116SI) at 30 fps with 1080p resolution in H.264 (.mp4) format for 15 minutes
502 (courtship) or 20 minutes (aggression) with electronic tablet (iPad Air, 5th Generation) or
503 smartphone (iPhone 13) running a 'white screen light app' as the background illumination
504 source. A transparent acrylic sheet with 4 mm spacers was kept on top of this 'backlight' to
505 create an air gap to ensure heat exchange and to prevent the DANCE arenas from getting
506 hot.

507 **Tracking flies using FlyTracker**

508 To successfully classify the behavior of the flies, we tracked their location and body parts
509 using Caltech FlyTracker (Eyjolfsdottir et al., 2014). It allows for detection and tracking of
510 flies based on their location, body position, orientation, and interactions across the frames.
511 Occurrence of identity switches between male and female flies was corrected using the
512 'identity correction' tool available in the "visualizer" program of the FlyTracker package. This
513 data was then pushed to the JAABA pipeline to develop DANCE classifiers.

514 **Developing the DANCE behavioral classifiers**

515 The Janelia Automatic Animal Behavior Annotator (JAABA) was used to develop DANCE
516 behavioral classifiers. The algorithm allows experienced users to 'encode their intuition' to
517 develop 'classifiers' that help to annotate various animal behaviors using tracking data from
518 the behavioral recordings (Kabra et al., 2013).

519 Tracking data generated using FlyTracker from training videos were processed using
520 JAABA. The stereotypic behavior bouts were labeled as 'true behavior', e.g., 'Lunge', 'Wing
521 extension' etc., and an approximately equal number of obvious non-behavior bouts
522 ('None'), were labeled. After each round of training, the false positives predicted by the
523 classifier would be specifically labeled as 'None.' This cycle of training and correcting the
524 predictions was done until there was no improvement in the classifier's performance. We
525 included courtship videos of males with decapitated virgins in the training set for
526 the attempted copulation and circling classifiers to increase the number of labeled bouts of
527 True behavior, as these behaviors occur more frequently with decapitated virgins.
528 Additionally, we included videos with decapitated virgins in the training sets for
529 the following and copulation classifiers to provide more examples of the 'None' behavior.
530 For the wing extension classifier, we used only mated females, as there were an adequate
531 number of bouts representing both the true behavior and the 'None' behavior.

532 Various DANCE classifiers were developed using following frames: lunge: 2730 frames (91
533 seconds in total); wing extension: 99,947 frames (3,331.57 seconds in total); attempted-
534 copulation: 39,513 frames (1,317.1 seconds in total); copulation: 56,979 frames (1899.3
535 seconds in total); circling: 14,396 frames (479.87 seconds in total); following: 25,787 frames
536 (859.57 seconds in total).

537 DANCE classifiers, custom codes and other analysis details are available in our GitHub
538 repository at- https://github.com/agrawallab/DANCE.

539 **Characterization of male aggression and courtship behaviors**

540 Lunge: It is defined as 'the attacking fly rises on his hind legs, lifting his long body axis by
541 45° then snaps down on his opponent with his head reaching a velocity of 200 mm/s'
542 (Dankert et al., 2009).

543 Courtship classifiers were developed for five male courtship behaviors based on existing
544 definitions (Dankert et al., 2009; Ribeiro et al., 2018). For the duration-based classifiers, a
545 filter of minimum bout length was added in the post-processing of each classifier as a
546 measure to eliminate the bouts shorter than 98% of manual bouts. Details of individual
547 courtship classifiers are mentioned below.

548 Wing Extension: Dankert et.al 2009, Ribeiro et al., 2018 define a wing extension, when the
549 angle between the body's main axis and the line from the body's center to the tip of the
550 wing exceeds 30°, with the behavior persisting for approximately 13 frames i.e. 0.5s, at 25
551 fps. We processed our data using MateBook as described (Ribeiro et al., 2018); but with a
552 reduced persistence filter for wing extension to 0.33 seconds from the default 0.50 seconds
553 to match our classifier's threshold of minimum 10 frames, since in all the courtship videos
554 that we analyzed, >98% of the bouts followed this threshold.

555 Attempted copulation and copulation: Existing definitions, such as (Ribeiro et al., 2018)
556 define copulation as 'occlusions' persisting for more than 45 seconds and does not quantify
557 'attempted-copulation'. We used a similar threshold while training our classifiers but
558 ensured that it is not merely an 'occlusion'— if the male curls it's abdomen without mounting
559 or mounts the female for a minimum of 0.33 seconds (10 frames at 30 fps) to a maximum
560 of <45 seconds, it was considered as 'attempted-copulation' whereas mounting for >45

561 seconds (>1350 frames at 30 fps) was considered as 'copulation'. The results were
562 compared with MateBook by matching its 'event settings'.

563 Circling: (Dankert et al., 2009) define circling as the fly drifting sideways in a circle with
564 approximately constant velocity. Additionally, (Ribeiro et al., 2018) considered this behavior
565 with a persistence of 13 frames (0.5s at 25 fps). We developed our 'DANCE circling
566 classifier' based on these existing definitions, but with a reduced persistence of 10 frames
567 (0.33s at 30 fps) and compared the results with MateBook by matching its 'event settings'.

568 Following: Ribeiro et al., 2018 define the event following as when the 'the male keeps
569 distance to the female between 2 and 5mm, while directly behind the female, while both
570 flies are walking with a minimum speed of 2mm/s' with the behavior persisting for 25 frames
571 i.e. 1.0s , at 25 fps. We developed our 'DANCE following extension classifier' based on
572 these existing definitions, but with a reduced persistence of 10 frames (0.33s, at 30 fps)
573 and compared the results with MateBook by matching its 'event settings'.

574 **Manual behavioral annotations**

575 To quantitatively measure the classifier's performance, manual behavioral annotations, or
576 'ground-truthing' were performed using the JAABA's ground-truthing mode (Kabra et al.,
577 2013). The classifier's performance is measured by its robustness in predicting the
578 behavioral bouts for the frames it was not trained on, which comprises the 'testing set.' In
579 the testing set, we first annotate the behavioral bouts as 'true behavior' without looking at
580 the predictions of the output. Then these videos are processed with the same classifier
581 using JAABAPlot and compared as described subsequently.

582 **Comparison of manual and DANCE annotations**

583 The classifier's output was compared with that of the manual ground-truth, based on both
584 the number of bouts for the one-frame lunge classifier and the bout duration for all duration-

585 based courtship classifiers (wing extension, following, circling, attempted-copulation, and
586 copulation). The duration of the bouts was used to derive the respective behavioral index.
587 A regression analysis and classification metrics were calculated between the bout numbers
588 or the behavior index based on the type of classifier for each of the ground-truth, DANCE,
589 and the existing software's scores using Prism 8 (GraphPad Software).

590 The manual scores and JAABA output were compared, bout by bout, for the total number
591 of true positives, false positives, and false negatives. Human annotated bouts (ground-
592 truth) were used as a reference to the number of true positives. A 'true positive' is called
593 when a JAABA bout overlaps with ground-truth bout for one or more frames (over 0.03 ms).
594 If one ground-truth bout overlapped with two or more JAABA bouts, then all the JAABA
595 bouts would be collectively counted as one bout, i.e., one true positive bout. Similarly, if two
596 or more ground-truth bouts overlapped with a single JAABA bout, then the number of
597 ground-truth bouts would be the count for the JAABA bout; i.e., two or more number of true
598 positives.

599 When a JAABA bout does not overlap with any ground-truth (human annotation), then it
600 would be considered a false positive. When a ground-truth bout does not overlap with any
601 JAABA bout, it is considered a false negative (Leng et al., 2020).

602 We used a custom Python code, to calculate the number of total overlaps between the
603 ground-truth, the DANCE classifier's, and the existing algorithm's output. Once we had the
604 total number of true positives, false positives, and false negatives, we derived the precision,
605 recall and F1 scores to gain a deeper understanding of the performance of DANCE
606 classifiers. Precision rate represents the ratio of correctly predicted positive observations to
607 the total predicted positives; the recall rate represents the ratio of correctly predicted
608 positive observations to all actual positives; and the F1 score is a single metric which is a

609 harmonic mean of both precision and recall. We used the following formulas to calculate
610 precision, recall, and F1 score:

611 $$precision = \frac{(number\ of\ true\ positives)}{(number\ of\ true\ positives) + (number\ of\ false\ positives)}$$

612 $$recall = \frac{(number\ of\ true\ positives)}{(number\ of\ true\ positives) + (number\ of\ false\ negatives)}$$

613 $$F1\ score = 2\ x\ \frac{(precision\ x\ recall)}{(precision +\ recall)}$$

614

615 **Statistical analysis**

616 For non-normally distributed data, Mann-Whitney U test or ANOVA with appropriate
617 posthoc correction was used. Statistical analysis of behavioral data was performed using
618 Prism 8 (GraphPad Software), custom scripts or Microsoft Excel.

619
620 **Supplementary Video legends:**

621 1. 3D rendered DANCE Aggression hardware.
622 2. Using DANCE hardware setup for recording aggression.
623 3. 3D rendered DANCE Courtship hardware
624 4. Using DANCE hardware setup for recording Courtship.
625 5. Aggression and courtship behaviors recorded in DANCE hardware.
626 6. Optogenetic silencing of dopaminergic neurons in DANCE shows increased
627 aggression.

628

629 **Supplementary File descriptions:**

630 1. Bill of material for DANCE set up and comparison with existing setups.
631 2. Comparison between DANCE lunge classifier, ground-truth and existing methods.
632 3. Definitions for behavioral classifiers.

4. Comparison between DANCE courtship classifiers, ground-truth and MateBook.

Acknowledgement

We thank Barry Dickson from Queensland Brain Institute, Australia for encouraging discussion and suggestions related to MateBook and Ben Arthur, HHMI, Janelia Research Campus, USA for helpful suggestions for setting up the MateBook analysis. We thank Mayank Kabra, Kristin Branson and colleagues for developing JAABA and providing excellent support to the user community. We thank Ulrike Heberlein, HHMI, Janelia Research Campus, USA and Gaurav Das, NCCS, Pune and Bloomington Drosophila Stock Center (NIH P40OD018537) for fly stocks. We thank Santhosh Chidangil, MAHE for help with LED power meter. We thank Santosh D'Mello, LSU Shreveport and Gaurav Das for helpful discussions; Gaurav Das and Toshiharu Ichinose for critical reading and feedback on the manuscript. Figures in the manuscript were created using Biorender.com.

Funding

This work was supported by funding to PA from the Department of Biotechnology (DBT), Ramalingaswami Re-entry Fellowship (BT/RLF, Re-entry/34/2018), and DBT, Research grant (BT/PR36166/BRB/10/1859/2020) by the DBT, Ministry of Science and Technology, Government of India. RSPY was supported by a TMA Pai fellowship from MAHE and DBT, Research grant to PA. FA was supported by Ramalingaswami fellowship, DBT, India to PA. TK was supported by DBT, Research grant to PA. MV is supported by NFST, Ministry of Tribal Affairs, Government of India. SA is supported by the Department of Science and Technology (DST) Inspire Fellowship. PPP is supported by a TMA Pai fellowship from MAHE, India. SBS is supported by the Anusandhan National Research Foundation (ANRF), Ministry of Science and Technology, Government of India, grant to PA (CRG/2022/006846).

Competing interests: RSPY, FA and PA are listed as inventors on a published patent related to this work by the Indian patent office, titled: 'Device For Measuring Complex Social Behaviors In Small Insects', Application No. 202441072884.

References

Agrawal P, Kao D, Chung P, Looger LL. 2020. The neuropeptide Drosulfakinin regulates social isolation-induced aggression in *Drosophila*. *J Exp Biol* **223**:jeb207407.

Alekseyenko OV, Chan Y-B, Li R, Kravitz E a. 2013. Single dopaminergic neurons that modulate aggression in Drosophila. *Proc Natl Acad Sci U S A* **110**:6151–6156.

Anderson DJ, Perona P. 2014. Toward a Science of Computational Ethology. *Neuron* **84**:18–31.

Asahina K. 2017. Neuromodulation and Strategic Action Choice in Drosophila Aggression. *Annu Rev Neurosci* **40**:51–75.

Asahina K, Watanabe K, Duistermars BJ, Hoopfer E, González CR, Eyjólfsdóttir EA, Perona P, Anderson DJ. 2014. Tachykinin-Expressing Neurons Control Male-Specific Aggressive Arousal in Drosophila. *Cell* **156**:221–235.

Bastock M, Manning A. 1955. The Courtship of Drosophila Melanogaster. *Behaviour* **8**:85–110.

Bennet-Clark HC, Ewing AW. 1969. Pulse interval as a critical parameter in the courtship song of Drosophila melanogaster. *Anim Behav* **17**:755–759.

Benzer S. 1967. BEHAVIORAL MUTANTS OF Drosophila ISOLATED BY COUNTERCURRENT DISTRIBUTION. *Proc Natl Acad Sci U S A* **58**:1112–1119.

Branson K, Robie A a., Bender J, Perona P, Dickinson MH. 2009. High-throughput ethomics in large groups of Drosophila. *Nat Methods* **6**:451–457.

Brown RGB. 1965. Courtship behaviour in the Drosophila obscura group. Part ii. Comparative studies. *Behaviour* **25**:281–322.

Chen S, Lee AY, Bowens NM, Huber R, Kravitz EA. 2002. Fighting fruit flies: a model system for the study of aggression. *Proc Natl Acad Sci U S A* **99**:5664–5668.

Chiu H, Hoopfer ED, Coughlan ML, Pavlou HJ, Goodwin SF, Anderson DJ. 2021. A circuit logic for sexually shared and dimorphic aggressive behaviors in Drosophila. *Cell* **184**:507-520.e16.

Chowdhury B, Wang M, Gnerer JP, Dierick HA. 2021. The Divider Assay is a high-throughput pipeline for aggression analysis in Drosophila. *Communications Biology 2021 4:1* **4**:1–12.

Cook R, Cook A. 1975. The attractiveness to males of female Drosophila melanogaster: Effects of mating, age and diet. *Anim Behav* **23**:521–526.

Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P. 2009. Automated monitoring and analysis of social behavior in Drosophila. *Nat Methods* **6**:297–303.

Datta SR, Anderson DJ, Branson K, Perona P, Leifer A. 2019. Computational Neuroethology: A Call to Action. *Neuron* **104**:11–24.

Davis SM, Thomas AL, Liu L, Campbell IM, Dierick HA. 2018. Isolation of aggressive behavior mutants in drosophila using a screen for wing damage. *Genetics* **208**:273–282.

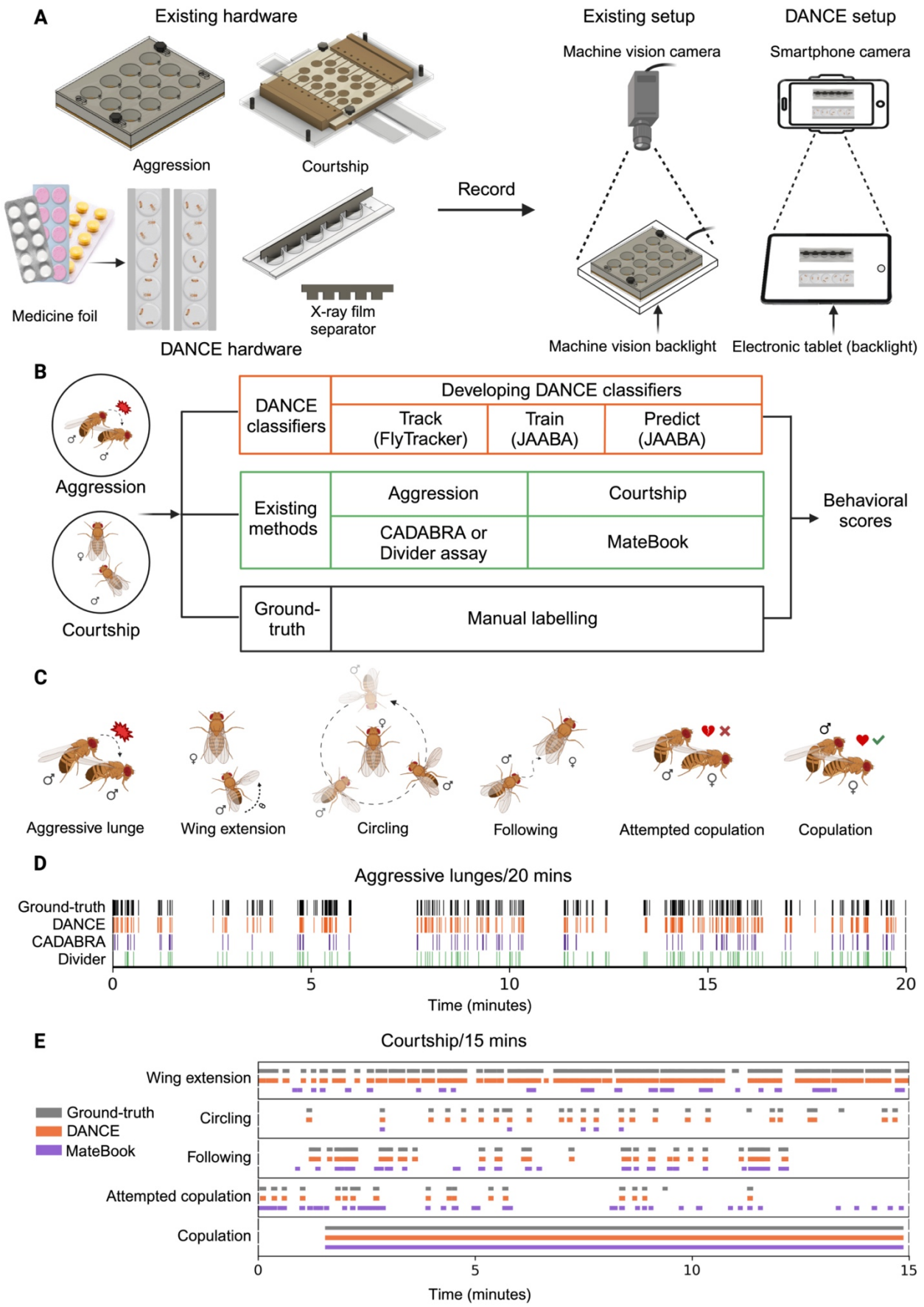
Davis SM, Thomas AL, Nomie KJ, Huang L, Dierick HA. 2014. Tailless and Atrophin control Drosophila aggression by regulating neuropeptide signalling in the pars intercerebralis. *Nat Commun* **5**:3177.

Dickson BJ. 2008. Wired for sex: The neurobiology of Drosophila mating decisions. *Science* **322**:904–909.
Dierick HA. 2007. A method for quantifying aggression in male drosophila melanogaster. *Nat Protoc* **2**:2712–2718.
Dow MA, Schilcher F von. 1975. Aggression and mating success in Drosophila melanogaster. *Nature* **254**:511–512.
Eyjolfsdottir E, Branson S, Burgos-Artizzu XP, Hoopfer ED, Schor J, Anderson DJ, Perona P. 2014. Detecting Social Actions of Fruit Flies. Computer Vision – ECCV 2014. pp. 772–787.
Fowler GL. 1973. Some Aspects of the Reproductive Biology of Drosophila: Sperm Transfer, Sperm Storage, and Sperm Utilization In: Caspari EW, editor. Advances in Genetics. Academic Press. pp. 293–360.
Gill KS. 1963. A mutation causing abnormal courtship and mating behavior in males of Drosophila melanogaster. *Am Zool* **3**:507.
Gil-Martí B, Barredo CG, Pina-Flores S, Poza-Rodriguez A, Treves G, Rodriguez-Navas C, Camacho L, Pérez-Serna A, Jimenez I, Brazales L, Fernandez J, Martin FA. 2023. A simplified courtship conditioning protocol to test learning and memory in Drosophila. *STAR Protoc* **4**:101572.
Gomez-Marin A, Paton JJ, Kampff AR, Costa RM, Mainen ZF. 2014. Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nat Neurosci* **17**:1455–1462.
Govorunova EG, Sineshchekov OA, Janz R, Liu X, Spudich JL. 2015. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **349**:647–650.
Greenspan RJ, Ferveur J-F. 2000. Courtship in Drosophila. *Annu Rev Genet* **34**:205–232.
Hall JC. 2002. Courtship lite: a personal history of reproductive behavioral neurogenetics in Drosophila. *J Neurogenet* **16**:135–163.
Hall JC. 1978. Courtship among males due to a male-sterile mutation in Drosophila melanogaster. *Behav Genet* **8**:125–141.
Hoffmann AA. 1987a. A laboratory study of male territoriality in the sibling species Drosophila melanogaster and D. simulans. *Anim Behav* **35**:807–818.
Hoffmann AA. 1987b. Territorial Encounters Between Drosophila Males of Different Sizes. *Anim Behav* **1**:70–71.
Hoopfer ED, Jung Y, Inagaki HK, Rubin GM, Anderson DJ. 2015. P1 interneurons promote a persistent internal state that enhances inter-male aggression in Drosophila. *eLife* **4**:1–27.
Hoyer SC, Eckart A, Herrel A, Zars T, Fischer S a., Hardie SL, Heisenberg M. 2008. Octopamine in male aggression of Drosophila. *Curr Biol* **18**:159–167.
Ishii K, Cortese M, Leng X, Shokhirev MN, Asahina K. 2022. A neurogenetic mechanism of experience-dependent suppression of aggression. *Sci Adv* **8**:eabg3203.
Jacobs ME. 1960. Influence of Light on Mating of Drosophila Melanogaster. *Ecology* **41**:182–188.
Jia Y, Jin S, Hu K, Geng L, Han C, Kang R, Pang Y, Ling E, Tan EK, Pan Y, Liu W. 2021. Gut microbiome modulates Drosophila aggression through octopamine signaling. *Nat Commun* **12**. doi:10.1038/s41467-021-23041-y
Jung Y, Kennedy A, Chiu H, Mohammad F, Claridge-Chang A, Anderson DJ. 2020. Neurons that Function within an Integrator to Promote a Persistent Behavioral State in Drosophila. *Neuron* **105**:322-333.e5.
Kabra M, Robie A a., Rivera-Alba M, Branson S, Branson K. 2013. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods* **10**:64–67.

755 Kessler S. 1962. Courtship rituals and reproductive isolation between the races or
756 incipient species of Drosophila paulistorum. *Am Nat* **96**:117–121.
757 Kim Y-K, Ehrman L. 1998. Developmental Isolation and Subsequent Adult Behavior of
758 Drosophila paulistorum.IV. Courtship. *Behav Genet* **28**:57–65.
759 Kim Y-K, Saver M, Simon J, Kent CF, Shao L, Eddison M, Agrawal P, Texada M, Truman
760 JW, Heberlein U. 2018. Repetitive aggressive encounters generate a long-lasting
761 internal state in *Drosophila melanogaster* males. *Proceedings of the National*
762 *Academy of Sciences* **115**:1099–1104.
763 Koemans TS, Oppitz C, Donders RAT, van Bokhoven H, Schenck A, Keleman K, Kramer
764 JM. 2017. Courtship Conditioning As a Measure of Learning and Memory. *J Vis*
765 *Exp*. doi:10.3791/55808
766 Kravitz EA, Fernandez M de la P. 2015. Aggression in Drosophila. *Behav Neurosci*
767 **129**:549–563.
768 Leng X, Wohl M, Ishii K, Nayak P, Asahina K. 2020. Quantifying influence of human
769 choice on the automated detection of Drosophila behavior by a supervised
770 machine learning algorithm. *PLoS One* **15**:e0241696.
771 Lim RS, Eyjólfsdóttir E, Shin E, Perona P, Anderson DJ. 2014. How food controls
772 aggression in Drosophila. *PLoS One* **9**.
773 MacBean IT, Parsons PA. 1967. Directional selection for duration of copulation in
774 Drosophila melanogaster. *Genetics* **56**:233–239.
775 Markow TA. 1987. Behavioral and sensory basis of courtship success in Drosophila
776 melanogaster. *Proceedings of the National Academy of Sciences* **84**:6200–6204.
777 Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. 2018.
778 DeepLabCut: markerless pose estimation of user-defined body parts with deep
779 learning. *Nat Neurosci* **21**:1281–1289.
780 Mohammad F, Stewart JC, Ott S, Chlebikova K, Chua JY, Koh TW, Ho J, Claridge-Chang
781 A. 2017. Optogenetic inhibition of behavior with anion channelrhodopsins. *Nat*
782 *Methods* **14**:271–274.
783 Nilsen SP, Chan Y-B, Huber R, Kravitz EA. 2004. Gender-selective patterns of aggressive
784 behavior in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **101**:12342–12347.
785 O'Dell KMC. 2003. The voyeurs' guide to Drosophila melanogaster courtship. *Behav*
786 *Processes* **64**:211–223.
787 Pan Y, Baker BS. 2014. Genetic Identification and Separation of Innate and Experience-
788 Dependent Courtship Behaviors in Drosophila. *Cell* **156**:236–248.
789 Pantalia M, Lin Z, Tener SJ, Qiao B, Tang G, Ulgherait M, O'Connor R, Delventhal R,
790 Volpi J, Syed S, Itzhak N, Canman JC, Fernández MP, Shirasu-Hiza M. 2023.
791 Drosophila mutants lacking the glial neurotransmitter-modifying enzyme Ebony
792 exhibit low neurotransmitter levels and altered behavior. *Sci Rep* **13**:10411.
793 Pavlou HJ, Goodwin SF. 2013. Courtship behavior in Drosophila melanogaster: towards a
794 "courtship connectome." *Curr Opin Neurobiol* **23**:76–83.
795 Reza MA, Mhatre SD, Morrison JC, Utreja S, Saunders AJ, Breen DE, Marenda DR.
796 2013. Automated analysis of courtship suppression learning and memory in
797 Drosophila melanogaster. *Fly* **7**:105–111.
798 Ribeiro IMA, Drews M, Bahl A, Machacek C, Borst A, Dickson BJ. 2018. Visual Projection
799 Neurons Mediating Directed Courtship in Drosophila. *Cell* **174**:607-621.e18.
800 Robie AA, Hirokawa J, Edwards AW, Simpson JH, Reiser MB, Branson K, Robie AA,
801 Hirokawa J, Edwards AW, Umayam LA, Lee A, Phillips ML, Card GM, Korff W,
802 Rubin GM, Simpson JH, Reiser MB, Branson K. 2017a. Mapping the Neural
803 Substrates of Behavior Resource Mapping the Neural Substrates of Behavior. *Cell*
804 **170**:393–406.
805 Robie AA, Seagraves KM, Egnor SER, Branson K. 2017b. Machine vision methods for

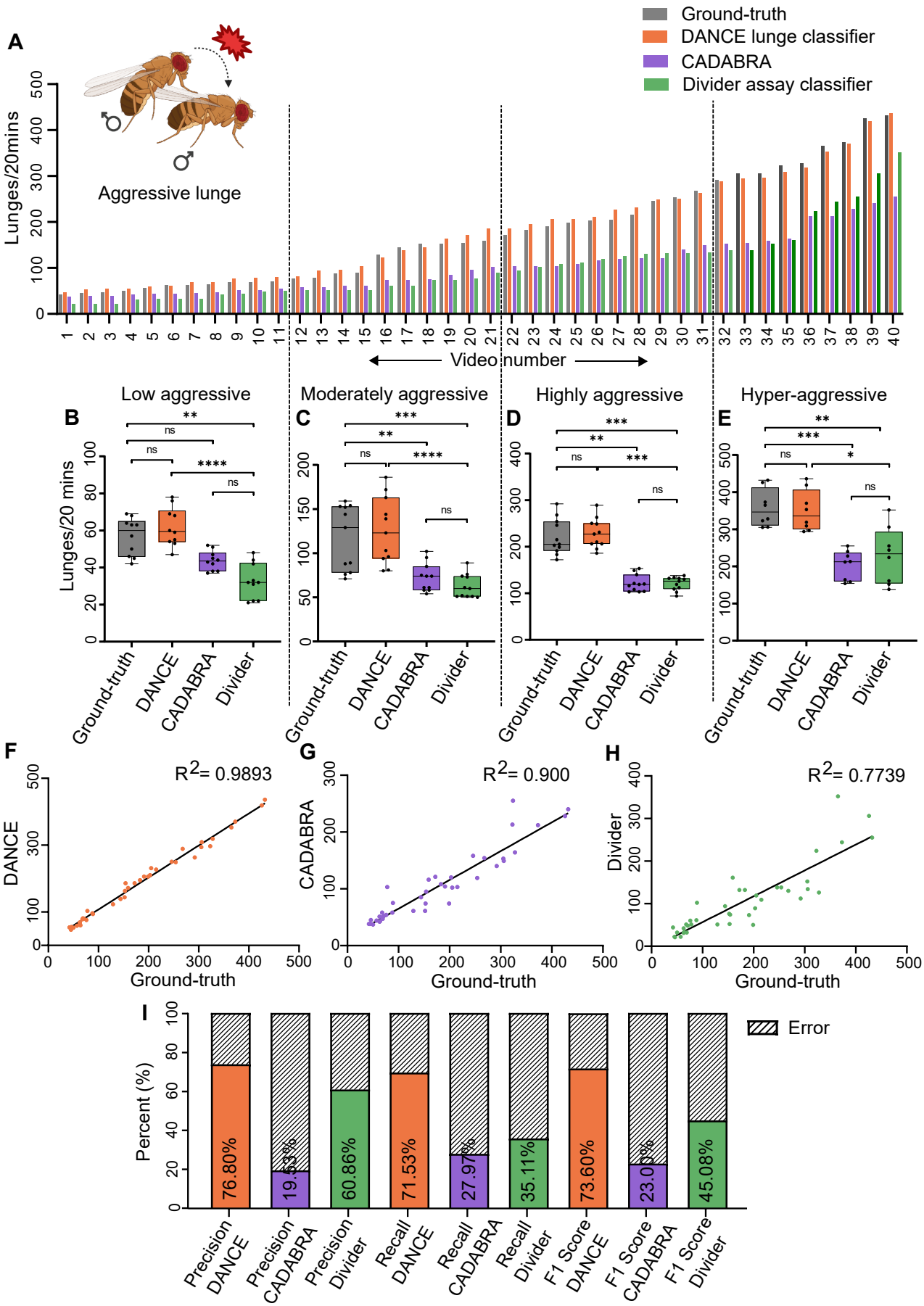
analyzing social interactions. *J Exp Biol* **220**:25–34.
Seeds AM, Ravbar P, Chung P, Hampel S, Midgley FM Jr, Mensh BD, Simpson JH. 2014. A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila. *eLife* **3**:e02951.
Shorey HH. 1962. Nature of the Sound Produced by Drosophila melanogaster during Courtship. *Science* **137**:677–678.
Simon JC, Heberlein U. 2020. Social hierarchy is established and maintained with distinct acts of aggression in male Drosophila melanogaster. *J Exp Biol* **223**:jeb232439.
Sokolowski MB. 2001. Drosophila: genetics meets behaviour. *Nat Rev Genet* **2**:879–890.
Spieth HT. 1974. Courtship behavior in Drosophila. *Annu Rev Entomol* **19**:385–405.
Spieth HT. 1968. Evolutionary Implications of Sexual Behavior in Drosophila In: Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary Biology: Volume 2. Boston, MA: Springer US. pp. 157–193.
Spieth HT. 1966. Drosophilid mating behavior: the behaviour of decapitated females. *Anim Behav* **14**:226–235.
Spieth HT. 1952. Mating Behavior Within the Genus Drosophila (Diptera). Bulletin of the AMNH.
Sturtevant AH. 1915. Experiments on sex recognition and the problem of sexual selection in Drosoophilia. *J Exp Psychol Anim Behav Process* **5**:351–366.
Swain B, von Philipsborn AC. 2021. Sound production in Drosophila melanogaster: Behaviour and neurobiology. Advances in Insect Physiology. pp. 141–187.
Tompkins L, Hall JC, Hall LM. 1980. Courtship-stimulating volatile compounds from normal and mutant Drosophila. *J Insect Physiol* **26**:689–697.
Wang L, Dankert H, Perona P, Anderson DJ. 2008. A common genetic target for environmental and heritable influences on aggressiveness in Drosophila. *Proc Natl Acad Sci U S A* **105**:5657–5663.
Watanabe K, Chiu H, Pfeiffer BD, Wong AM, Hoopfer ED, Rubin GM, Anderson DJ. 2017. A Circuit Node that Integrates Convergent Input from Neuromodulatory and Social Behavior-Promoting Neurons to Control Aggression in Drosophila. *Neuron* **95**:1112–1128.
Wohl MP, Liu J, Asahina K. 2023. Tachykininergic Neurons Modulate the Activity of Two Groups of Receptor-Expressing Neurons to Regulate Aggressive Tone. *J Neurosci* **43**:3394–3420.
Yadav RSP, Ansari F, Bera N, Kent C, Agrawal P. 2024. Lessons from lonely flies: Molecular and neuronal mechanisms underlying social isolation. *Neurosci Biobehav Rev* **156**:105504.
Zhang N, Guo L, Simpson JH. 2020. Spatial comparisons of mechanosensory information govern the grooming sequence in Drosophila. *Curr Biol* **30**:988-1001.e4.

Figures and Figure Legends



847 **Figure 1. DANCE assay overcomes the limitation of existing methods for**
848 **quantification of aggression and courtship behaviors.** (A) Existing hardware for
849 acquisition of aggression and courtship behavior using machine vision cameras and the
850 simplified DANCE hardware. (B) Steps for developing DANCE classifiers and
851 benchmarking with existing methods and manual ground-truthing to produce behavioral
852 scores. (C) Various behavioral classifiers developed to quantify male aggression (lunge)
853 and male courtship (wing extension, circling, following, attempted copulation, and
854 copulation). (D) Raster plots comparing performance of ground-truth *vs*. DANCE classifier
855 *vs*. CADABRA *vs.* Divider assay for aggression in a representative video. (E) Raster plots
856 for various courtship behaviors comparing performance of ground-truth *vs*. DANCE
857 classifiers *vs*. MateBook from representative videos.

858


A
Ground-truth
DANCE lunge classifier
CADABRA
Divider assay classifier
Aggressive lunge
Lunges/20mins
Video number
Low aggressive
Moderately aggressive
Highly aggressive
Hyper-aggressive
B
C
D
E
Lunges/20 mins
Ground-truth
DANCE
CADABRA
Divider
F
G
H
R2= 0.9893
R2= 0.900
R2= 0.7739
DANCE
CADABRA
Divider
Ground-truth
I
Percent (%)
Error
76.80%
19.53%
60.86%
71.53%
27.97%
35.11%
73.60%
23.00%
45.08%
Precision DANCE
Precision CADABRA
Precision Divider
Recall DANCE
Recall CADABRA
Recall Divider
F1 Score DANCE
F1 Score CADABRA
F1 Score Divider

859

Figure 2. DANCE lunge classifier outperforms existing methods of quantifying male aggression. (A) Lunge scores from 20 minute long videos from ground-truth (grey), DANCE lunge classifier (orange), CADABRA (purple) and Divider assay classifier (green). (B-E) Comparison of lunge scores from different levels of aggression. Male flies showing lunges ranging from 0 to 500 were compared manually, with a new DANCE lunge classifier, CADABRA, and Divider assay classifier. (B) 'low aggressive', (n=10; **P<0.0017, ****P<0.0001). (C) 71–160 lunges, 'moderately aggressive', (n=11; **P<0.0102, ***P<0.0002, ***P<0.0001). (D) 161–300 lunges, 'highly aggressive', (n = 11; **P<0.0057, ***P<0.0002, ***P<0.0004). and (E) >300 lunges, 'hyper-aggressive', (n = 8; *P<0.0402, **P<0.0029, ***P<0.0006; Friedman's ANOVA with Dunn's test) (F) Regression analysis of DANCE 'lunge classifier' *vs.* manual scores (R^2=0.9893, n=40). (G) CADABRA *vs.* DANCE lunge classifier (R^2=0.9, n=40). (H) Divider assay lunge classifier *vs.* manual scores (R^2=0.7739, n=40). (I) F1 score, precision, and recall between DANCE lunge classifier, CADABRA, and Divider assay classifier.

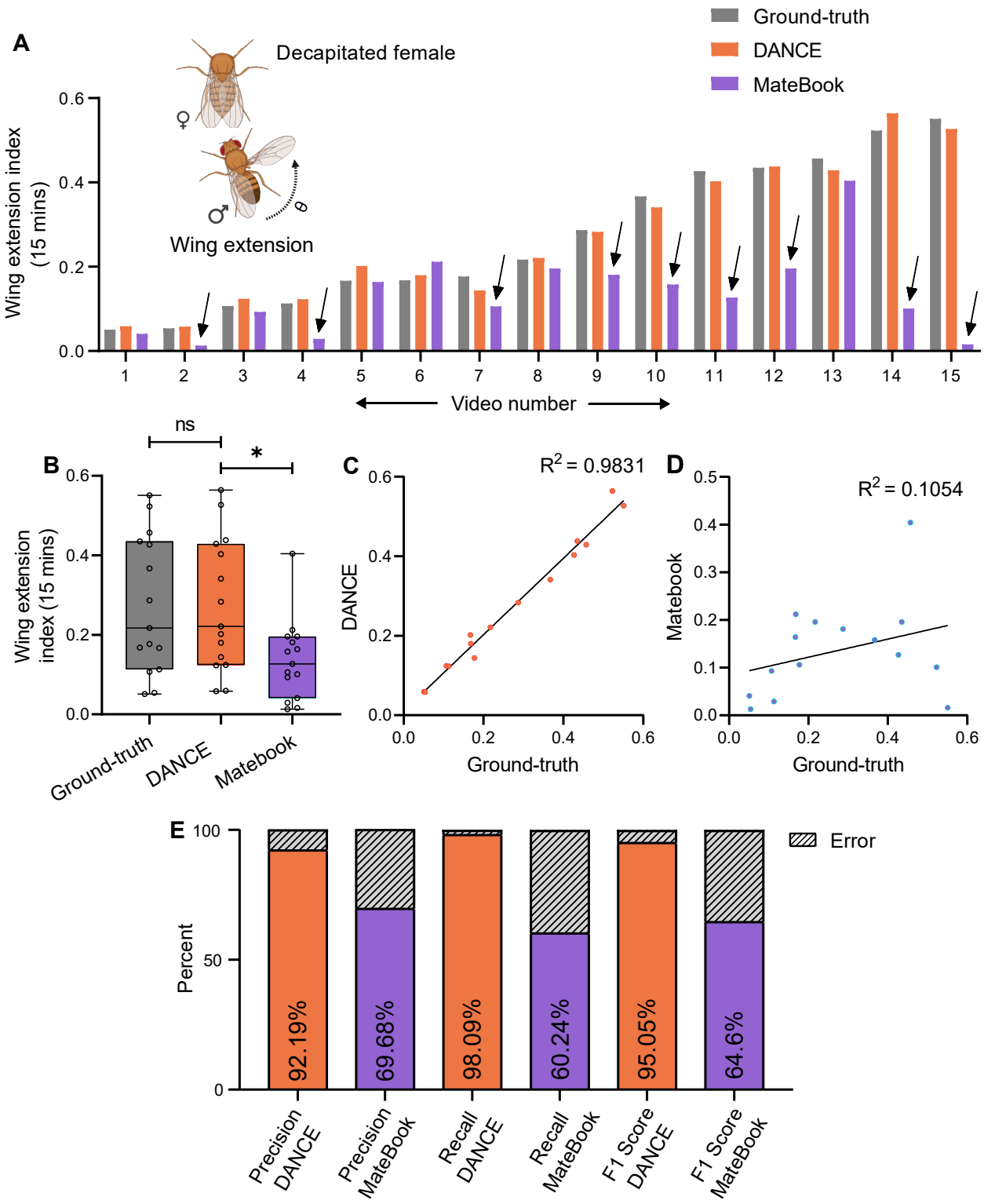


Figure 3. DANCE wing extension classifier outperforms existing methods of quantification. (A) Wing extension index of males from 15 minutes long videos from the ground-truth (grey), DANCE wing extension classifier (orange) and MateBook (purple), against decapitated virgin females. MateBook underscored wing extension across multiple videos (black arrows). (B) Comparison of ground-truth *vs.* DANCE *vs.* MateBook wing extension classifier (Kruskal-Wallis ANOVA with Dunn's test, ns, p>0.9999, *p=0.0436; n=15). (C) Regression analysis of the DANCE wing extension classifier *vs.* ground-truth

(R^2=0.9831, n=15). (D) MateBook *vs.* ground-truth (R^2=0.1054, n=15). (E) F1 score, precision, and recall of DANCE wing extension classifier and MateBook against ground-truth scores.

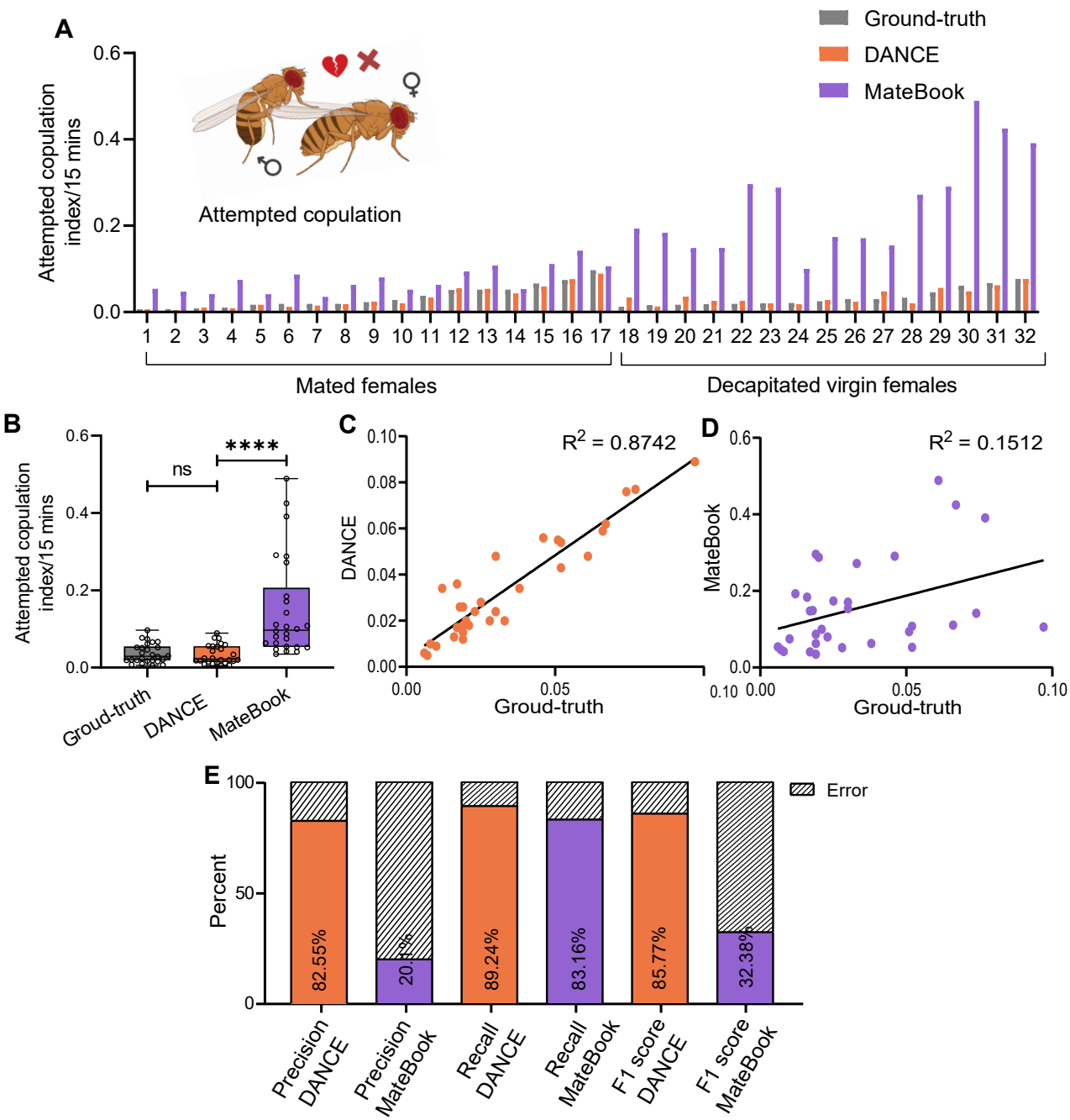


Fig. 4. DANCE attempted-copulation classifier outperforms existing methods of quantification. (A) Attempted copulation index of males from 15 minute long videos from the ground-truth (grey), 'DANCE attempted copulation classifier' (orange) and MateBook (purple) against both mated and decapitated females. (B) Comparison of ground-truth *vs.* DANCE attempted-copulation classifier *vs.* MateBook (Kruskal-Wallis ANOVA with Dunn's test, ns, p>0.9999; ****p<0.0001, n=32). (C) Regression analysis of the attempted-

copulation classifier *vs.* ground-truth (R^2=0.8742, n=32). (D) Regression analysis of MateBook *vs.* ground-truth (R^2=1512, n=32). (E) F1 score, precision, and recall of DANCE and MateBook attempted-copulation classifiers against ground-truth scores.

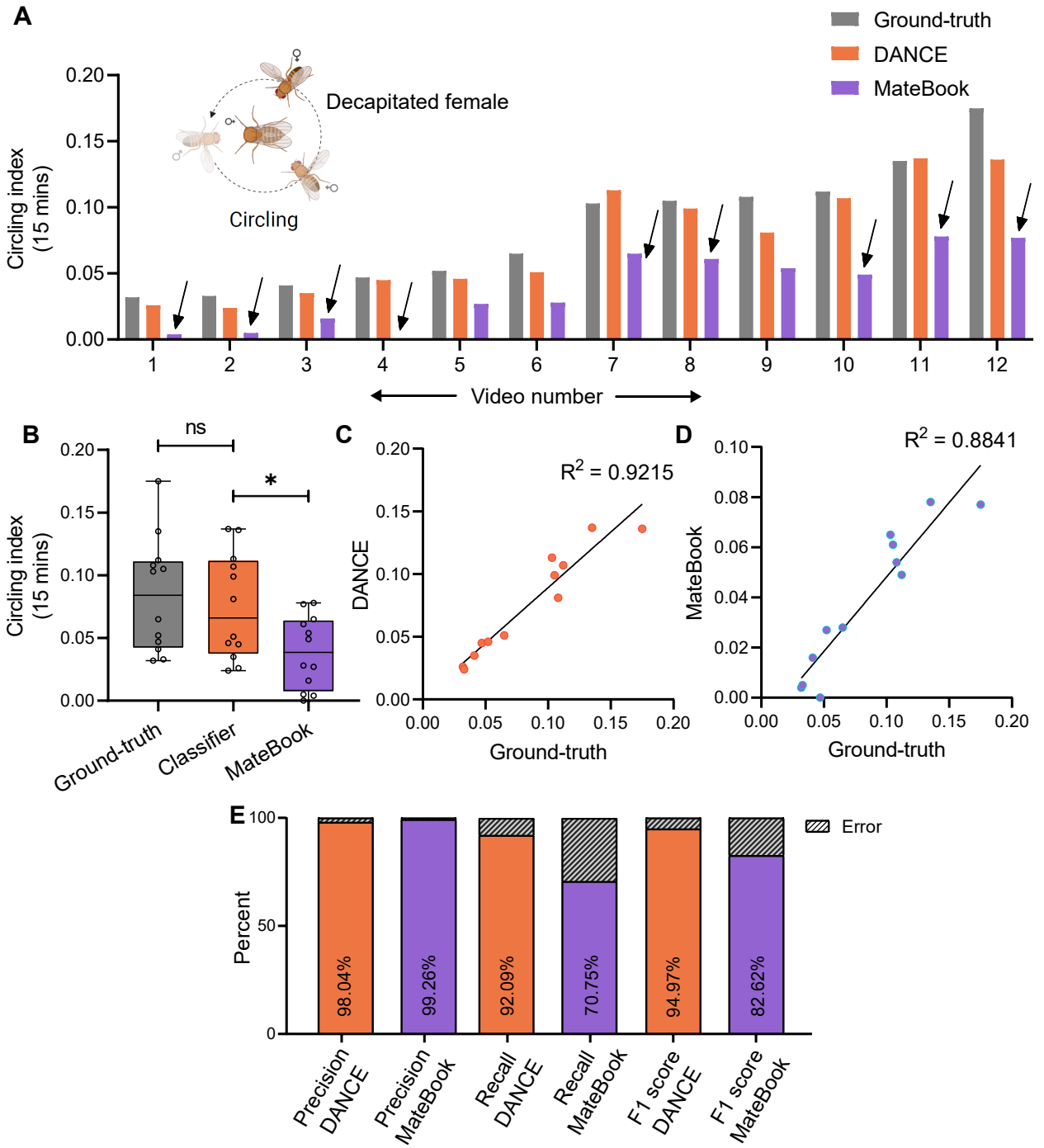


Fig. 5. DANCE circling classifier outperforms existing methods of quantification. (A) Circling index of males from 15 minutes long videos from the ground-truth (grey), 'DANCE circling classifier' (orange) and MateBook (purple) against decapitated virgin females. (B) Comparison of manual *vs.* DANCE *vs.* MateBook circling classifier (Ordinary one-way ANOVA with Dunnett's test, ns, p=0.8014; *p=0.0157, n=12). (C) Regression analysis of the circling classifier *vs.* ground-truth (R^2=0.92, n=12). (D) MateBook *vs.* ground-truth (R^2=0.88, n=12). (E) F1 score, precision, and recall of DANCE and MateBook circling classifiers against ground-truth scores.

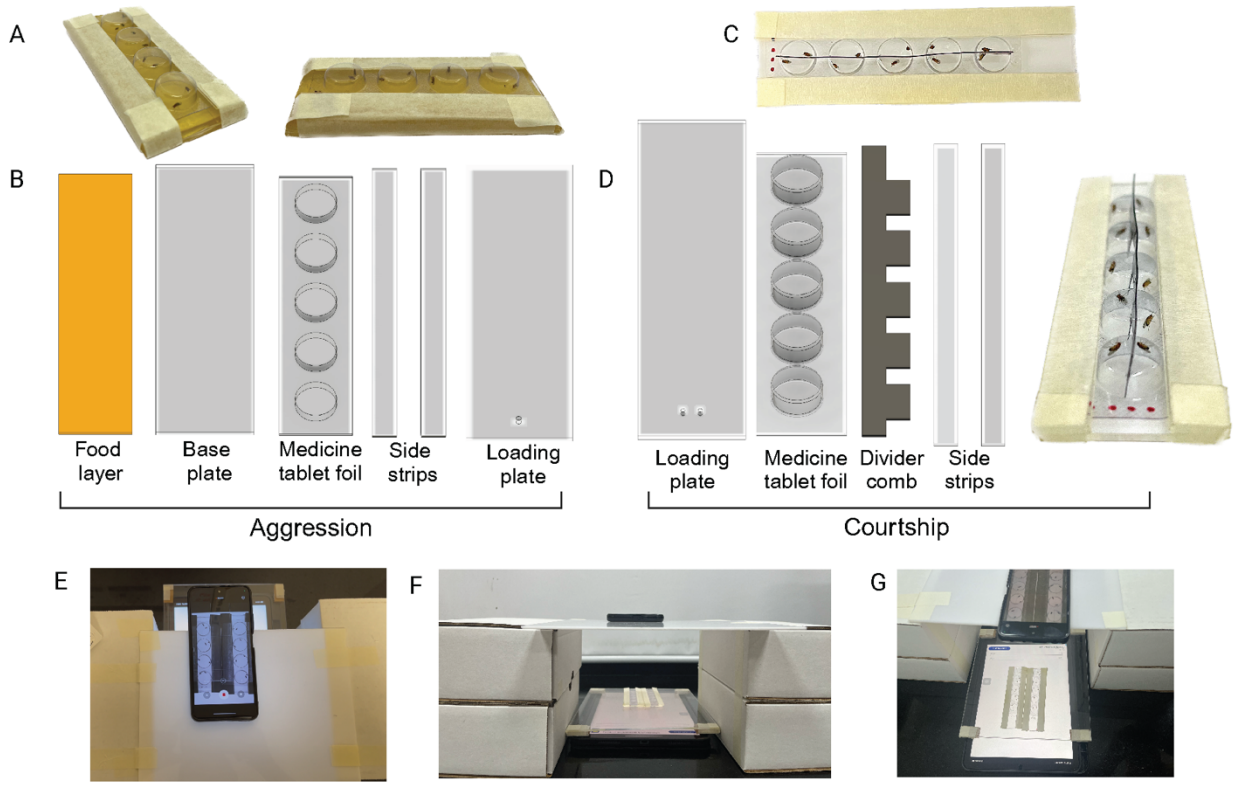
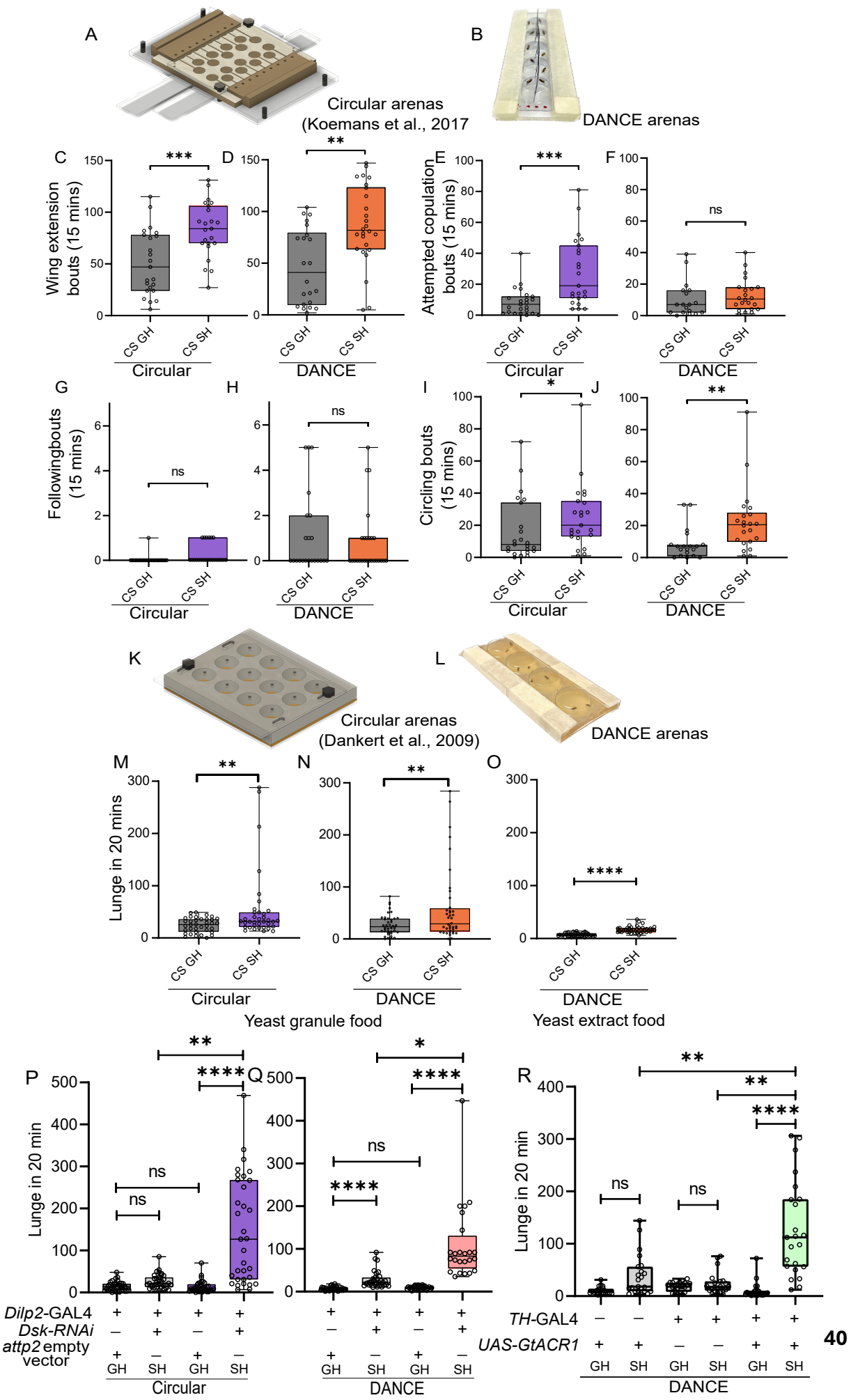


Fig. 6. DANCE hardware and recordings setup. (A) DANCE aggression set up (B) 3D-rendered components of the aggression set up (C) DANCE courtship set up (D) 3D-rendered components of the courtship set up, male and female are separated on either side using X-ray film separator or 'divider comb'. (E-G) Top and side views of the DANCE setup with a smartphone camera for recording and electronic tablet being used as a backlight.


A
Circular arenas
(Koemans et al., 2017
B
DANCE arenas
C
D
E
F
Wing extension bouts (15 mins)
Attempted copulation bouts (15 mins)
CS GH
CS SH
Circular
DANCE
G
H
I
J
Followingbouts (15 mins)
Circling bouts (15 mins)
K
Circular arenas
(Dankert et al., 2009)
L
DANCE arenas
M
N
O
Lunge in 20 mins
Yeast granule food
Yeast extract food
P
Q
R
Lunge in 20 min
Dilp2-GAL4
Dsk-RNAi
attp2 empty vector
TH-GAL4
UAS-GtACR1
GH
SH

917

Fig. 7. Benchmarking DANCE hardware and testing various neurogenetic tools. (A-B) Courtship behaviors recorded using a pre-existing (circular) and DANCE set up, from group-housed (GH) and single-housed (SH) flies, for (C-D) wing extension, (C) ***$p<0.0010$, n=23; (D) **$p<0.0013$, GH, n=22 and SH, n=26. (E-F) Attempted copulation, (E) ***$p<0.0002$, n=23; (F) ns, $p<0.1907$, GH, n=18 and SH, n=22. (G-H) Following, (G) ns, $p>0.0959$, n=23; (H) ns, $p<0.6589$, GH, n=22 and SH, n=26. (I-J) Circling, (I) *$p<0.012$, n=23; (J) **$p<0.0021$, GH, n=19 and SH, n=22. (K-L) Aggressive lunges recorded using a pre-existing (circular) and DANCE set up. (M-N) Lunges in SH flies compared to GH flies reared on food with yeast granules, (M) **$p<0.0138$, n=36; (N) **$p<0.0372$, n=40. (O) Effect of yeast extract food on aggressive behavior; ****$p<0.0001$, n=38. (P-Q) Genetic knockdown of the neuropeptide Drosulfakinin (Dsk) in insulin-producing neurons using *dilp2*-GAL4. (P) ns, $p<0.0502$, ns, $p>0.9999$, ****$p<0.0001$, **$p<0.0040$, n=35. (Q) ****$p<0.0001$, ns, $p>0.9999$, ****$p<0.0001$, *$p>0.0210$, n=30. (R) Optogenetic silencing of dopaminergic neurons by *UAS-GtACR1* driven by *TH*-GAL4 driver, ns, $p<0.0986$, ns, $p>0.9999$, ****$p>0.0001$, **$p>0.0012$, **$p>0.0013$, n=24. (C-J and M-O) Mann-Whitney U test; (P-R) Kruskal-Wallis test with Dunn's multiple comparisons.

Supplementary information

Supplementary Figures and Legends

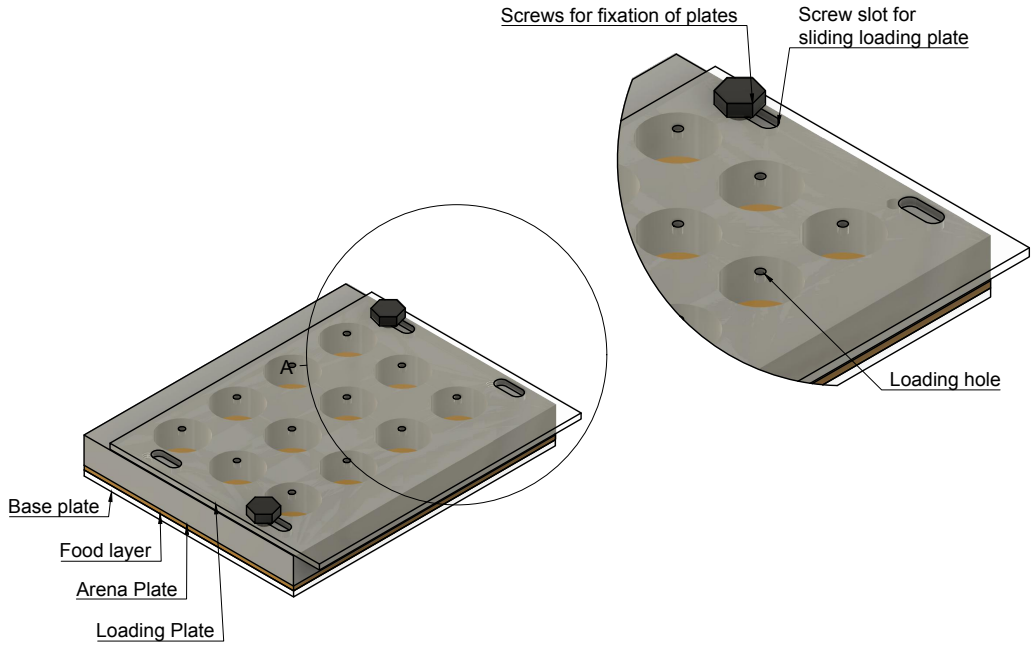


Figure 1— Figure Supplement 1. Aggression chamber described by Dankert *et al.*, 2009. It consists of a bottom food plate, 12 wells (aggression arenas), a top plate with fly loading holes and a screw slot for sliding the loading plate.

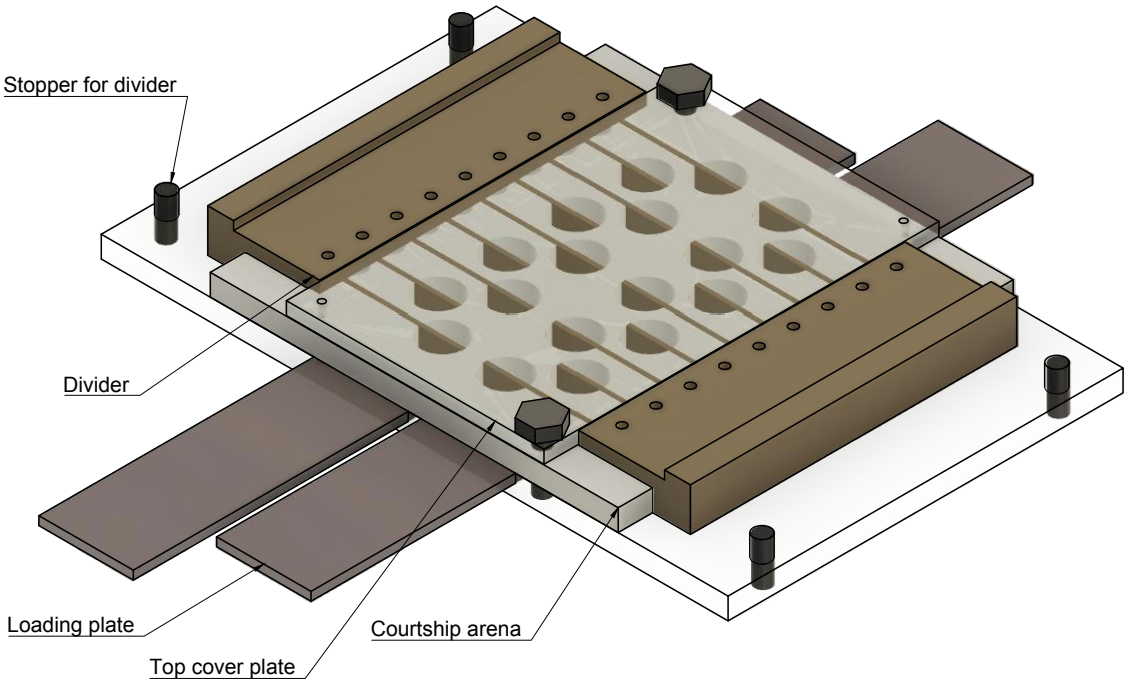


Figure 1— Figure Supplement 2. Courtship set-up described by Koemans *et al.*, 2017). It consists of 18 wells (courtship arenas), a top cover plate, a sliding loading plate, and a sliding divider assembly to separate male and female flies.

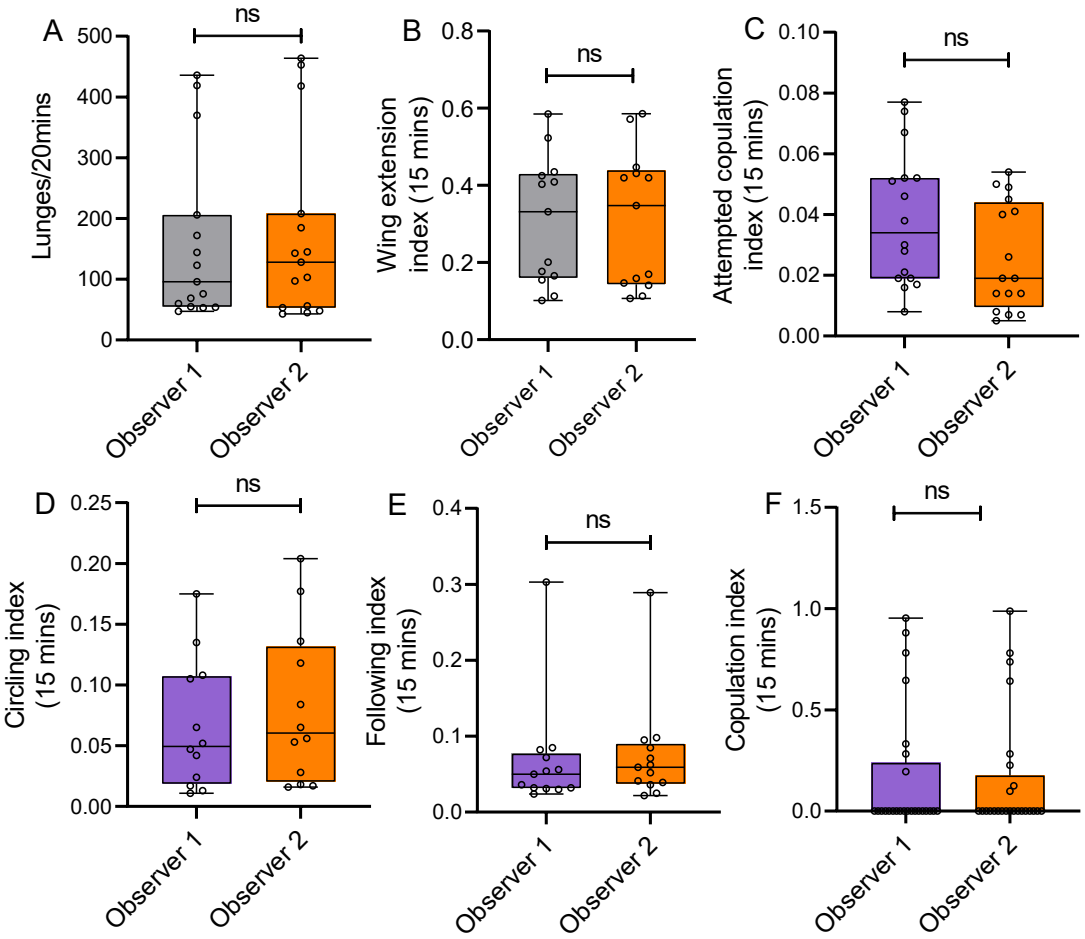


Figure 2— Figure Supplement 1. Comparison of annotations by two independent evaluators to assess observer bias during classifier ground-truthing. (A) Aggressive lunge, p=0.8789, n=15. (B) Courtship including wing extension, p=0.9999, n=13. (C) Attempted copulation, p=0.0571, n=16. (D) Circling, p=0.4343, n=12. (E) Following, p=0.4405, n=13. (F) Copulation, p=0.9221, n=25. (A-F) Mann-Whitney U test.

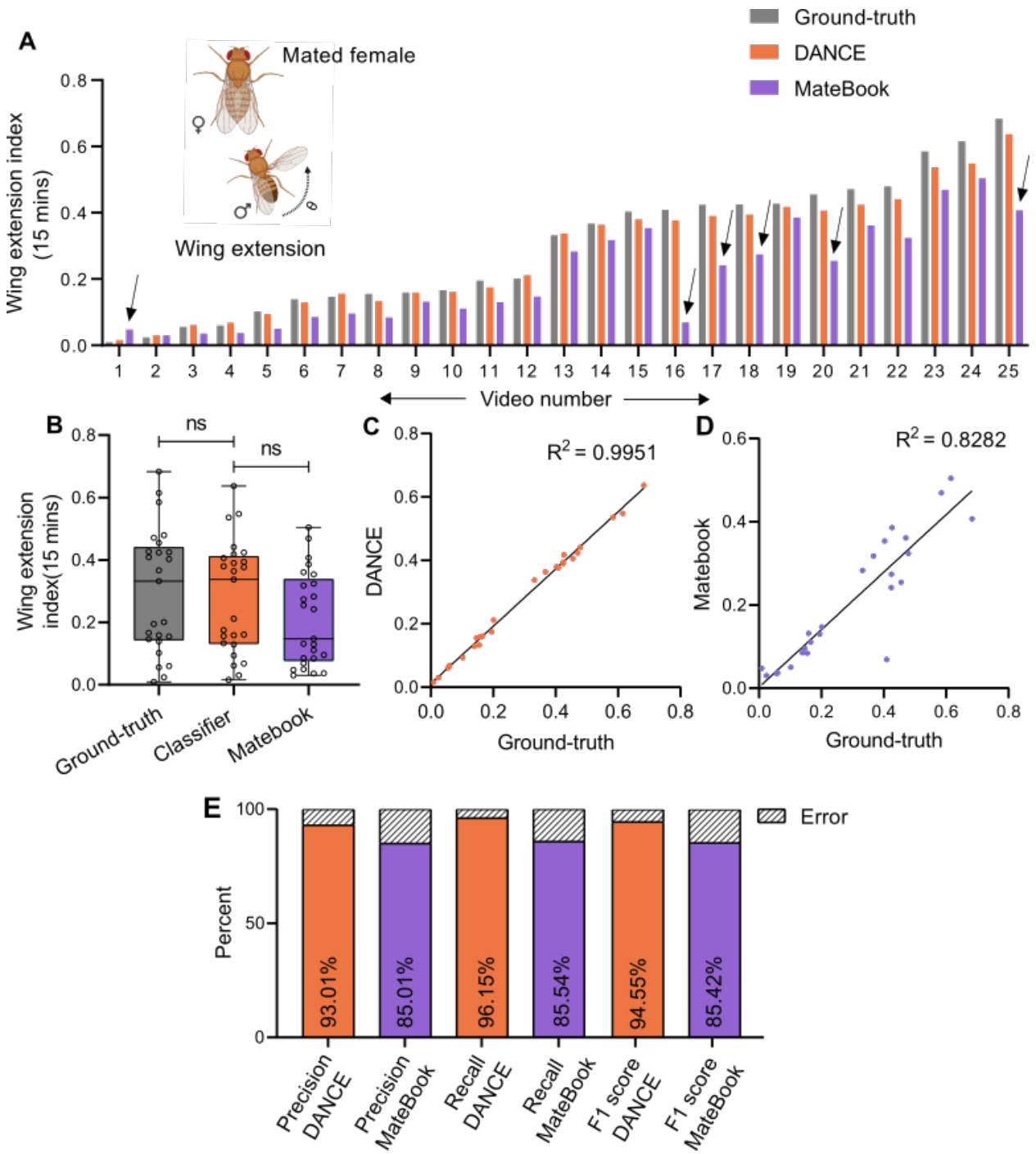


Figure 3— Figure Supplement 1. DANCE wing extension classifier outperforms existing quantification methods in videos with mated females. (A) Wing extension of males from 15-minute-long videos from the ground-truth (grey), the DANCE classifier (orange) and MateBook (purple) against mated females MateBook underscored wing extension across multiple videos (black arrows) (Friedman's ANOVA with Dunn's test, p=0.3582; p<0.0001, n=25). (B) Comparison of ground-truth *vs.* DANCE *vs.* MateBook wing extension classifier (Kruskal-Wallis ANOVA with Dunn's test, p>0.9999; p=0.1039, n=25). (C) Regression analysis of the DANCE classifier *vs.* ground-truth (R^2=0.9951, n=25). (D) MateBook *vs.* ground-truth (R^2=0.8282, n=25). (E) F1 score, precision, and recall of DANCE classifier and MateBook against ground-truth scores.

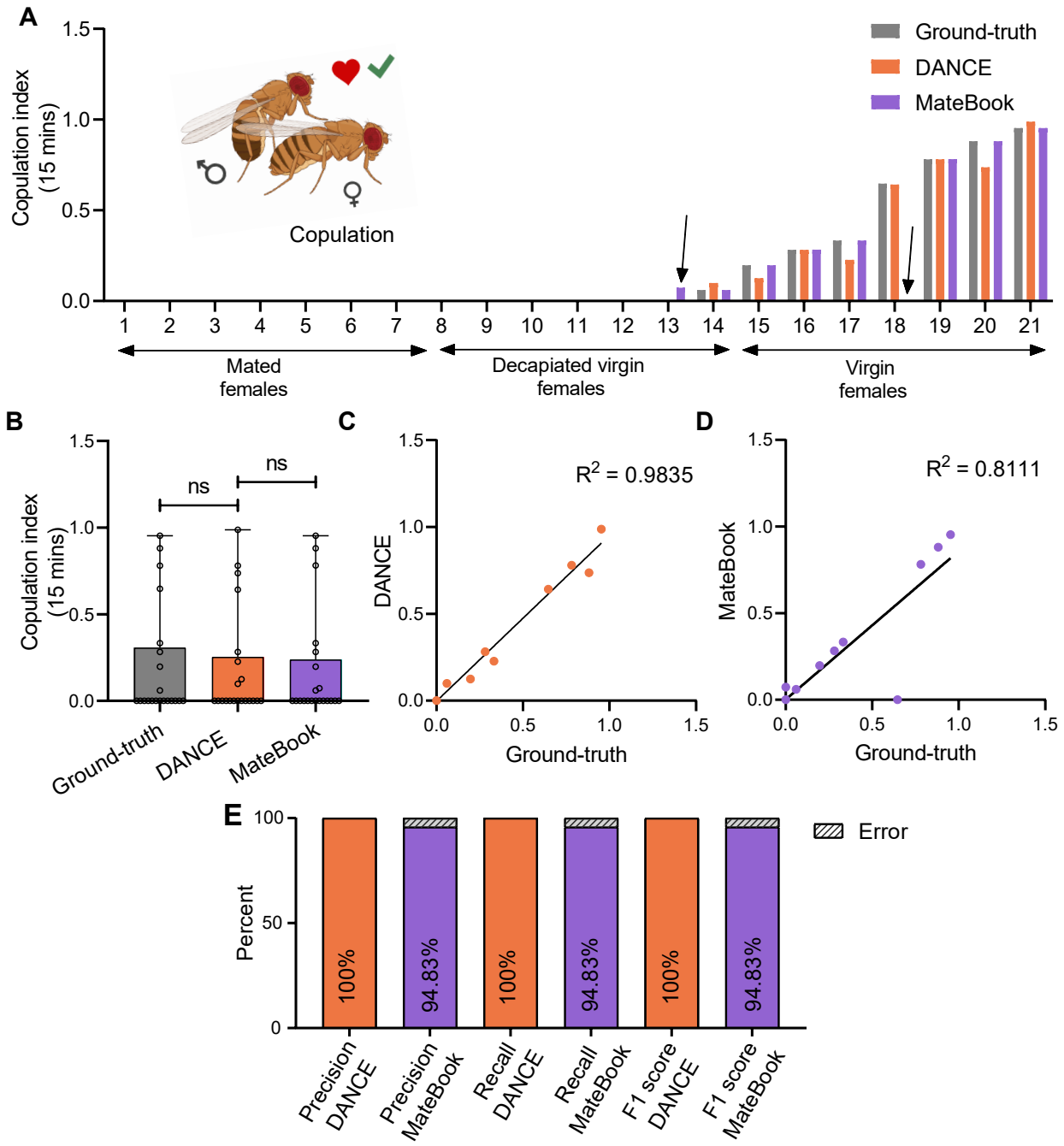


Figure 4— Figure Supplement 1. DANCE copulation classifier evaluation in mixed female dataset. (A) Quantification of copulation in 15 minutes from individual videos showing scores from the manual method (grey), DANCE copulation classifier (orange), and MateBook (purple) (Friedman ANOVA with Dunn's test, p>0.9999; p>0.9999, n=21). (B) Box plot comparison of manual *vs.* DANCE copulation classifier *vs.* MateBook (Kruskal-Wallis ANOVA with Dunn's test, p>0.9999; p>0.9999, n=21). (C) Bar plots showing F1 score, precision, and recall of DANCE copulation classifier and MateBook against ground-truth scores. (D) Regression analysis of the copulation classifier *vs.* manual scores (R^2=0.98, n=21). (E) MateBook *vs.* manual scores (R^2=0.81, n=21).

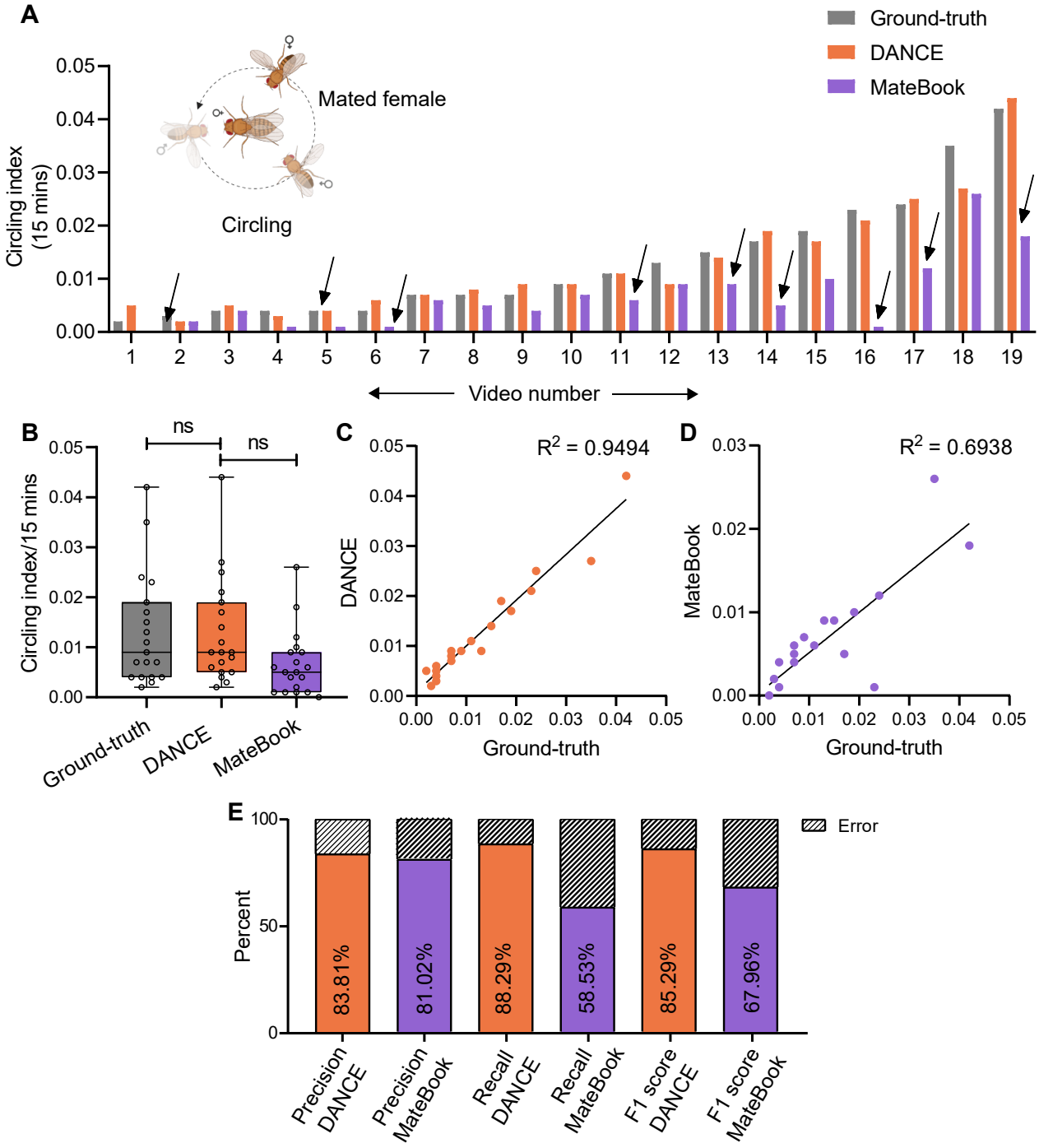


Figure 5— Figure Supplement 1. DANCE circling classifier evaluation in the mated female dataset. (A) Circling index of males from 15-minute videos from the ground-truth (grey), 'DANCE circling classifier' (orange) and MateBook (purple) against mated female dataset (Friedman's ANOVA with Dunn's test, p>0.9999; p <0.0001, n=19). (B) Comparison of manual *vs.* DANCE *vs.* MateBook circling classifier (Kruskal-Wallis ANOVA, p>0.9999; p=0.0822, n=19). (C) Regression analysis of the circling DANCE classifier *vs.* ground-truth (R2 = 0.9494, n = 19). (D) MateBook *vs.* ground-truth (R^2=0.6938, n=19). (E) F1 score, precision, and recall of DANCE and MateBook circling classifiers against ground-truth scores.

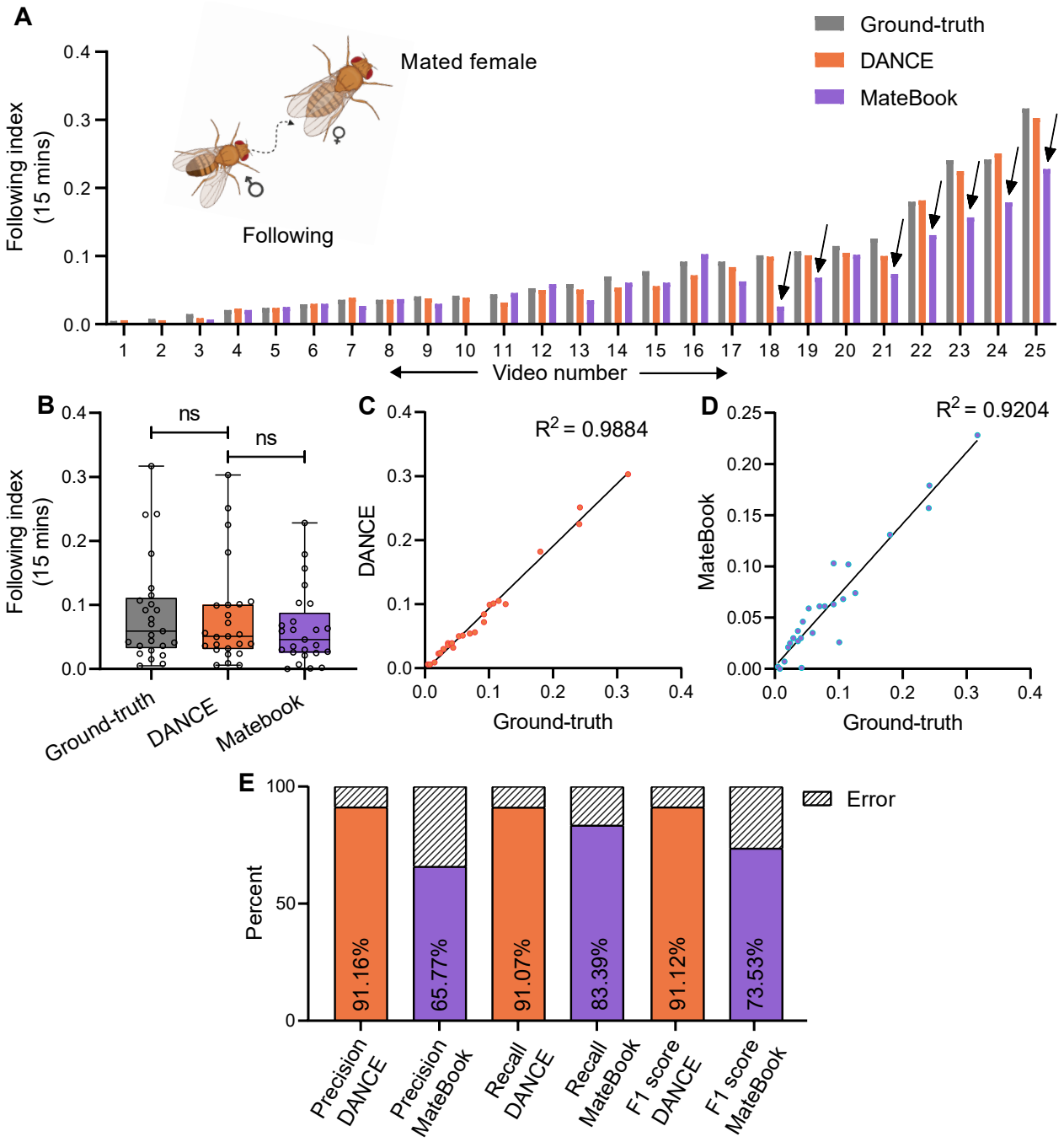


Figure 5— Figure Supplement 2. DANCE following classifier evaluation in the mated female dataset. (A) Following index of males from 15 minutes long videos from the ground-truth (grey), 'DANCE following classifier' (orange) and MateBook (purple) against mated females (Friedman's ANOVA with Dunn's test, p=0.1794; p=0.0029, n=25). (B) Box plot comparison of manual *vs.* DANCE following classifier *vs.* MateBook (Kruskal-Wallis ANOVA with Dunn's test, p>0.9999; p=0.5287, n=25). (C) Regression analysis of the following classifier *vs.* ground-truth (R^2=0.9894, n=25). (D) MateBook *vs.* ground-truth (R^2=0.9204, n=25). (E) F1 score, precision, and recall of DANCE and MateBook following classifiers against ground-truth scores.

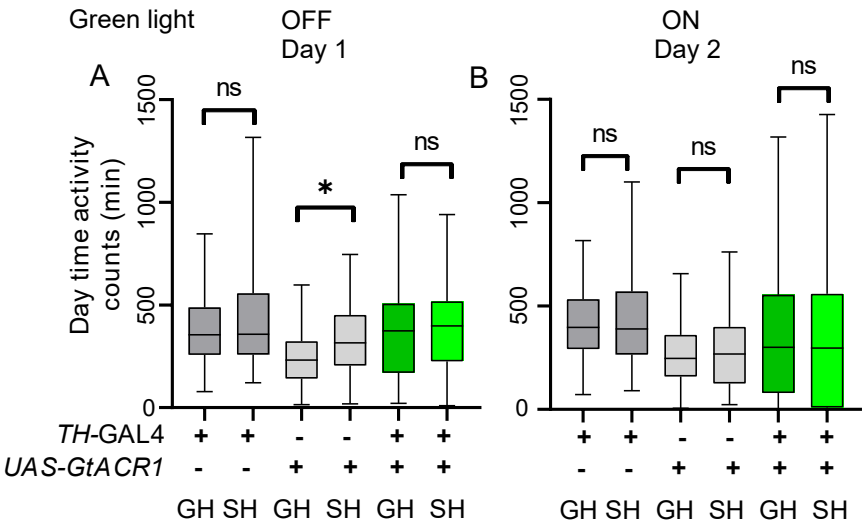


Figure 7— Figure Supplement 1. Effect of optogenetic silencing of dopaminergic neurons on *Drosophila* activity. (A, B) Transient silencing of dopaminergic neurons using *UAS-GtACR1* driver didn't affect daytime activity between SH and GH flies. (A) Without silencing mediated by green light on day 1; *TH*-GAL4, ns, p=0.7383, GH: n=43, SH: n=43; *UAS-GtACR1*, ns, p=0.4812, GH: n=51, SH: n=42; T*H*-GAL4>*UAS-GtACR1*, ns, p=0.9942, GH: n=54, SH: n=51. (B) With silencing mediated by green light on day 2; *TH*-GAL4, ns, p=0.9976, GH: n=43, SH: n=43; *UAS-GtACR1*, ns, p=0.9779, GH: n=51, SH: n=42; T*H*-GAL4>*UAS-GtACR1*, ns, p=0.9974, GH: n=54, SH: n=51. One-way ANOVA with Tukey's multiple comparisons test for within each day comparison. Two-way ANOVA for comparison across days; interaction, ns, p=0.5504, silencing, ns, p=0.5172, housing, ns, p=0.1602, GH: n=43, SH: n=41; *UAS-GtACR1*, interaction, ns, p=0.4533, silencing, ns, p=0.3602, housing, *p=0.255, GH: n=51, SH: n=42; T*H*-GAL4>*UAS-GtACR1*, interaction, ns, p=0.9977, silencing, ns, p=0.2454, housing, ns, p=0.5868, GH: n=54, SH: n=51.