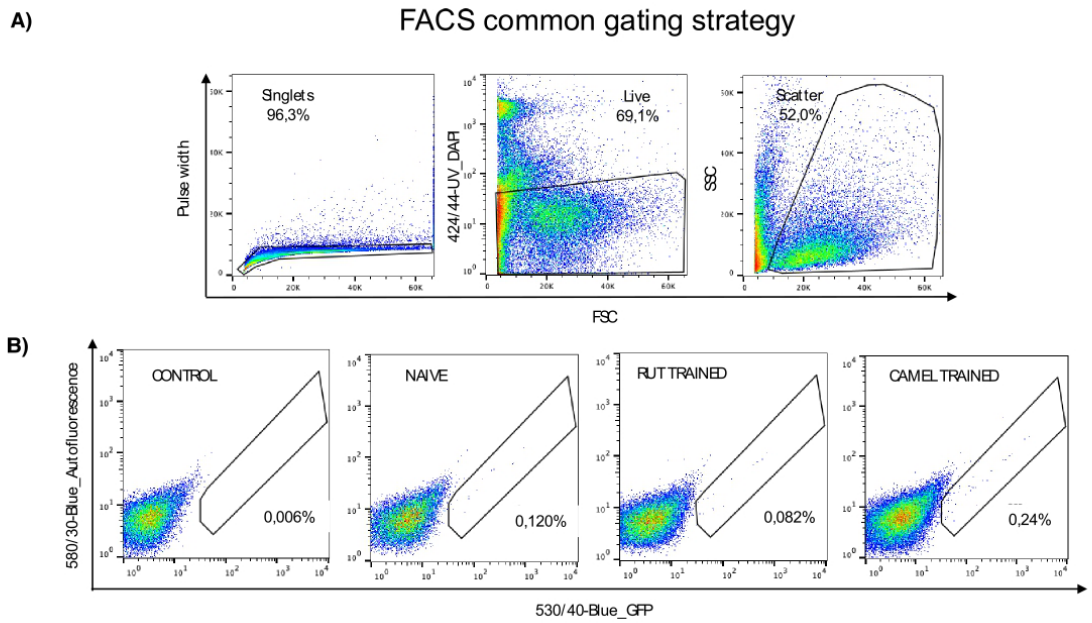
Supplementary tables (attached)

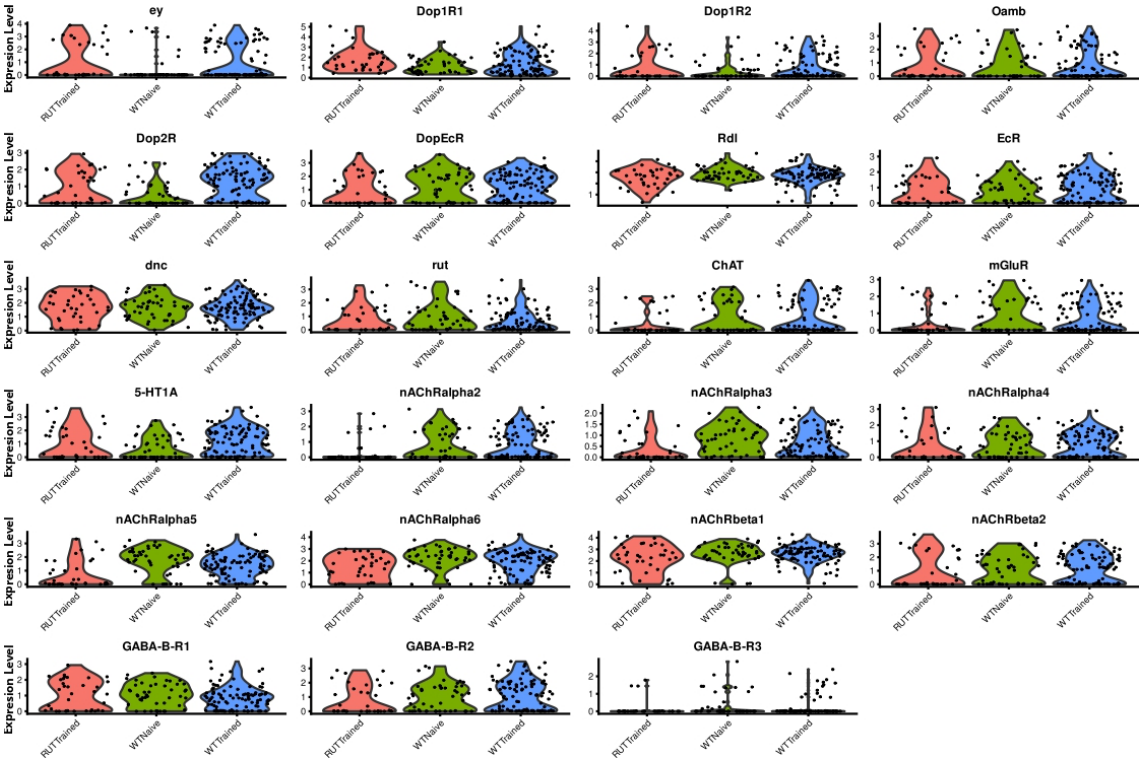
Table S1 – Whole head TIGs and TRGs
Table S2 – Mushroom Body TIGs and TRGs
Table S3 - List of TIGs in WH and MB+WHMB and the associated GO enrichment
Table S4 – Table of overlapping TIGs from different datasets
Table S5 – Associated GO enrichment of cluster 1-5
Table S6 – Table of MB- and lobe- specific genes
Table S7 – Associated GO enrichment of cluster ABC
Table S8 – List of TIGs with transcription factor binding within accessible regions of MB chromatin
Table S9 – Experimental Crosses and Genotypes

Supplementary figures



Supplementary figure 1 – CAMEL-GFP cell sorting of dissociated *Drosophila* brains.

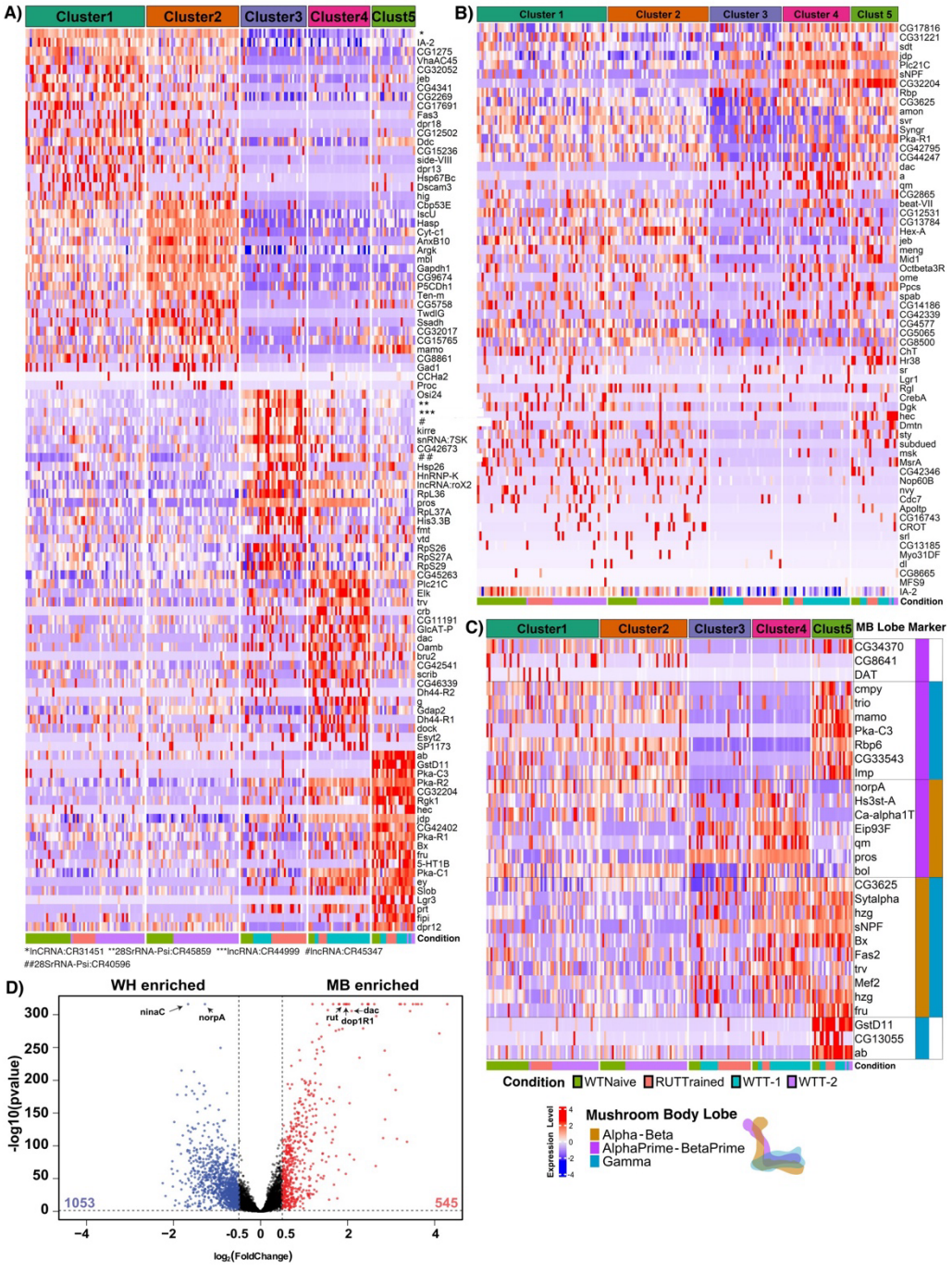
(A) The common gating strategy followed up on the sort to exclude cell aggregates, dead cells and cellular debris. (B) Percentage of the GFP positive events for each of control, WT naive, rutabaga mutant and WT trained conditions. GFP percentage is referred to those single live non debris events.



Supplementary figure 2 – Violin plots of MB markers

Violin plots of MB specific marker expression in isolated CAMEL neurons from trained rut (red), naive (green) and trained animals.

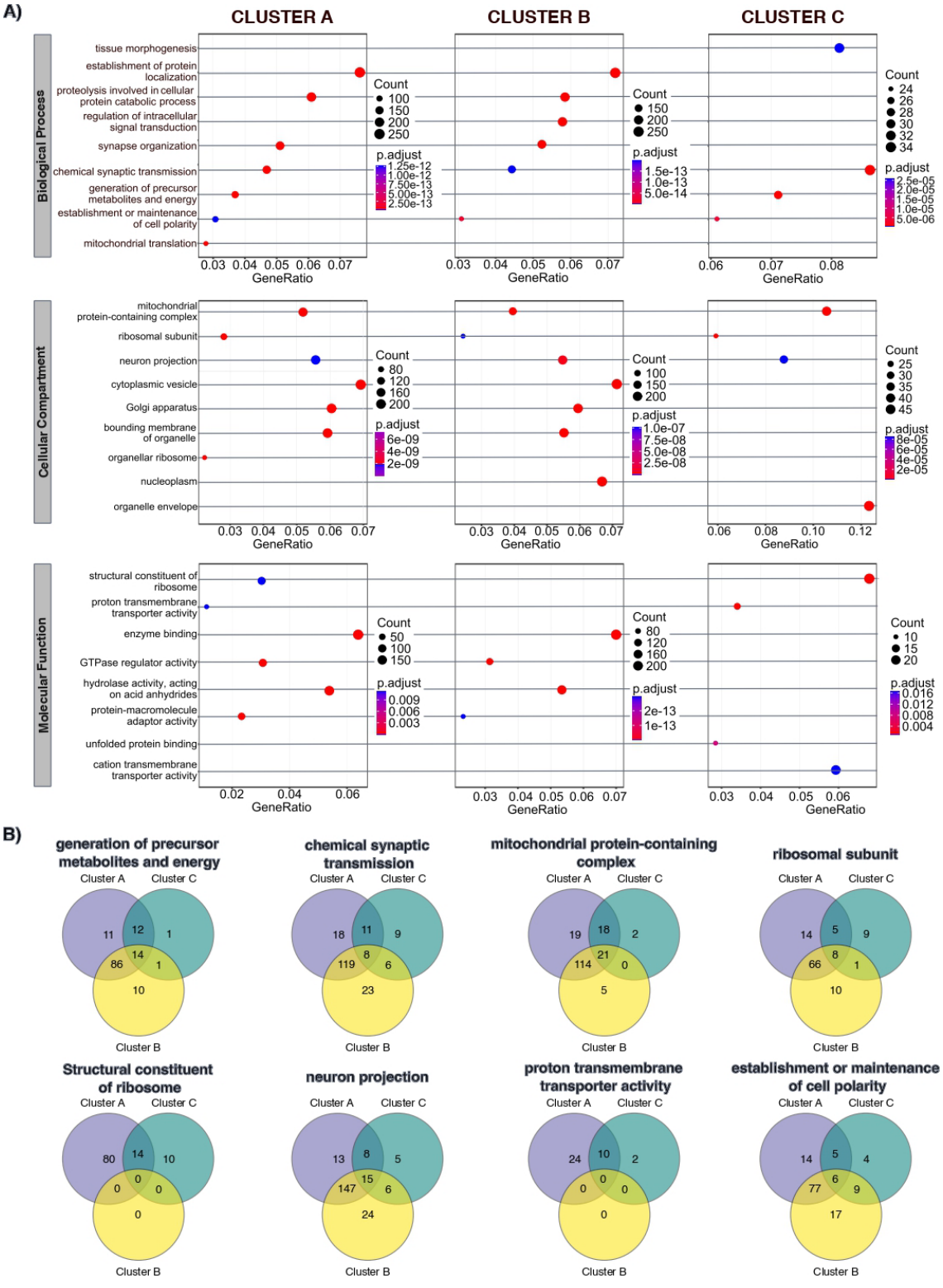
Figure Sup. 3



Supplementary figure 3 – Heatmaps of cluster 1-5

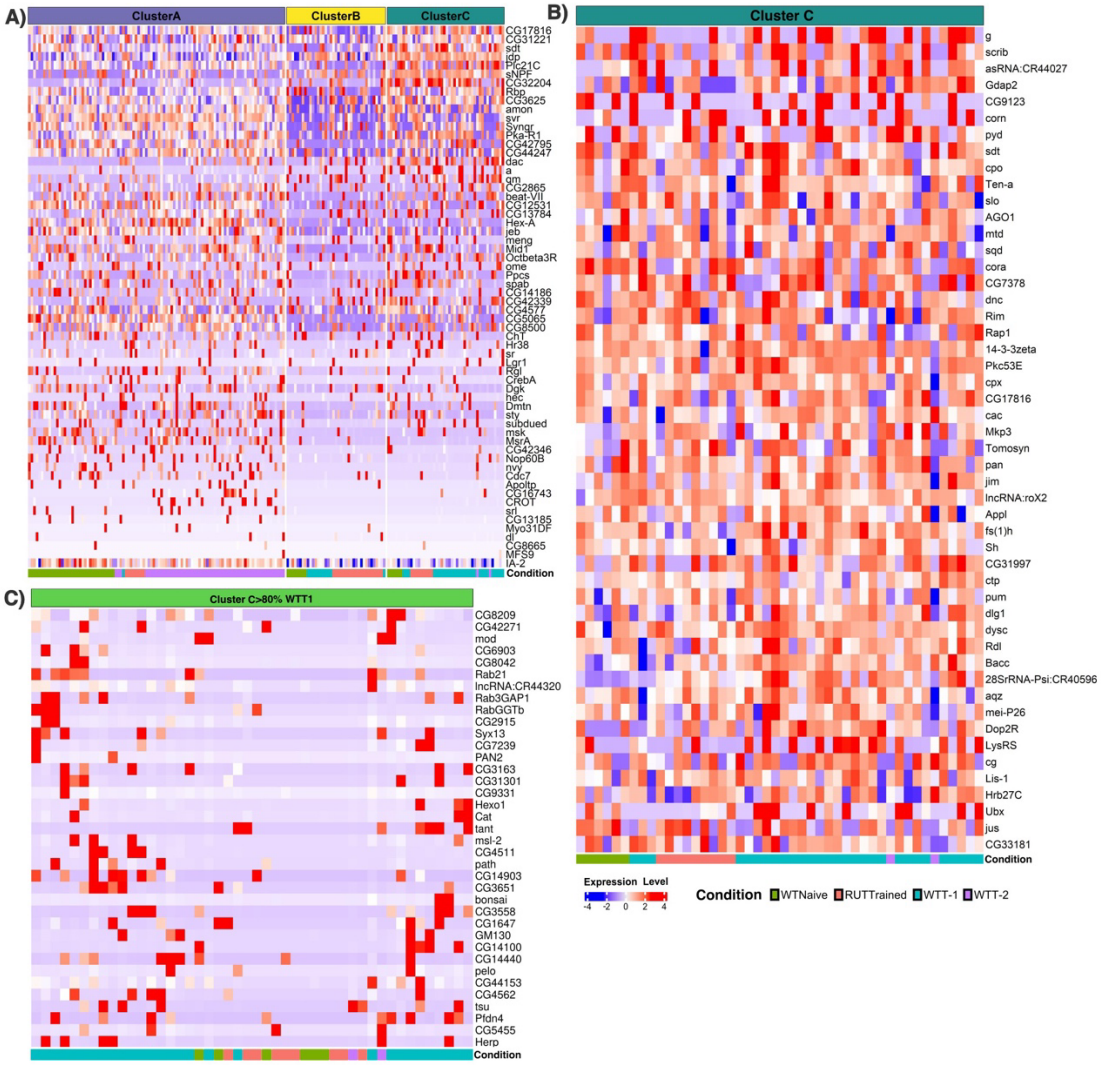
(A) Heatmap of top 30 most differentiating genes between cluster 1-5. (B) Heatmap of up-regulated genes after learning obtained by bulk RNAseq following

courtship conditioning. (C) Heatmap of 30 MB lobe markers. The lobe identity for markers is shown in lateral bars ($\alpha\beta$ in brown, $\alpha'\beta'$ in purple and γ in blue). Overlapping means that the marker is expressed in both lobes, although at different expression levels. (D) Volcano plot of differential expression analysis results between all whole-head (WH) (n=38) and INTACT-obtained mushroom-body (MB) samples (n=38). 6965 genes were used for differential expression analysis. Differentially expressed genes FDR<0.05) MB enriched are highlighted in red ($\log_2$ fold change >0.5) and WH-enriched genes are highlighted in b blue ($\log_2$ fold change <-0.5).



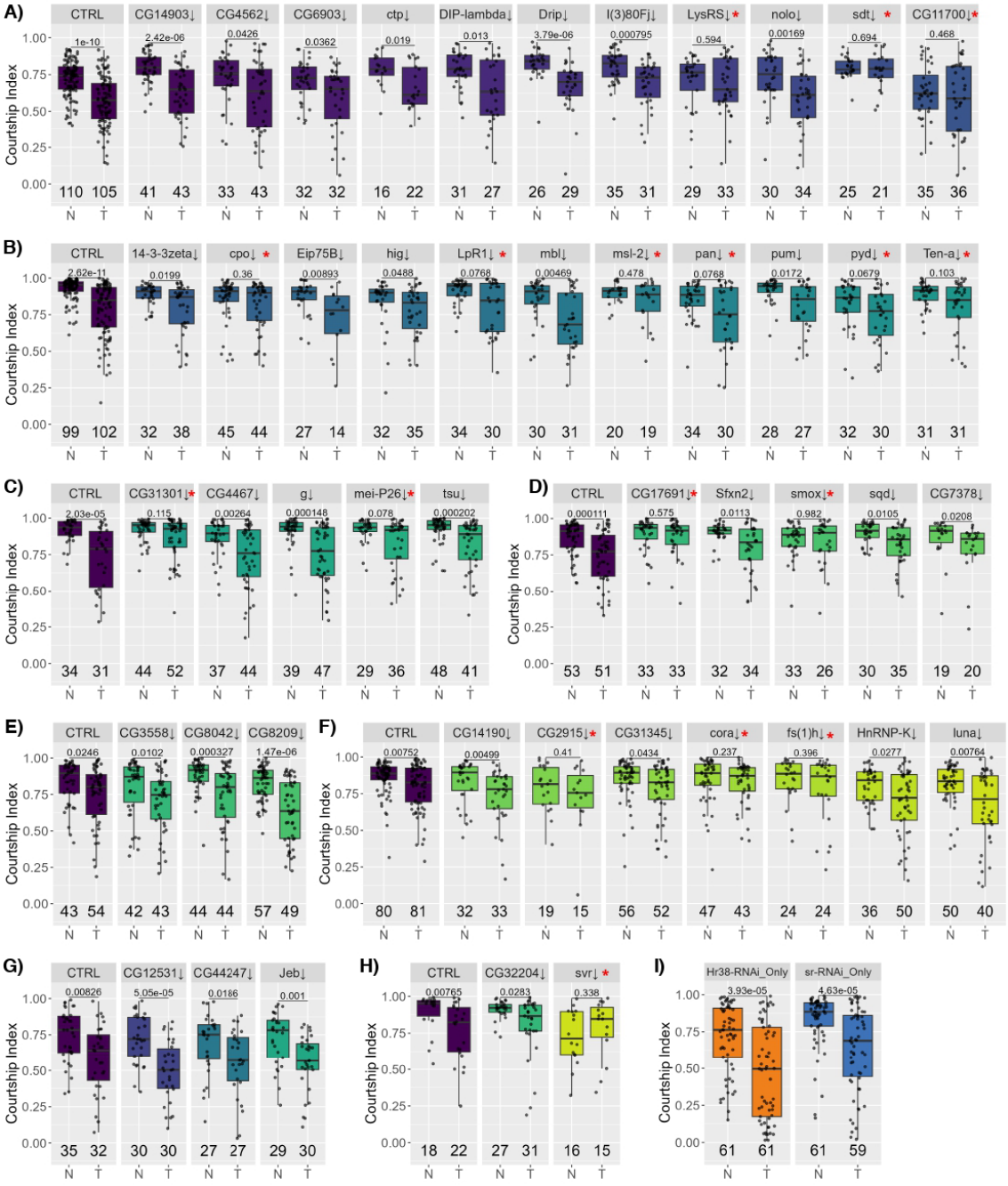
Supplementary figure 4.- Similar GO terms between cluster A-C.

(A) Most representative of significative GO terms between cluster A, B and C, categorized in biological processes, cellular component and molecular function. (B) Venn Diagram showing the number of genes for selected common GO terms between the three clusters (see also Suppl table 7).



Supplementary figure 5 – Heatmaps of ABC clusterization

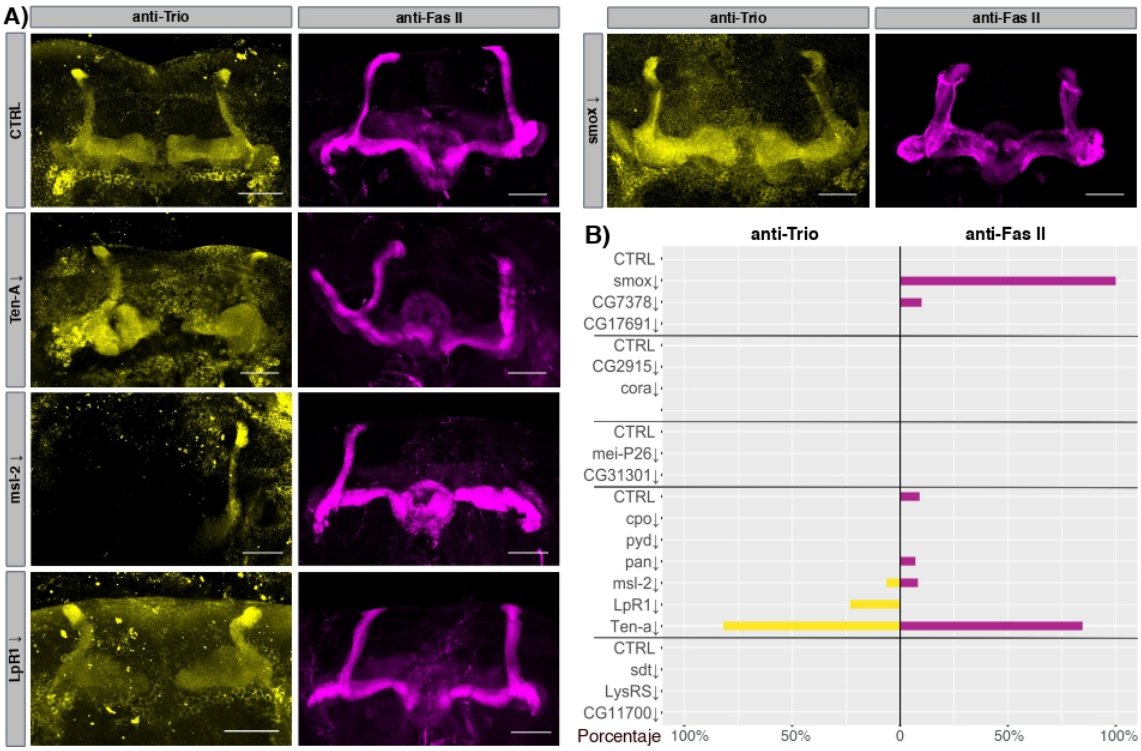
(A) Heatmap of up-regulated genes after learning obtained by bulk RNAseq following courtship conditioning under the ABC clusterization. (B,C) Heatmaps of candidate genes for cluster C (A) and cluster C>80% (B) neurons: genes that were up-regulated in more than 80% of WTT-1 neurons and in less than 20% of neurons from naive/rut flies within the cluster C.



Supplementary figure 6 – The full set of data for the functional screen

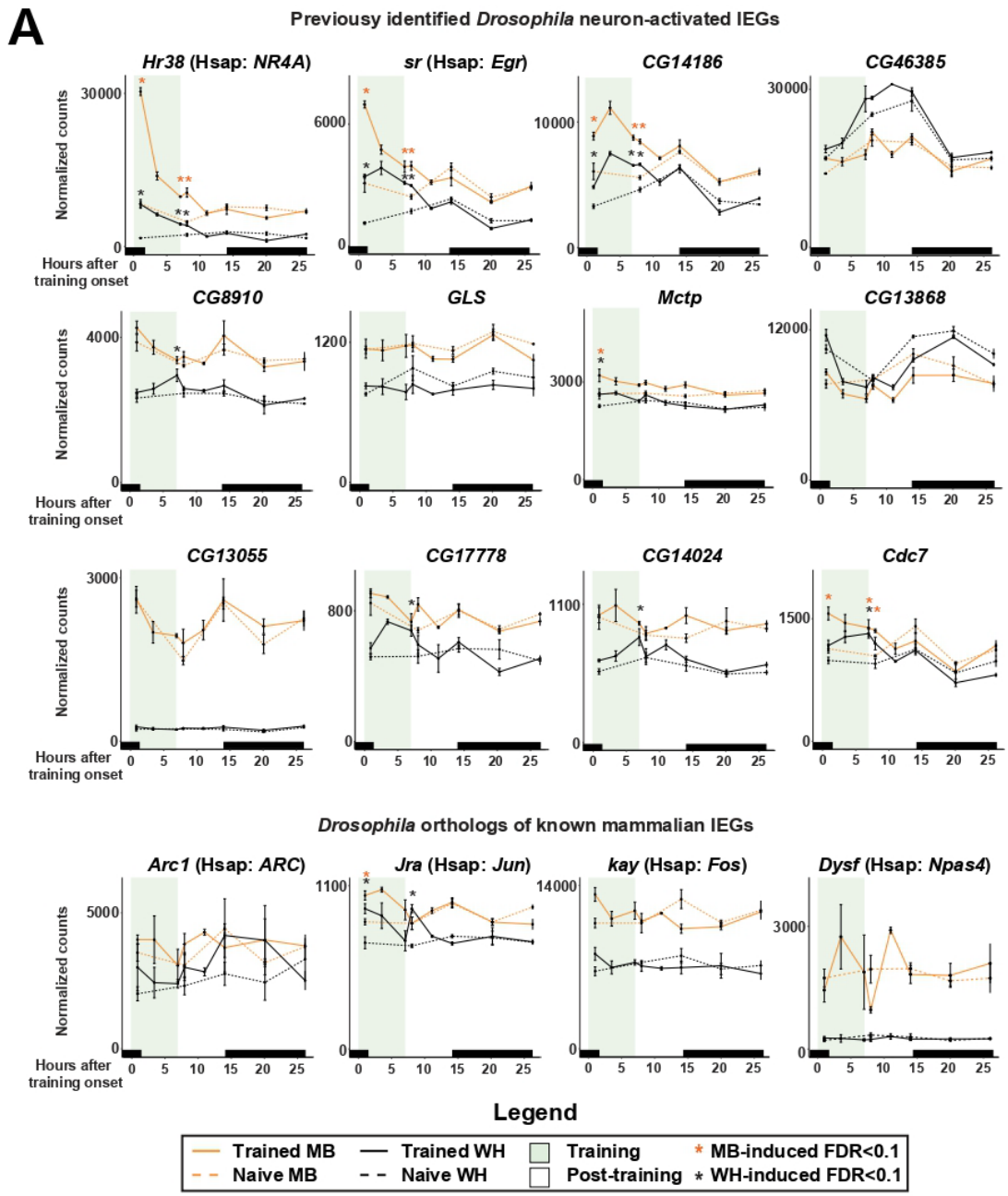
(A-H) Dot plot showing courtship indices (CIs) for naive (N) and trained (T) flies of 50 candidate genes in LTM assays. Each condition was repeated between 2-4 times. Red asterisks showed the positive hits that abolished learning/retrieval. The combination of the driver (MB247-gal4 or Rut-gal4), tub-gal80^{ts} and control (containing the transgene without an cDNA) lines was different for each panel (see suppl table 9 for details). (I) Dot plot showing courtship indices (CIs) for naive (N) and trained (T) flies of UAS-only controls for Hr38-RNAi and sr-RNAi in LTM assays. Statistical significance between naive and trained flies was determined

using a Mann-Whitney test. Mean is displayed with error bars indicating +/- SEM. Number of flies tested for each condition is shown under corresponding dot plot.



Supplementary figure 7 – MB staining

(A) MB staining of positive hits with anti-TRIO (yellow) and fascicline II (cyan), which distinguish α´β´ and γ and αβ and γ lobes, respectively. Only a representative control and a representative altered MB are shown. Scale bar represents 50 µm. (B) Percentage of MB with alterations in the α´β´ or γ lobes (yellow) and αβ or γ (magenta) respect to normal MB.


A
Previousy identified Drosophila neuron-activated IEGs
Hr38 (Hsap: NR4A)
sr (Hsap: Egr)
CG14186
CG46385
CG8910
GLS
Mctp
CG13868
CG13055
CG17778
CG14024
Cdc7
Normalized counts
Hours after training onset
Drosophila orthologs of known mammalian IEGs
Arc1 (Hsap: ARC)
Jra (Hsap: Jun)
kay (Hsap: Fos)
Dysf (Hsap: Npas4)
Legend
Trained MB
Naive MB
Trained WH
Naive WH
Training
Post-training
MB-induced FDR<0.1
WH-induced FDR<0.1

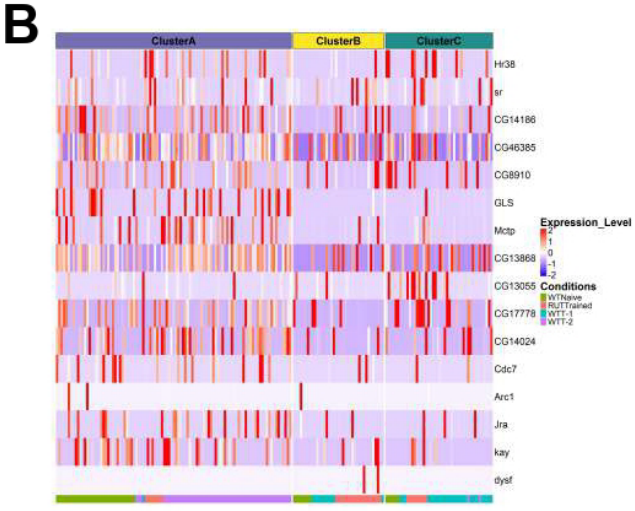

B
ClusterA
ClusterB
ClusterC
Hr38
sr
CG14186
CG46385
CG8910
GLS
Mctp
CG13868
CG13055
CG17778
CG14024
Cdc7
Arc1
Jra
kay
dysf
Expression_Level
Conditions
WTNaive
RUTTrained
WTT-1
WTT-2

Figure S8 – Candidate immediate early genes in courtship LTM

Normalized transcript levels of (A) 12 *Drosophila* immediate-early genes (IEGs) previously identified[55] and (B) for fly orthologs of known mammalian IEGs. Significance identified from differential expression analysis between trained and time-of-day matched naive flies, at time-points noted on the x-axis, is displayed (*=FDR <0.1). Bold black line indicates incubator dark period. (C) Heatmap of the 12 *Drosophila* immediate-early genes (IEGs) and for fly orthologs of known mammalian IEGs previously identified among the three ABC clusters. The lower bar in the graph indicates the experimental condition for each sequenced neuron.

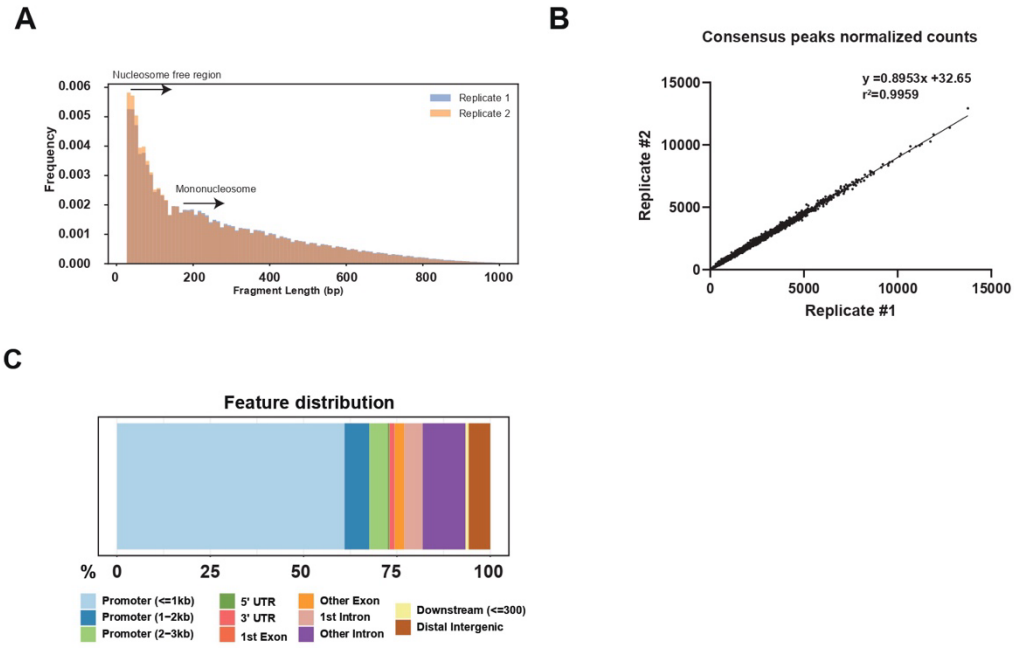


Figure S9 – ATAC-seq quality control

(A) Fragment size distribution of ATAC-seq libraries generated from INTACT-isolated MB nuclei. Distribution for two biological replicates is shown. Peak signal of nucleosome free regions (80-120 bp) and mononucleosomes (~180 bp) is marked. (B) Scatter plot of normalized counts for consensus peaks from MB ATAC-seq samples. Each dot represents a consensus peak, with positions indicating normalized counts in replicate 1 and replicate 2. Line of best fit is shown, with corresponding linear equation, and coefficient of determination (R^2). (C) Feature distribution of annotated regulatory regions from significantly accessible peaks and is predominantly located near to the transcriptional start site (TSS) of genes.