Supplementary Information: A machine-learning enhanced data fusion pipeline identifies new G-protein based regulators of *Drosophila* wing growth and morphogenesis

Nilay Kumar[1,†], Maria F. Unger[1,†], Francisco J. Huizar[1,2,†], Vijay Velagala[1], Qinfeng Wu[1], Pavel A. Brodskiy[1] and Jeremiah J. Zartman[1,2,3]

[1]Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[2]Department of Bioengineering, University of Notre Dame, Notre Dame, IN, 46556, USA
[3]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, 46556, USA
[†]These authors contributed equally to this work.
∗Corresponding author: Department of Chemical and Biomolecular Engineering, University of Notre Dame, 100 McCourtney Hall, jzartman@nd.edu

Table of contents

Supplementary Tables **2-3**

1. Supplementary Table 1. Off-target analysis of RNAi lines generating severe wing phenotypes.
2. Supplementary Table 2. Comparison of hits across multiple studies.

Supplementary Figures **4-12**

1. Supplementary Figure 1. A deep-learning-based pipeline fully automates the identification of severe wing phenotypes in the loss-of-function GPCR screening dataset.
2. Supplementary Figure 2. Training and validation across all epochs
3. Supplementary Figure 3. Confusion matrix and intervein label schematics
4. Supplementary Figure 4. Knockdown of selected GPCRs and G proteins causes severe morphological defects in the adult Drosophila wing.
5. Supplementary Figure 5. Multiple Gal4 drivers produce similar phenotypes upon *CCHa1-R* knockdown.
6. Supplementary Figure 6. Evaluation of trichome density associated with severe wing phenotypes shown in Figure 3.
7. Supplementary Figure 7. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) confirms expression of positive hits.
8. Supplementary Figure 8. Quantification of adult wing area in the GPCR and G-protein RNAi screen.
9. Supplementary Figure 9. Quantification of anterior and posterior crossvein defects in the GPCR and G-protein RNAi screen.
10. Supplementary Figure 10. Evaluation of trichome density associated with severe wing phenotypes shown in Figure 4.

MAPPER 2.0 documentation **13-18**

Supplementary Tables:

Supplementary Table 1. Off-target analysis of RNAi lines generating severe wing phenotypes.

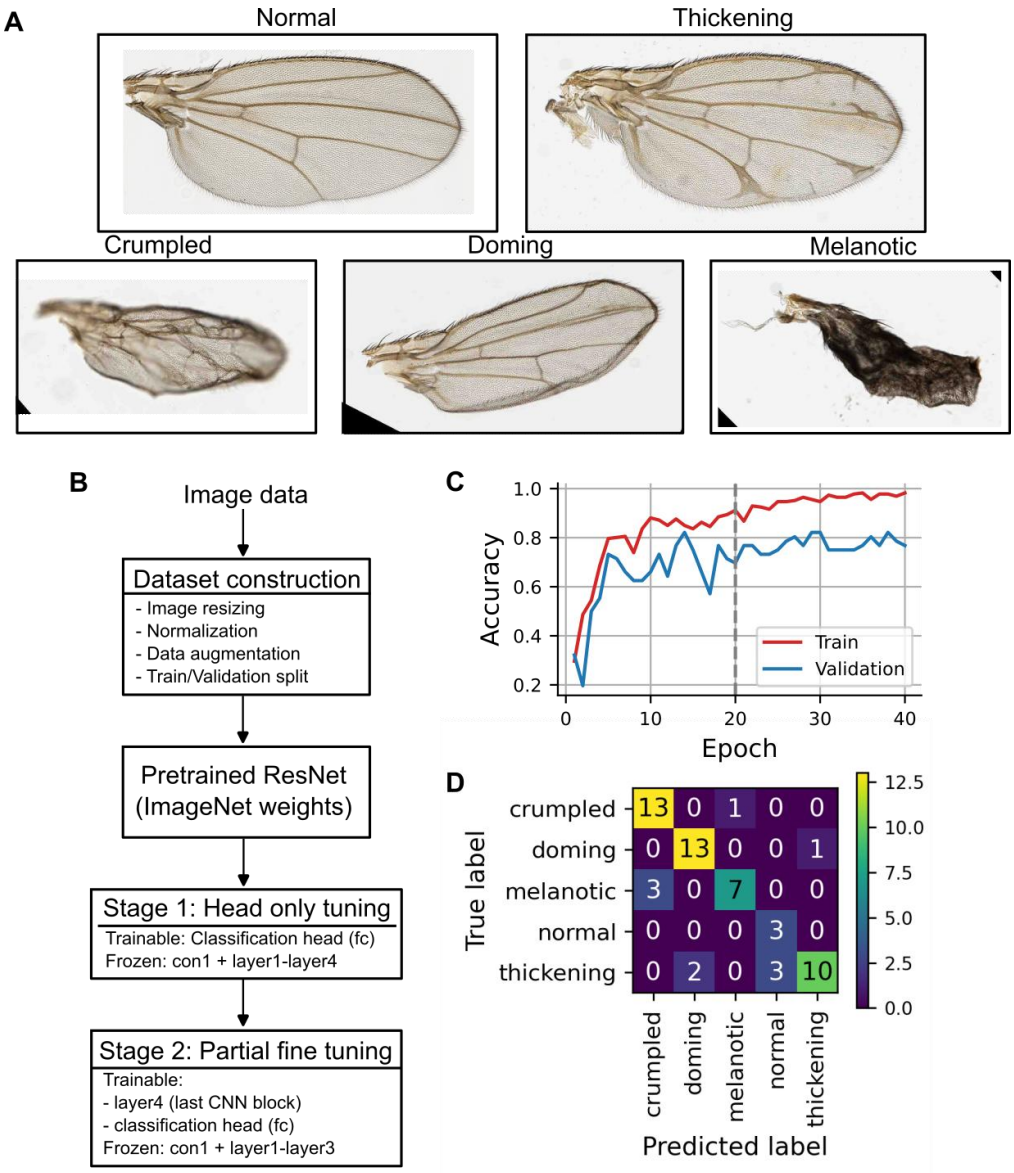
Gene symbol	dsRNA sequence	Off -target analysis
5-HT1B (* CG15113) (BL #54006)	CGGCAGCAACACGAACGACAA	Off-target analysis of the designed dsRNA targeting 5-HT1B (CG15113) using dsCheck revealed three perfect 21-nt matches (0 mismatches) to CG30126 (CG30126-PA), indicating potential unintended silencing of this gene. No perfect or near-perfect matches were detected for the intended target CG15113, suggesting ineffective on-target recognition. Additionally, multiple transcripts, including mam (CG8118) and CG9932, showed repeated 2-mismatch alignments (3 hits each), indicating a low but notable risk of off-target effects. These findings suggest that this dsRNA may fail to target CG15113 effectively while potentially silencing unrelated transcripts and should be redesigned or experimentally validated.
CCHa1-R (*CG30106) (BL #51168)	GCCGATCTGCTGGTTATTATA	An off-target analysis of the designed dsRNA (21 nt in length) targeting CCHa1-R (CG30106) was performed using dsCheck. The analysis revealed three perfect 21-nt matches (0 mismatches) to the target gene CG30106-PA, confirming robust and specific targeting. No additional transcripts were identified with 1 or 2 mismatches, indicating a high degree of sequence specificity and minimal risk of off-target silencing in other *Drosophila melanogaster* genes.
CG13579 (BL #51053)	CTAGTCAGTGCGGTTAGCAAA	No off targets detected
CG 30340 (BL #28652)	ATGTGATAGCCACCAACCGTTCTTT GCGCTCGCCCACCAACTTGATTATT GCAAATATGGCCGTTGCGGATCTAC TGACCCTGGCCATTTGTCCTGCCAT GTTCATGGTGAACGACTTCTATCAG AACTACCAGTTGGGATGCGTGGGC TGCAAGCTGGAGGGCTTTTTGGTG GTGGTCTTCCTCATCACAGCCGTGC TCAATCTCTCGGTAGTCAGCTACGA TCGTCTCACGGCCATCGTGCTGCC AATGGAGACGCGCCTCACCATAAG GGGCGTCCAGATAGTGGTGGTCTG CACCTGGGTCTCGGGTATCCTATTG GCATCACCATTGGCTTTCTATCGCT CTTACAGGGTGCGAGTGTGGAAGA ACTTTACGGAGCGATATTGCAAGGA GAACACCTCTGTGCTGCCCAAGTAC TG	Off-target analysis of the designed dsRNA (length: 421 nt) using dsCheck revealed multiple 21-nt segments with partial complementarity (0–2 mismatches) to other Drosophila melanogaster transcripts. Notably, the dsRNA exhibited 403 perfect 21-nt matches to the intended target gene CG30340. Potential off-target matches with up to 2 mismatches were found in other transcripts, including multiple isoforms of CG3304 (PA, PB, PC), which shared as many as 5 mismatched 21-mers with 1 mismatch and 2 with 2 mismatches. Additional transcripts showing moderate off-target potential included CG13724, CG11191, CG11949 (cora), CG6671 (AGO1), and CG7769 (DDB1), each of which shared multiple 21-mers with 1–2 mismatches.
mtt (*CG30361) (BL #44076)	AAGGAGCAGCAGCGACAACAA	Off-target analysis of the designed dsRNA (length: X nt) using dsCheck revealed multiple 21-nt segments with partial complementarity (0–2 mismatches) to other *Drosophila melanogaster* transcripts. Notably, the dsRNA exhibited perfect 21-nt matches to three isoforms of the target gene (CG30361) and potential off-target matches with up to 2 mismatches in other genes such as CG32529, Vap-33-1, and kel,

Supplementary Table 2. Comparison of hits across multiple studies.

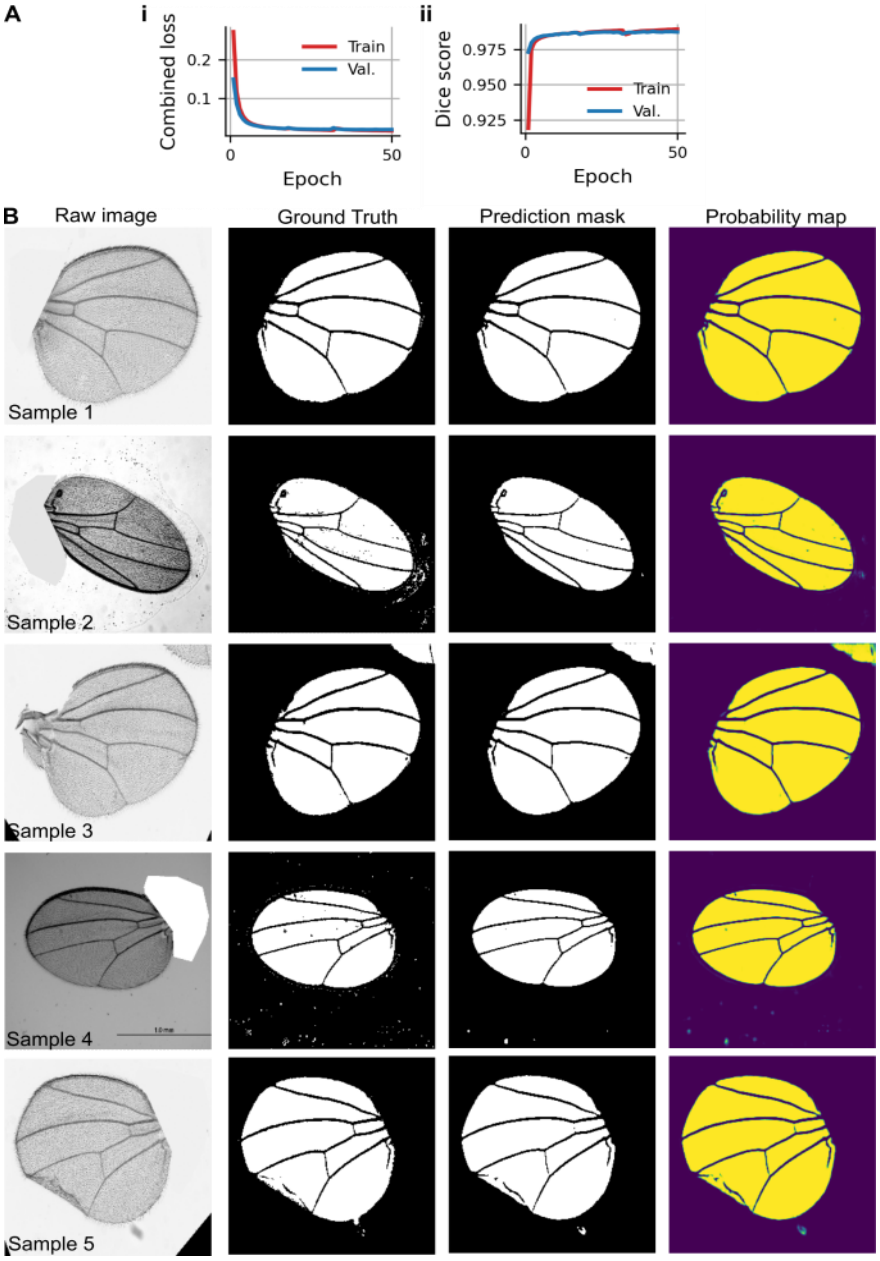
	Gene Symbol	Saad and Hipfner[a]		This study	
1	GABA-B-R2	V1785	increased Hh signaling	P{TRiP.HMC02975}attP2	Small wings; Pigmentation phenotype (Figure 3)
		TRiP.JFO2779	increased Hh signaling		
		V110268	increased Hh signaling	P{TRiP.JF02779}attP2	WT
		TRiP.HMC02975	WT		
2.A	NPFR76F	Not tested		P{TRiP.JF02657}attP2	Small wings (Figure 3)
2.B	NPFR1	Not tested		P{TRiP.JF01959}attP2	Small wings (Figure 3)
3.A	AstA-R1	Not tested		P{TRiP.JF02578}attP2	Small wings (Figure 3)
3.B	AstA-R2	Not tested		P{TRiP.JF01955}attP2	Small wings (Figure 3)
3.C	AstC-R2	Not tested		P{TRiP.GL01063}attP2	Small wings (Figure 3)
4.A	mthl9	V108967	strong inflation defects, decreased Hh signaling	P{TRiP.HMJ24136}attP40	ACV defects (Figure 4)
		V2769	WT		
		V2770	WT		
				P{TRiP.HMC03141}attP40	ACV, PCV, and L5 end defects (Figure 4)
		TRiP.HMC03141	WT		
		TRiP.HMJ24136	WT		
4.B	mthl10	V100829	WT	P{TRiP.HMC03304}attP2	ACV/PCV defects (Figure 4)
		V51425	WT		
4.C	mthl14	Not tested		P{TRiP.HMC06446}attP2	ACV defects (Figure 4)
5.A	fz3	Not tested		P{TRiP.GLC01626}attP2	ACV defects (Figure 4)
5.B	fz4	Not tested		P{TRiP.HMC05864}attP40	ACV defects (Figure 4)
6	CG4313	V107434	WT	P{TRiP.HMS00755}attP2	ACV/PCV defects (Figure 4)
7	Dh44-R2	Not tested		P{TRiP.JF03289}attP2	ACV/PCV defects (Figure 4)
8	ETHR	Not tested		P{TRiP.HMC03400}attP2	ACV/PCV defects (Figure 4)
9	CCHa1-R	Not tested		P{TRiP.HMJ21029}attP40	Vein thickening (Figure 2)
					WT
10	CG13579	Not tested		P{TRiP.HMJ21188}attP40	Crumpled wings (Figure 2)
					WT
11	CG30340	Not tested		P{TRiP.JF03067}attP2	Doming around the periphery (Figure 2)
					WT
12	Mtt	Not tested		P{TRiP.HMS02793}attP40	Crumpled/Melanized wings (Figure 2)
				P{TRiP.HMS00367}attP2	15/36 males showed bifurcation defects at L5. Partially formed PCV for a small percentage of adults.
13	Cirl	V100749	Increased ratio of L3-L4 w.r.t. to the wing area	P{TRiP.JF02674}attP2	No change in wing area reported. (Supplementary Figure 7). Bifurcation defects comparable to MS1096 Gal4 (Supplementary Figure 8)
14	mth	V102303	Increased ratio of L3-L4 w.r.t. to the wing area	P{TRiP.GL01055}attP2	Reduction in wing area (Supplementary Figure 7)
15	mthl5	V101593	Increased ratio of L3-L4 w.r.t. to the wing area	P{TRiP.HMC06002}attP40	No change in wing area reported. (Supplementary Figure 7).
16	stan	V107993	Decrease in wing area	P{TRiP.HMS01464}attP2	No change in wing area reported. (Supplementary Figure 7).
17	mthl4	V50752	Decrease in wing area	P{TRiP.HMC02422}attP2	No change in wing area or bifurcation defects (Supplementary Figure 7, 8)
18	rk	V105360	Decrease in wing area	P{TRiP.JF02678}attP2	No change in wing area reported (Supplementary Figure 7).
		V905			

[a] Saad F, Hipfner DR. Extensive crosstalk of G protein-coupled receptors with the Hedgehog signalling pathway. Development. 2021 Apr 1;148(7):dev189258

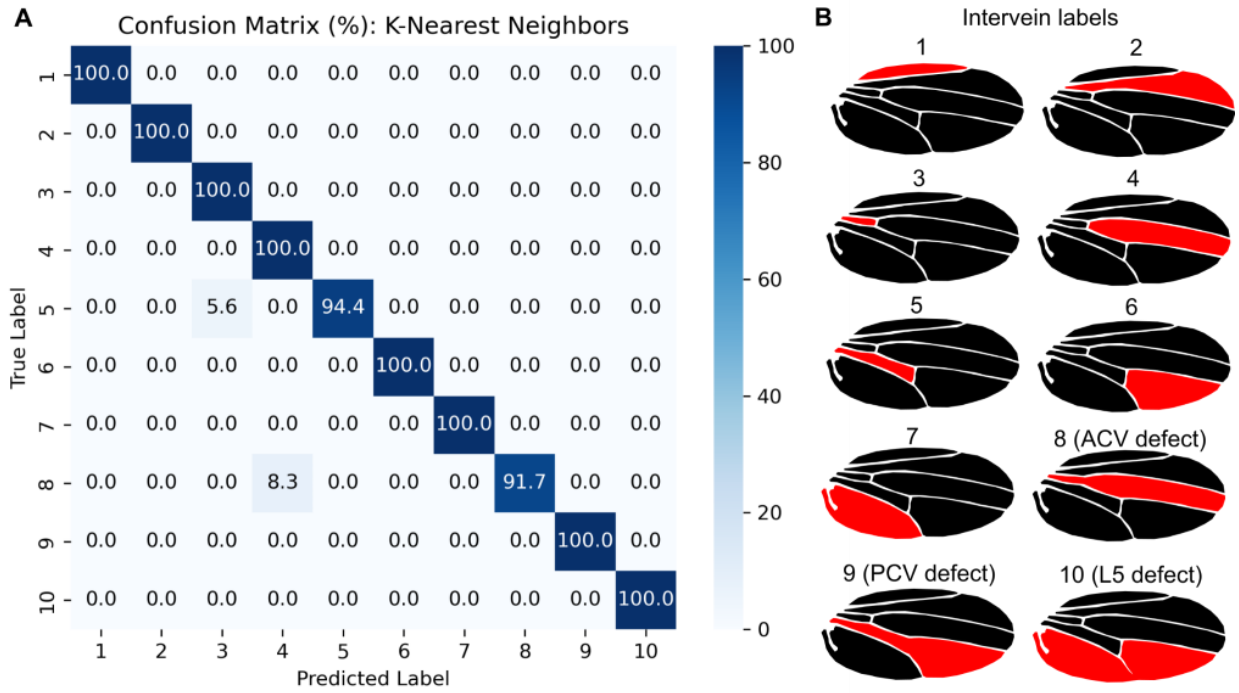
Supplementary Figures



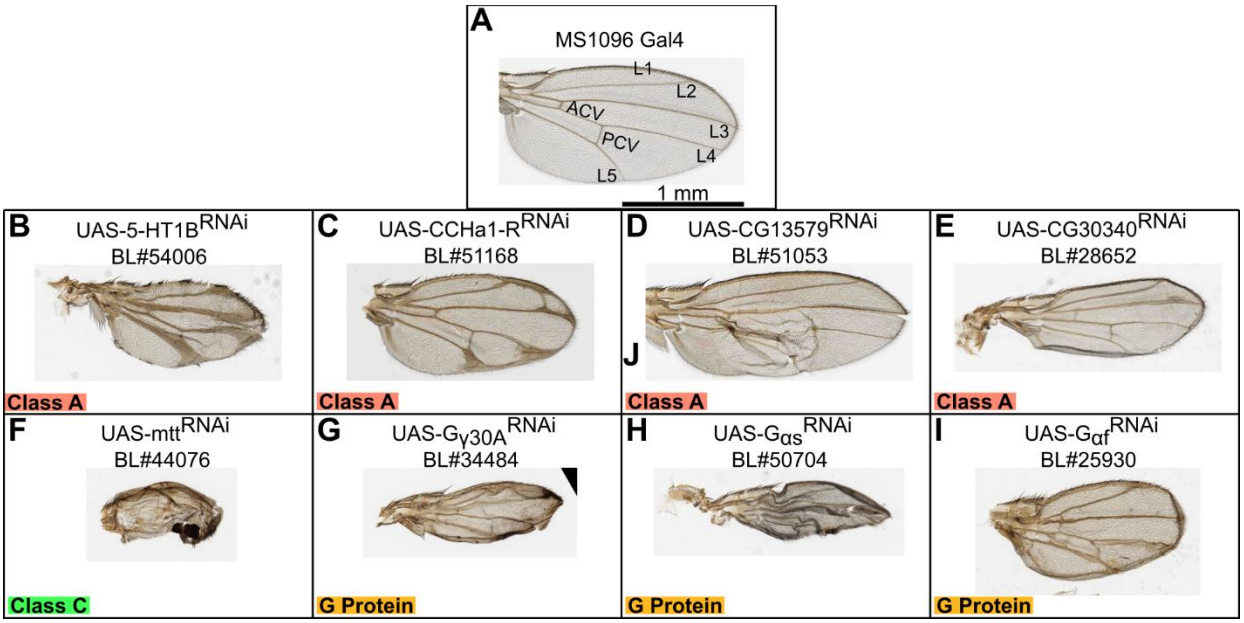
Supplementary Figure 1. A deep-learning-based pipeline fully automates the identification of severe wing phenotypes in the loss-of-function GPCR screening dataset. (A) Representative examples of the five wing phenotypic classes used in this study: Normal, Thickening, Crumpled, Doming, and Melanotic. (B) Schematic of the training methodology. (C) Combined training and validation accuracy across all epochs, with dashed lines indicating the transition between Stage 1 and Stage 2. (D) Confusion matrix on the validation set showing classification performance across all five classes.



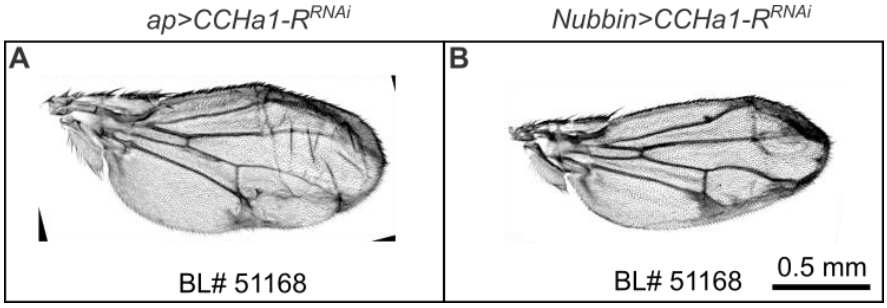
Supplementary Figure 2. Training and validation across all epochs. (A) Training and validation across all epochs, shown as **(i)** combined loss and **(ii)** Dice scores. **(B)** Qualitative performance of the trained U-Net model on the test dataset. Each row, from left to right, shows the raw image, the corresponding ground-truth labeling of the intervein compartments, the segmentation mask predicted by the U-Net model, and the probability map indicating the likelihood that each pixel belongs to the intervein tissue or the background. Each row represents a unique test sample. For compatibility with the U-Net input layer, images were resized to have equal x- and y-dimensions; therefore, the displayed images may not preserve the original aspect ratios of the raw images.



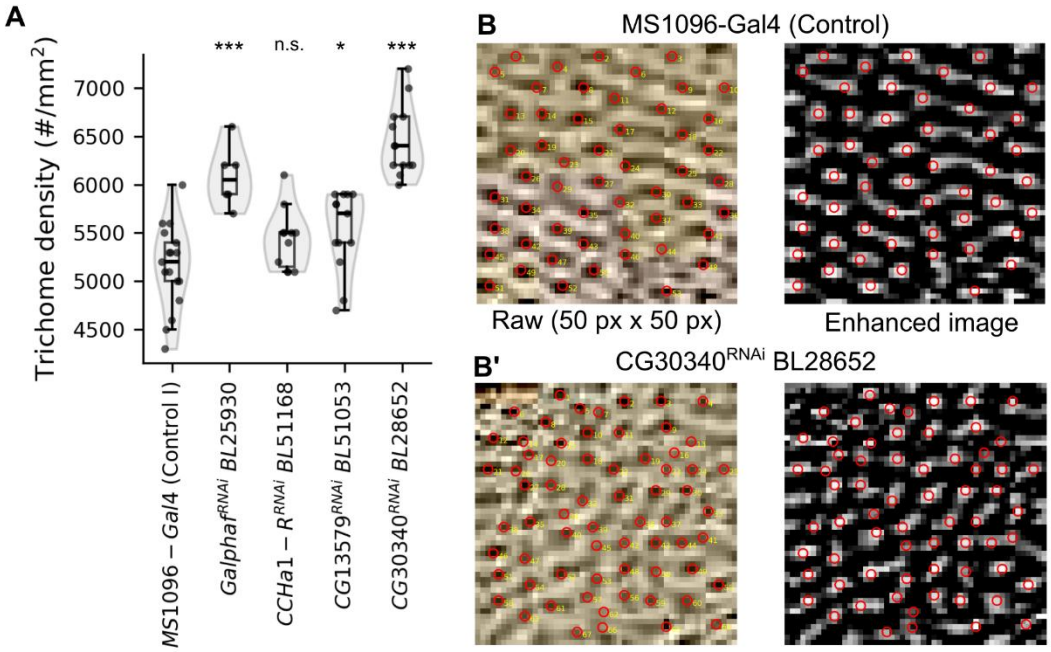
Supplementary Figure 3. Confusion matrix and intervein label schematics. **(A)** Confusion matrix showing the performance of K-nearest neighbor (KNN) classifier evaluated on the hold-out test dataset. Columns correspond to the ground truth intervein labels obtained from manual annotations, and rows correspond to the labels predicted by the classifier. Each matrix element reports the number (or fraction) of test samples from a given true class that was assigned to each predicted class, with diagonal entries indicating correct classifications and off-diagonal entries highlighting the misclassification patterns between intervein regions. **(B)** Schematics defining the anatomical intervein label used for classification. Each region is assigned a unique identity based on its position relative to the longitudinal veins and the anterior and posterior cross veins.



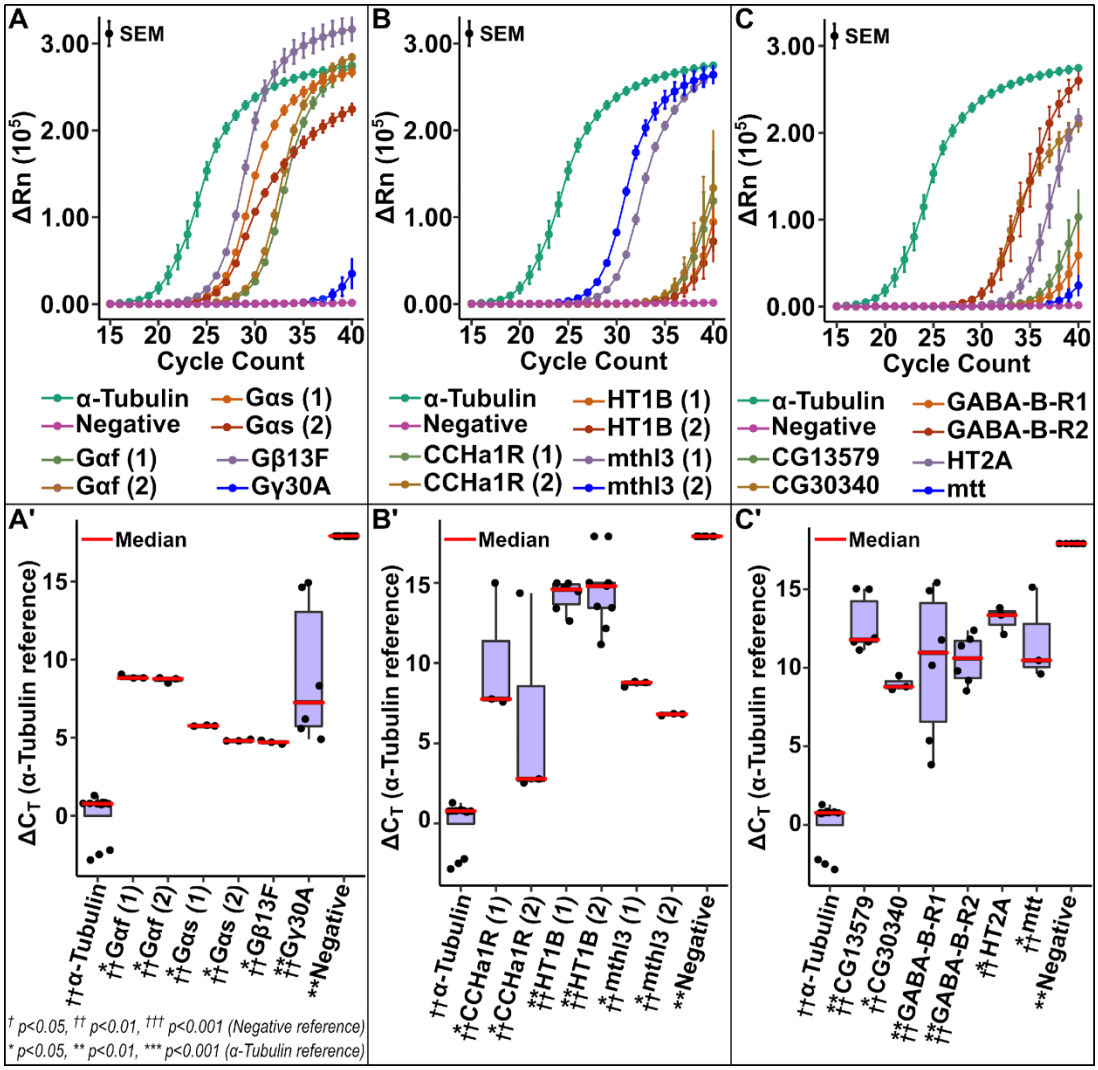
Supplementary Figure 4. Knockdown of selected GPCRs and G proteins causes severe morphological defects in the adult *Drosophila* wing. The MS1096-Gal4 driver was used to express UAS-RNAi constructs targeting GPCRs and G proteins in the larval wing imaginal disc as part of the loss-of-function screen. **(A)** Representative adult wing from the parental MS1096-Gal4 control, highlighting key structures of the adult wing blade**. (B–G)** Representative adult wings from GPCR and G-protein knockdown conditions that produced severe wing morphological defects.



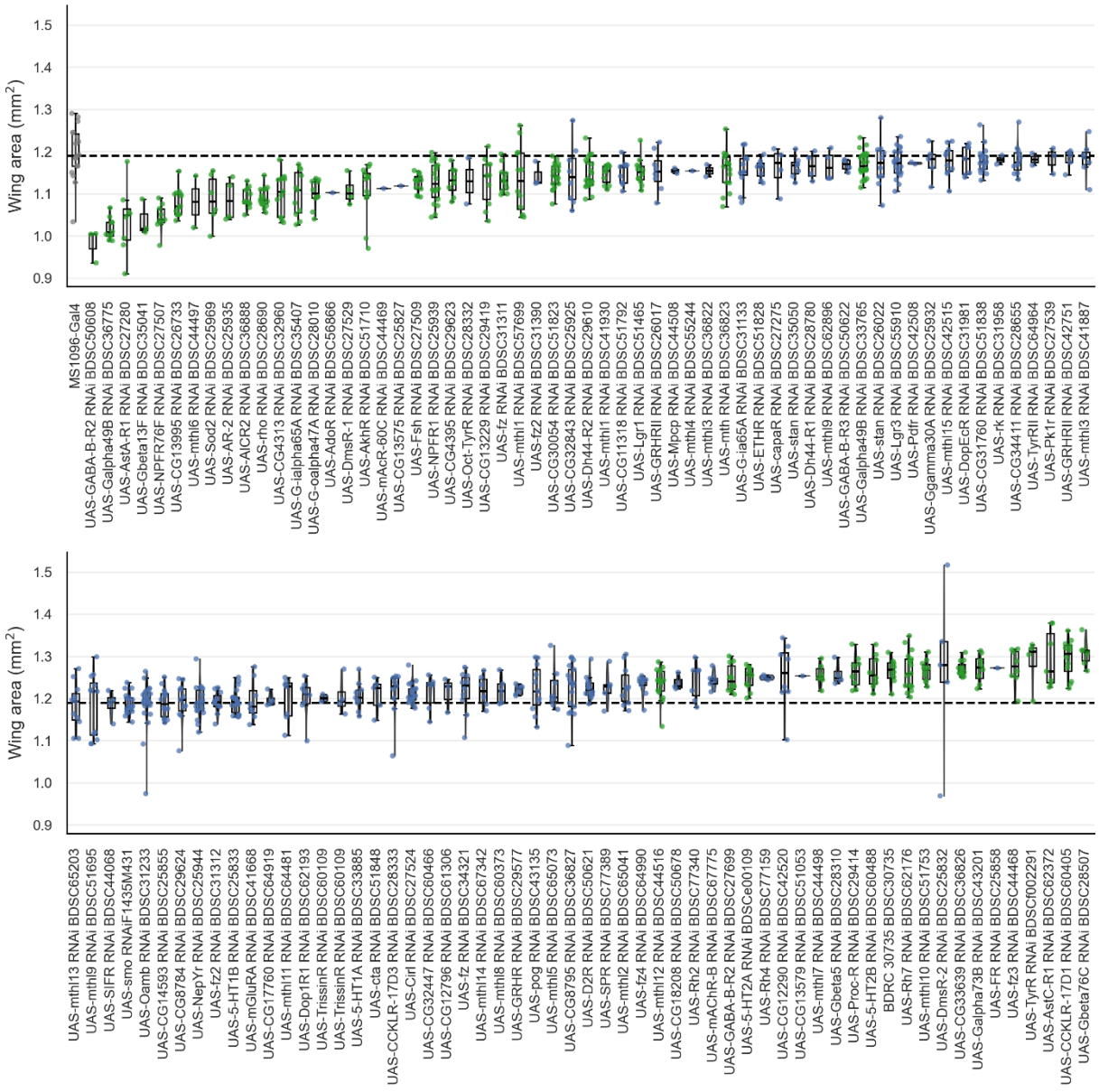
Supplementary Figure 5. Multiple Gal4 drivers produce similar phenotypes upon *CCHa1-R* knockdown. (A, B) An apterous-Gal4 (ap) and Nubbin-Gal4 drivers were used to knock down *CCHa1-R*. The figure shows representative wing samples from adult males.



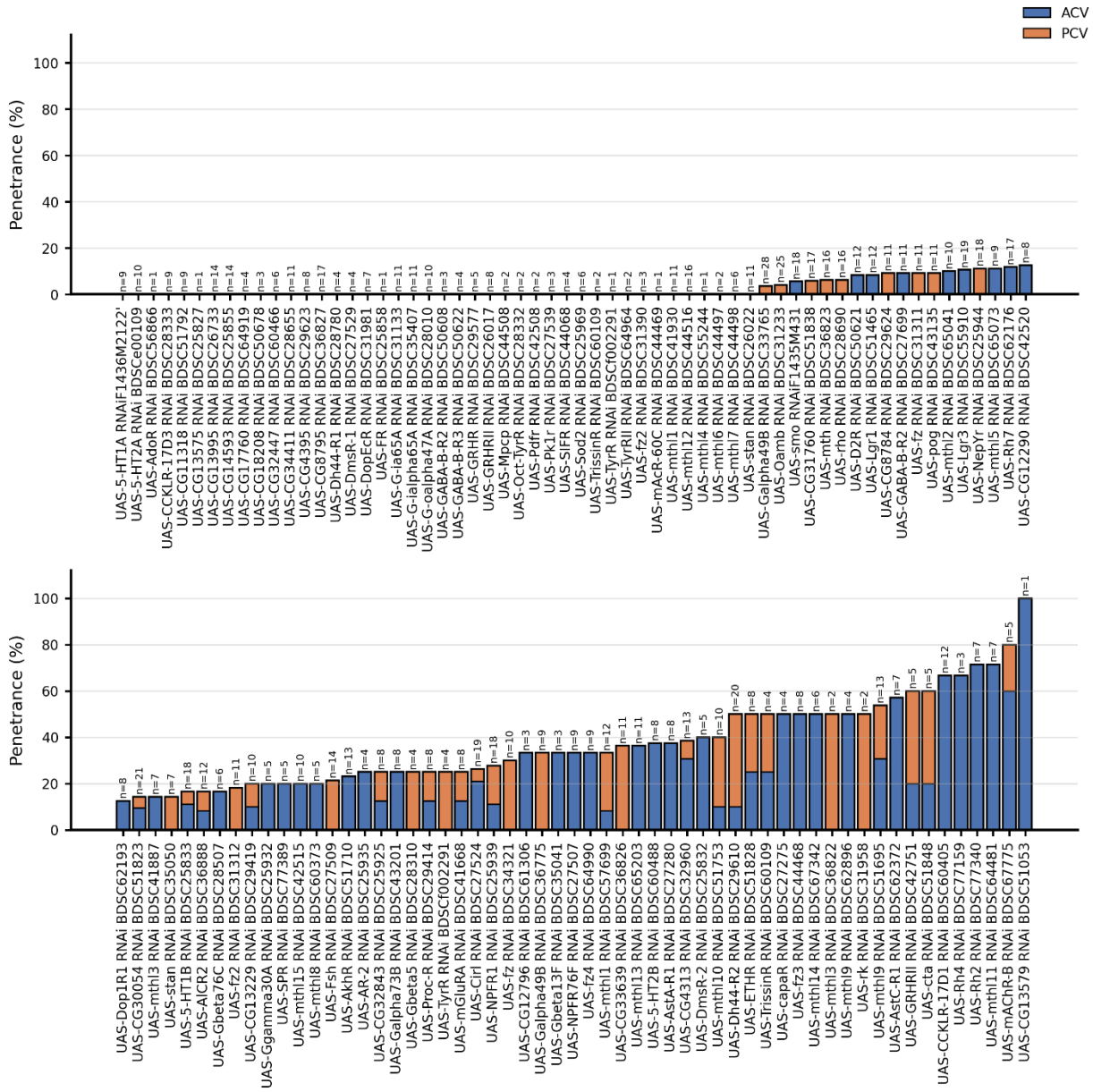
Supplementary Figure 6. Evaluation of trichome density associated with severe wing phenotypes shown in Figure 3. (A) Combined box and violin plots showing trichome density, defined as the number of trichomes within a 50 px × 50 px region of interest selected from the center of the wing in intervein region 4. Individual data points are overlaid. A t-test was used to determine whether the distributions were statistically significantly different. **(B, B')** Representative 50 px × 50 px crops from the parental MS1096-Gal4 control and CG30340^{RNAi} wings (1 px = 2 x 10^{-3} mm).



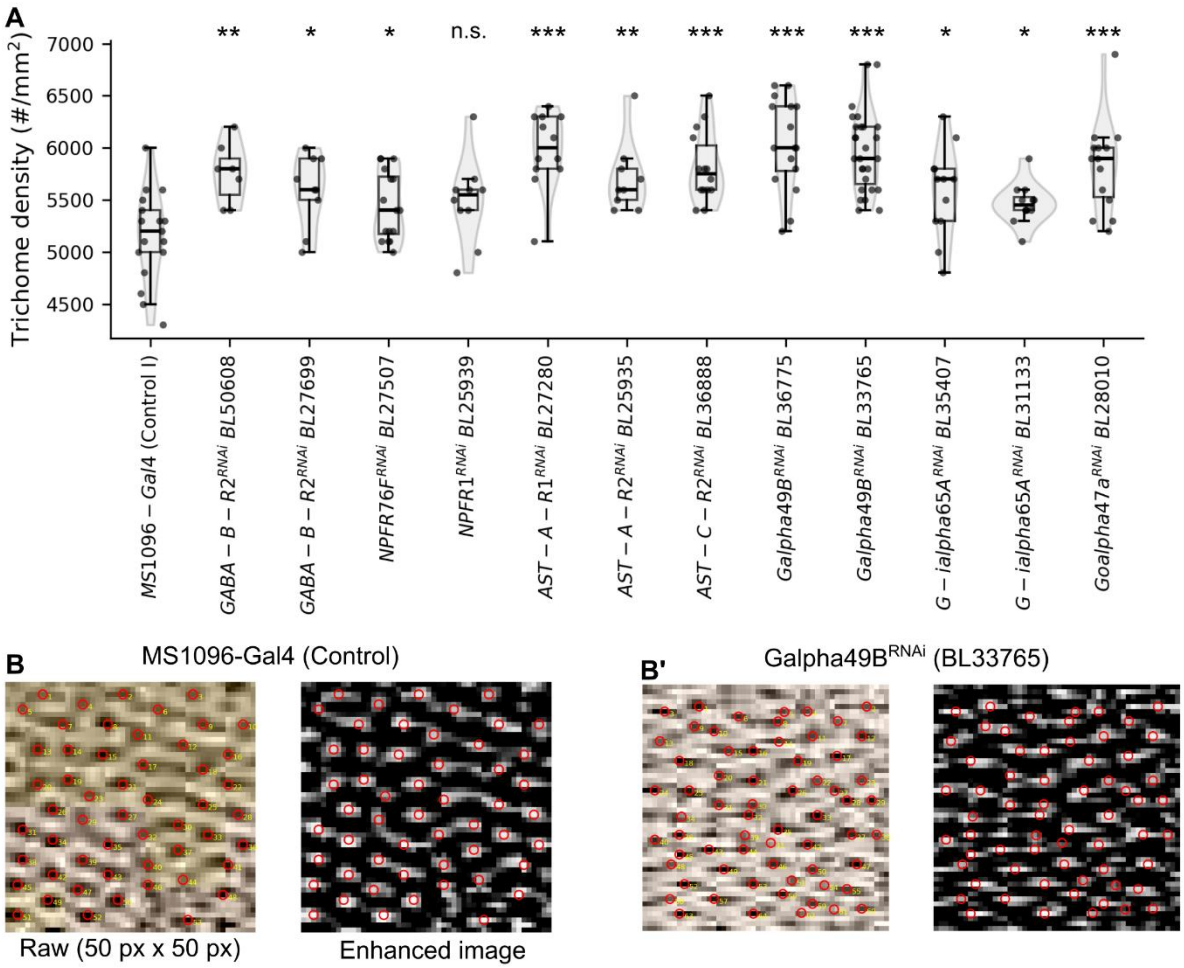
Supplementary Figure 7. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) confirms expression of positive hits. RT-qPCR was performed on cultured third instar larval wing disc cells (CME W1 Cl.8+, DGRC Stock 151). Select positive hit genes that exhibited severe wing phenotypes were tested to confirm expression in the *Drosophila* imaginal wing disc cell line. α-Tubulin at 84B (NCBI Reference Sequence: NM_057424), a ubiquitously expressed gene, was used as a positive control. No template control (NTC) wells containing no template DNA were used as a negative control. **(A-C)** The difference in normalized fluorescent signal from the experimental reaction and the baseline signal generated by the StepOneTM Software are plotted (ΔRn). Error bars represent the standard error of the mean (SEM) of each gene at a given cycle count. **(A'-C')** The difference in the number of amplification cycles required to reach the StepOneTM Software-generated threshold between the experimental gene and the positive control is plotted (ΔCT). Experimental runs in which no amplification was detected were assigned CT values of 40, as this was the limit of the experimental runs. Pairwise Mann-Whitney U tests were performed for statistical analysis with false discovery rate (FDR) correction (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ compared to positive control; † $p < 0.05$, †† $p < 0.01$, ††† $p < 0.001$ compared to negative control). **(A-A')** ΔRn and ΔCT plots for various G proteins. Numbers in parentheses indicate the use of multiple, distinct oligonucleotide primers. **(B-B')** ΔRn and ΔCT plots for various GPCRs where multiple, distinct (numbers in parentheses) oligonucleotide primers were evaluated. **(C-C')** ΔRn and ΔCT plots for various GPCRs in which only one pair of oligonucleotide primers was assessed.



Supplementary Figure 8. Quantification of adult wing area in the GPCR and G-protein RNAi screen. Adult wing area distributions were visualized for all GPCR and G-protein knockdown genotypes using combined violin and box plots with individual data points overlaid. The MS1096-Gal4 driver-only condition was used as parental control. An unpaired t-test was used to compare each knockdown genotype with the MS1096-Gal4 control. Data points shown in green indicate genotypes that differ statistically from the control group, defined as $p \leq 0.05$.



Supplementary Figure 9. Quantification of anterior and posterior crossvein defects in the GPCR and G-protein RNAi screen. Bar graphs show the total penetrance of crossvein defects for each genotype, defined as the percentage of adult wing samples displaying either anterior crossvein (ACV) or posterior crossvein (PCV) defects. The stacked bars indicate the relative contribution of ACV and PCV defects within each genotype. Each label represents a loss-of-function GPCR or G-protein RNAi condition driven by MS1096-Gal4 in the larval wing imaginal disc, with crossvein defects quantified in the adult wing blade.



Supplementary Figure 10. Evaluation of trichome density associated with severe wing phenotypes shown in Figure 4. (A) Combined box and violin plots showing trichome density, defined as the number of trichomes within a 50 px × 50 px region of interest selected from the center of the wing in intervein region 4. Individual data points are overlaid. A t-test was used to determine whether the distributions were statistically significantly different. **(B, B')** Representative 50 px × 50 px crops from the parental MS1096-Gal4 control and Galpha49B^{RNAi} wings (1 px = 2 x 10^{-3} mm).

User manual for MAPPER 2.0

1. **Load Image button**: Opens a file dialog to load a wing image (PNG, JPG, or TIFF). Displays the raw image immediately. Predicts the phenotype and shows the class probabilities on right hand side panel.
2. **Segment & Classify button**: Runs the full image processing pipeline. A trained U-Net based segmentation of intervein compartments followed by classification of intervein regions and feature extraction was carried out. This generates the intervein masks, assigns intervein labels, computes wing geometry, saves output masks, and displays the feature table as a new popup window.
3. **Batch process folder**: Allows you to select a master folder containing multiple wing images. All images in the folder (and subfolders) are automatically processed, masks are saved, and a single CSV feature file is generated summarizing features of all wings. A progress bar shows how many files have been processed and how many remain.
4. **Overlay binary**: Shows the segmented intervein compartment silhouette in red on top of the raw image
5. **Overlay labeled**: Toggles display of the labeled intervein mask. Each intervein compartment is shown with a unique color corresponding to its predicted identity.
6. **Overlay whole wing:** Toggles display of the reconstructed whole-wing binary mask. The whole-wing contour is obtained by dilating intervein regions, merging them across veins, then eroding back to recover the wing outline.
7. **Show AP axis:** Displays the Anterior-Posterior (AP) axis. Note: Current algorithm for extraction of any biological axes requires the wing blade to have 7 intervein identity similar to a normal looking wing.
8. **Show PD Axis:** Displays the Proximal-Distal (AP) axis.
9. **Show d(L3-L4):** Displays the distance between the L3 an L4 vein boundaries
10. **Scale (µm per pixel)**: User-defined conversion factor from pixels to microns. All reported areas and distances are scaled using this value.
11. **Erode radius:** Controls the erosion radius applied during segmentation cleanup. Higher values remove more noise but may shrink intervein regions.
12. **Open radius:** Control the morphological opening radius used to smooth ,masks and remove small objects.
13. **Wing Mask Radius:** Controls the dilation/erosion radius used to reconstruct the whole-wing contour from intervein regions.

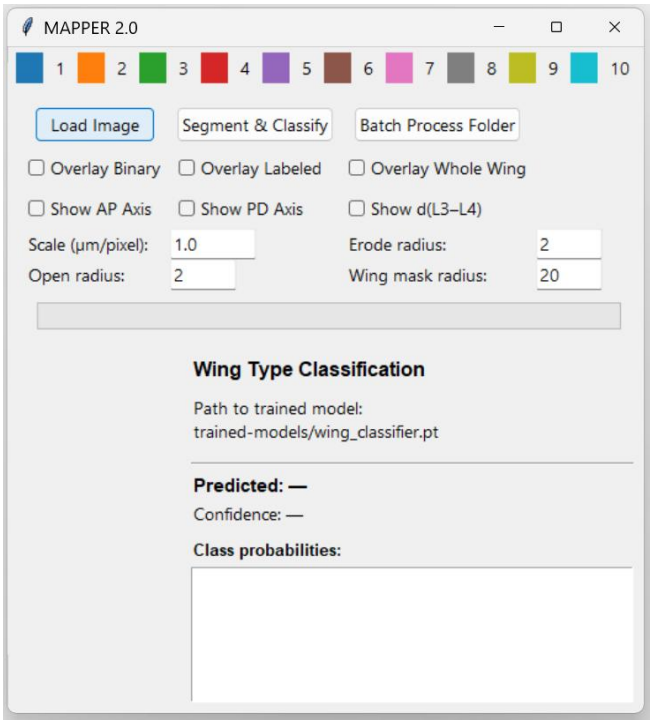
Feature table for individual wing processing: After running "Segment & Classify", a two-column table is generated. The first column contains the feature description, and the second column has the values. Each intervein labels (1-10) has two placeholders: (i) Intervein "i" area (μm^2) (ii) Intervein "i" centroid distance to wing center (µm). If an intervein label is absent, its value is reported as zero. Additional global features include wing area and lengths of biological axes. The table can also be exported as a CSV file manually.

Feature table for batch processing: During batch processing, the app generates a single CSV summary file in the selected master folder, where each wing image corresponds to one row in the table. The CSV file is organized to provide both phenotype-level and geometric quantitative outputs in a consistent format. For every sample, the file includes basic identifiers such as the filename, relative path and processing statues, followed by the predicted wing phenotype class, overall confidence, and the full set of class probabilities from the trained severe phenotype classifier. Quantitative morphometric features are reported in a standardized way: for each intervein label (1-10), the CSV contain the intervein area (in μm^2) and the centroid distance from the wing (in µm), with missing compartments automatically filled as zero to preserve consistent

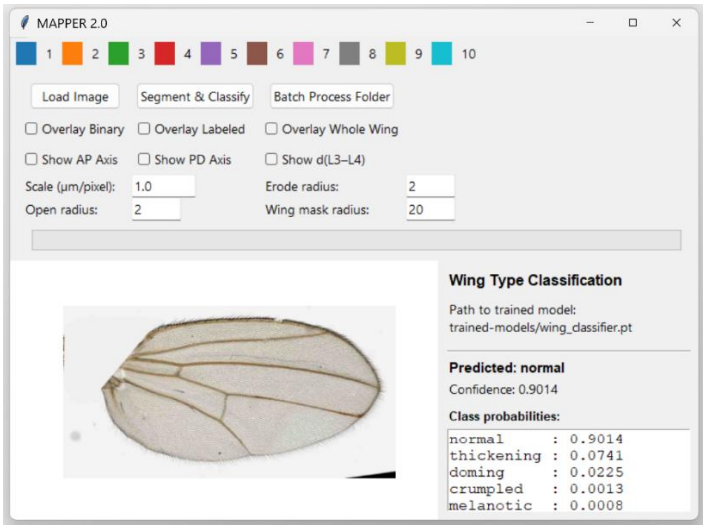
column structure. In addition, global wing0pscale measurements such as the total wing area and biological axis metrics (AP length. PD length, and d(L3-L4), when applicable) are included. Alongside the CSV, the pipeline automatically saves the corresponding binary, labeled and whole wing masks for every image, and if the masks already exist. The app skips re-running inference and directly uses the saved outputs for faster processing.

1 Steps to perform individual wing analysis

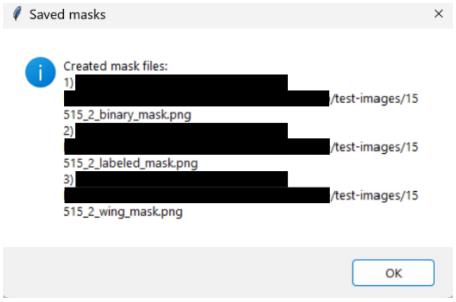
1.1 Follow the installation instructions in the root directory of the application to launch the GUI. Once launched, the GUI window should appear similar to the figure below,



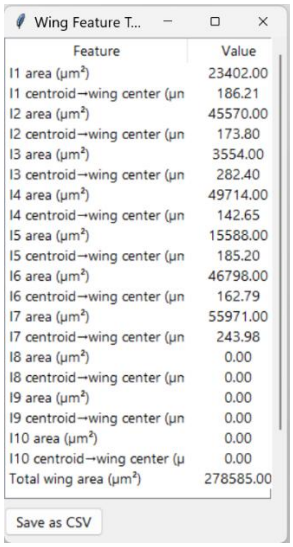
1.2 Click **Load Image** button. This opens a file explorer window, allowing you to select the wing image you want to process. Upon loading the raw image, the qualitative classification pipeline is executed and outputs the predicted phenotype class along with the associated probabilities. The raw image and classification output are displayed as shown below:



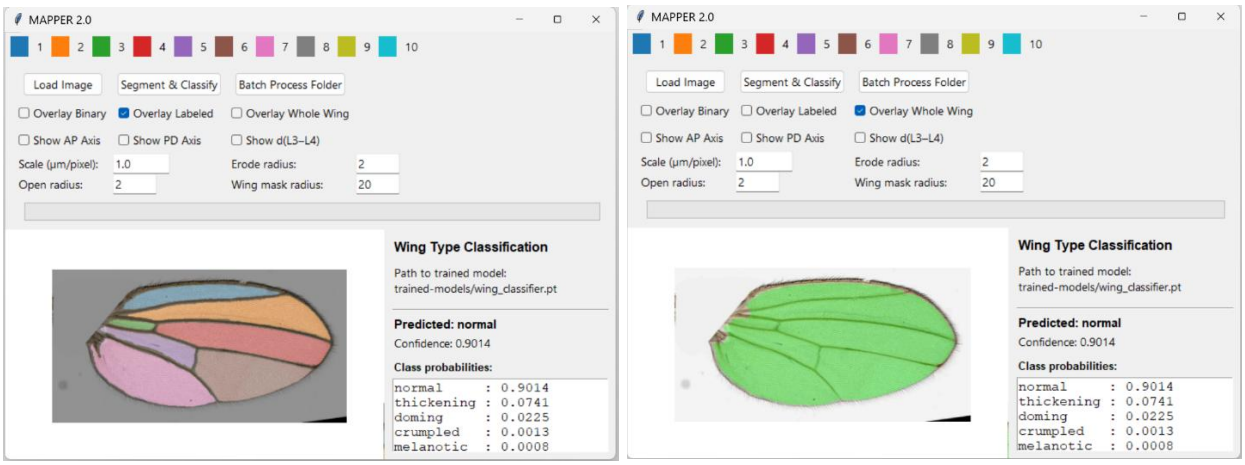
1.3 Click **Segment & Classify**. This performs (i) wing segmentation and (ii) classification of the intervein regions. Once processing is complete, a popup window appears with the file locations where the segmentation masks are saved.



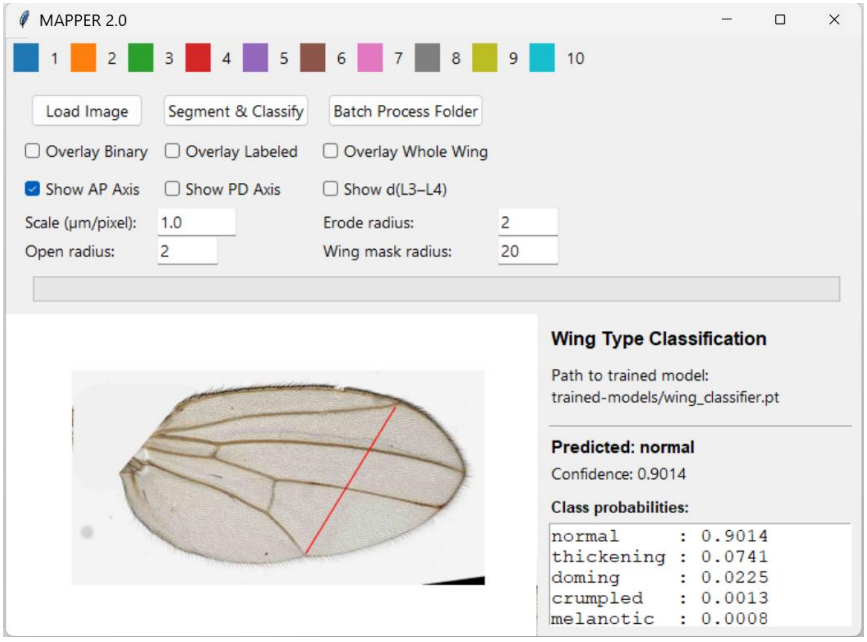
1.4 A feature table is also generated, as shown in the image below. The **Save as CSV** button allows the user to export this table as a .csv file.



1.5 Selecting the toggle switches **Overlay Labelled** and **Overlay Whole Wing** allows the user to display (i) the labeled intervein compartments (left) or (ii) the whole-wing mask (right) overlaid on the raw image. Only one overlay can be active at a time. Selecting one toggle automatically deactivates the other

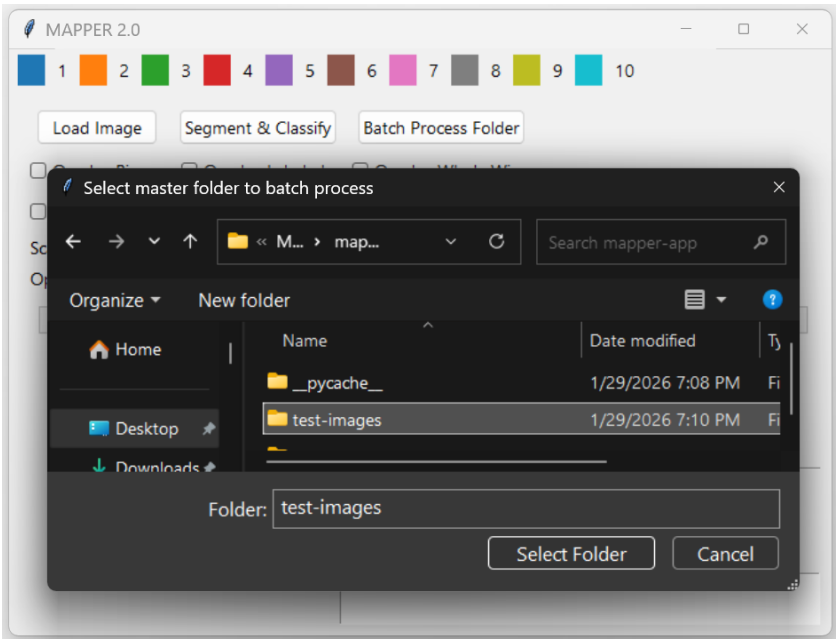


1.6 Enabling the **Show AP Axis** toggle displays the anterior-posterior (AP) axis overlaid on the image, as shown below:

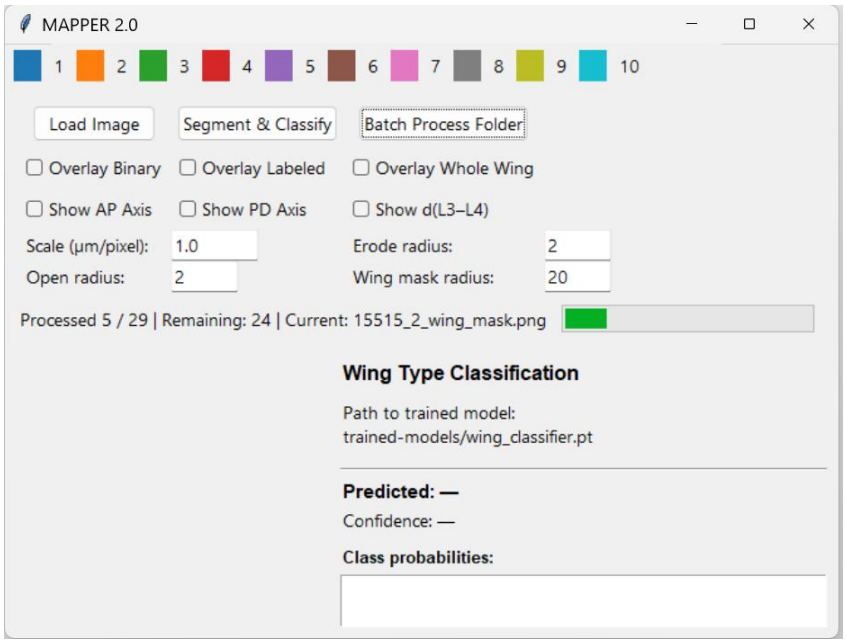


2. **Steps to perform batch processing**

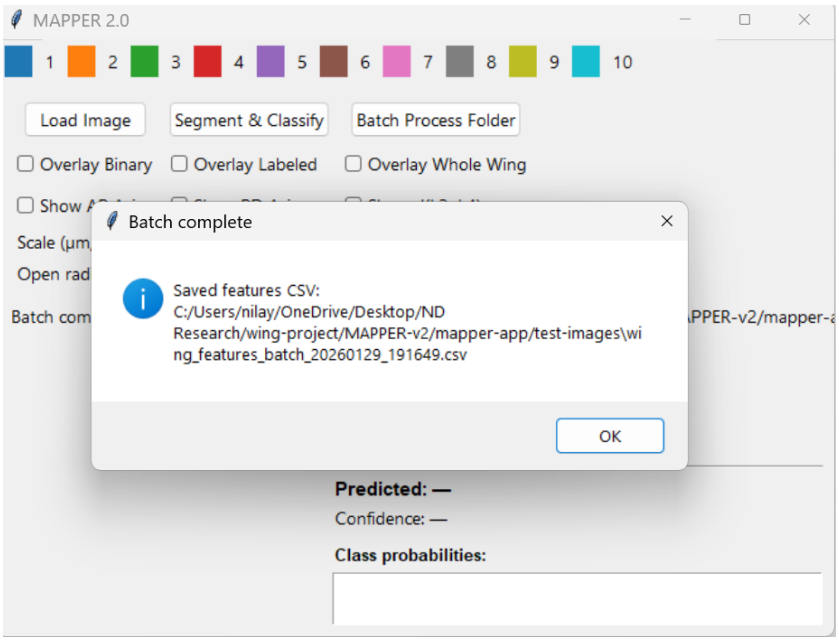
2.1 Click **Batch Process Folder** button. This opens a popup file explorer window, allowing the user to select the master folder containing all wing data, as shown below



2.2 Selecting the folder and pressing **Select Folder** initiates batch processing. Each wing image within the subfolders is automatically analyzed, and the corresponding segmentation masks are generated and saved. A progress bar in the center of GUI indicates the processing status and overall progress.



2.3 Once batch processing is complete, a pop-up window appears indicating the location where the summary statistics file has been saved.


MAPPER 2.0
1 2 3 4 5 6 7 8 9 10
Load Image
Segment & Classify
Batch Process Folder
Overlay Binary
Overlay Labeled
Overlay Whole Wing
Batch complete
Saved features CSV:
C:/Users/nilay/OneDrive/Desktop/ND Research/wing-project/MAPPER-v2/mapper-app/test-images\wing_features_batch_20260129_191649.csv
OK
Predicted: —
Confidence: —
Class probabilities: