Title: Evolution of decreased aggression in honey bees targeted brain cell gene regulatory networks

Ian M. Traniello[1,2†*], Syed Abbas Bukhari[2], Payam Dibaeinia[3], Guillermo Serrano[4], Arian Avalos[5], Amy Cash Ahmed[2], Alison Sankey[2], Mikel Hernaez[4], Saurabh Sinha[2,3], Sihai Dave Zhao[2,6], Julian Catchen[2,7], Gene E. Robinson[1,2,8*]

Affiliations:

[1]Neuroscience Program, University of Illinois at Urbana-Champaign; Urbana, IL (UIUC).

[2]Carl R Woese Institute for Genomic Biology, UIUC.

[3]Department of Computer Science, UIUC.

[4]Computational Biology Program, CIMA University of Navarra; Pamplona, Spain.

[5]United States Department of Agriculture – Agricultural Research Services, Honey Bee Breeding, Genetics and Physiology Research Laboratory; Baton Rouge, LA.

[6]Department of Statistics, UIUC.

[7]Department of Evolution, Ecology and Behavior, UIUC.

[8]Department of Entomology, UIUC.

[†]Current Address: Lewis-Sigler Institute for Integrative Genomics, Princeton University; Princeton, NJ.

*Corresponding authors. IMT: it4770@princeton.edu; GER: generobi@illinois.edu

Abstract: A major challenge in biology is to understand how inherited variation results in phenotypic differences, especially for complex polygenic traits like behavior. Integrating genetic variants associated with evolutionary differences in honey bee aggression with genes identified via brain single-cell (sc) transcriptomics and sc gene regulatory network (GRN) modeling, we achieved robust genotype-to-phenotype mapping. We identified genetic differences in GRN structure associated with evolutionary differences in aggression, located in specific brain cell populations related to olfaction and vision. Some involve serotonin, which already was known to influence bee aggression, but not yet from a genetic perspective. These results demonstrate that, as theory predicts, GRN analysis can integrate evolutionary and systems biology to identify molecular mechanisms underlying adaptive polygenic traits.

One Sentence Summary: Genetic variation affects behavioral variation by causing changes in brain cell gene regulatory networks.

Main text: Recent theory predicts that the genetic architecture of polygenic traits is best understood at the level of the gene regulatory network (GRN), which models regulatory relationships between transcription factors (TFs) and their target genes (TGs) (*1*, *2*). This prediction is consistent with an emerging appreciation that behaviorally related GRNs are an important level of organization in the brain (*3*). Developmental GRNs have proven to be very useful in the study of species evolutionary changes in morphological traits and have inspired similar studies of brain and behavior (*4*, *5*). We tested this prediction (*1*), using the recent evolution of decreased aggression in a Puerto Rican population of the otherwise highly aggressive Africanized honey bee (AHB, *Apis mellifera scutellata*) (*6*).

A prior genome-wide association study (GWAS) of the Puerto Rican AHB population failed to detect single nucleotide polymorphisms (SNPs) that associate with the behavior of "soldiers" (*7*) (highly aggressive individuals that act as a colony's first line of defense) or similarly aged but less aggressive nectar and pollen foragers. By contrast, that study did detect 253 genes with SNPs that associated with colony-level variation in aggression, and many of them showed strong signals of selection (*6*, *7*). These results suggested that these "colony aggression genes" reflect heritable differences between soldiers and foragers that are dependent upon colony environment. We used these colony aggression genes, bulk and single-cell (sc) transcriptomics, and GRN analysis to explore how variation in group genetics can affect variation in individual brain function to influence colony aggression.

We compared whole-brain transcriptomic profiles of 81 soldiers and 82 foragers from nine Puerto Rican honey bee colonies, using the same individuals previously subjected to whole-genome analysis (*7*). We coupled this analysis with scRNA-Sequencing to profile the transcriptomes of ~24,000 cells from 40 soldier and forager brains pooled across eight samples collected from four additional high- and low-aggression colonies in Urbana, Illinois (fig. S1; fig. S2A-B; Data S1). This allowed us to detect differentially expressed genes (DEGs) between soldiers and foragers, both in bulk brain samples and resolved to the level of specific brain cell populations. However, neither DEG list overlapped significantly with the colony aggression genes (Fig. 1A; fig. S2C; Data S2-S3). We therefore looked for connections between genetic variation in colony aggression and gene regulation.

We modeled GRNs from the sc data and developed a regulatory dissimilarity score, a measure of network difference that integrates across all TF-TG relationships. There was greater GRN dissimilarity between soldiers and foragers from more aggressive colonies relative to less aggressive colonies (Fig. 1C-E; fig. S3). These genetic differences were prominent in two cell populations of note: the non-compact Kenyon Cells (KC) of the mushroom bodies (MB) and astrocyte-like glia. The former have been associated with honey bee aggression via *in situ* hybridization (*8*) and the latter have been associated with behavioral plasticity in honey bees (*9*) and ants (*10*), but never before from a genetic perspective.

Having established a relationship between gene regulation and genetic variation in colony aggression, we used expression quantitative trait locus (eQTL) mapping to search for specific genes linking variation in group genetics to variation in individual brain function (*11*). We first identified thousands of eQTL-genes (eGenes) containing *cis*-located SNPs (eSNPs) in soldiers and foragers, and filtered this list to 30 eGenes by requiring them to have eSNPs that were also SNPs in the colony aggression genes (*7*) (Data S4). There was no significant overlap with bulk or sc soldier-forager DEGs (Fig. 1B; fig. S2C), nor did any eGenes overlap with genes implicated in a recent genomic analysis of aggression in a Brazilian AHB population (*12*). We thus next studied the influence of these 30 eGenes on the brain gene regulatory architecture underlying the soldier and forager behavioral states.

To do this, we developed a new computational method, similar to ref. (*13*), to test the prediction that if eGenes act in concert they will form tightly connected regulatory subnetworks, measured by the strength and degree of TF-TG relationships. We analyzed GRN subnetworks composed of either the 30 eGenes as TGs or, as a negative control, 15 random permutations of equally sized subnetworks using eGenes with eSNPs that did not overlap with SNPs in colony aggression genes (*7*). As predicted, the eGene subnetworks were highly similar to each other in terms of TF-TG relationships, more so than the random subnetworks, which did not show any cohesive signal (Fig. 2A-B; fig. S4). This finding points toward a mechanism by which group genetics can shape brain and behavior, by influencing the functional integration of GRNs.

The eGene subnetworks featured four highly connected TFs, each with a robust one-to-one ortholog in *Drosophila melanogaster*: *Single stranded-binding protein c31A* (*Ssb-c31a*), *apontic* (*apt*), *ecdysoneless* (*ecd*), and *CG31460* (Fig. 2C; Data S5). *Ssb-c31a* was most prominent in astrocyte-like glia, already identified above, and perineurial glia. *apt* was most prominent in the inner-compact KC, a small population of MB neurons flanking the above-mentioned non-compact KC which have also been implicated in honey bee aggression (*8*), and in the antennodorsal projection neurons of the antennal lobe, which communicate olfactory sensory information to the KC. There were 15 TF-TG relationships in the eGene subnetworks that showed strong soldier-forager differences, including *Ssb-c31a* and *apt*, which also showed the strongest expression differences between soldiers and foragers (Data S6). These results suggest that inherited changes in honey bee aggression are based, in part, on changes in the structure and function of olfactory pathways that respond to or process aggression-triggering olfactory stimuli.

Six of the 15 TF-TG relationships involved the target gene *tryptophanyl-tRNA synthetase mitochondrial* (*TrpRS-m*), the expression of which has been correlated with the expression of genes associated with availability and reuptake of tryptophan (*14*), the precursor to the

neurotransmitter serotonin. Changes in mitochondrial activity are known to regulate neural energetics underlying honey bee aggression (*15*). In addition, serotonin is involved in the regulation of aggression in diverse species of vertebrates and invertebrates (*16*), including honey bees, with serotonin-related genes located in a previously described aggression QTL (*17*). Moreover, serotonin levels in the optic lobe are correlated with the decision to sting and serotonin treatment has been shown to trigger stinging (*18*). The strongest soldier-forager differences in TF-TG relationships for *TrpRS-m* were in cell types that comprise the optic lobes and a population of cells primarily expressing *leukocyte-antigen-related-like* (*Lar*), a receptor tyrosine phosphatase required for neuronal development of the visual system in *D. melanogaster* (*19*). These results suggest that inherited changes in honey bee aggression also are based, in part, on changes in the structure and function of visual pathways that respond to or process aggression-triggering visual stimuli.

The evolution of decreased colony aggression in the Puerto Rican population involved a soft selective sweep (*6*), so it is likely that selection has had multiple influences on the brains of colony members. The following three results suggest selection might have affected the brains of foragers as well as soldiers: 1) genes with expression levels that correlated with colony aggression were significantly enriched for eGenes in foragers, but not soldiers (fig. S5; Data S7); 2) the regulation of *TrpRS-m* expression was significantly stronger in foragers compared to soldiers (Fig. 2D; Data S6); and 3) serotonin-related genes were significantly differentially expressed in the brains of foragers compared to soldiers (fig. S1A; Data S1), and specifically correlated with the expression of *TrypRS-m* in foragers from highly aggressive colonies (fig. S6). Division of labor among honey bee workers is not based on morphology, so variation in individual aggression is in degree and not kind; foragers will join in defending their colony if triggering stimuli exceed their threshold of response, and colony aggression and foraging intensity are inversely correlated (*20*). Perhaps variation in colony genetics affects the propensity of foragers to participate in defending their colony, in addition to affecting the propensity of soldiers to initiate and sustain the group response. This is consistent with the fact that as a prime example of collective behavior, several subgroups within a honey bee colony coordinate to defend against attack.

Our understanding of how genetics influences the evolution of behavior lags relative to morphology (*21*). This poses challenges in both fundamental and applied contexts, as anthropogenic environmental change frequently provokes behavioral responses (*22*). We have shown that decreased colony aggression (which may have been anthropogenic; ref. *6*) stems from evolutionary pressures targeting features of brain cell GRN architecture that distinguish high- and low-aggression colony members, namely changes in specific cell populations and genes associated with olfaction and vision. This genotype-phenotype mapping was not possible with analyses based only on gene expression. The approach used here, which uses scGRNs to integrate population genetics and systems biology, provides tools to better understand the evolution of behavior.

Figures:

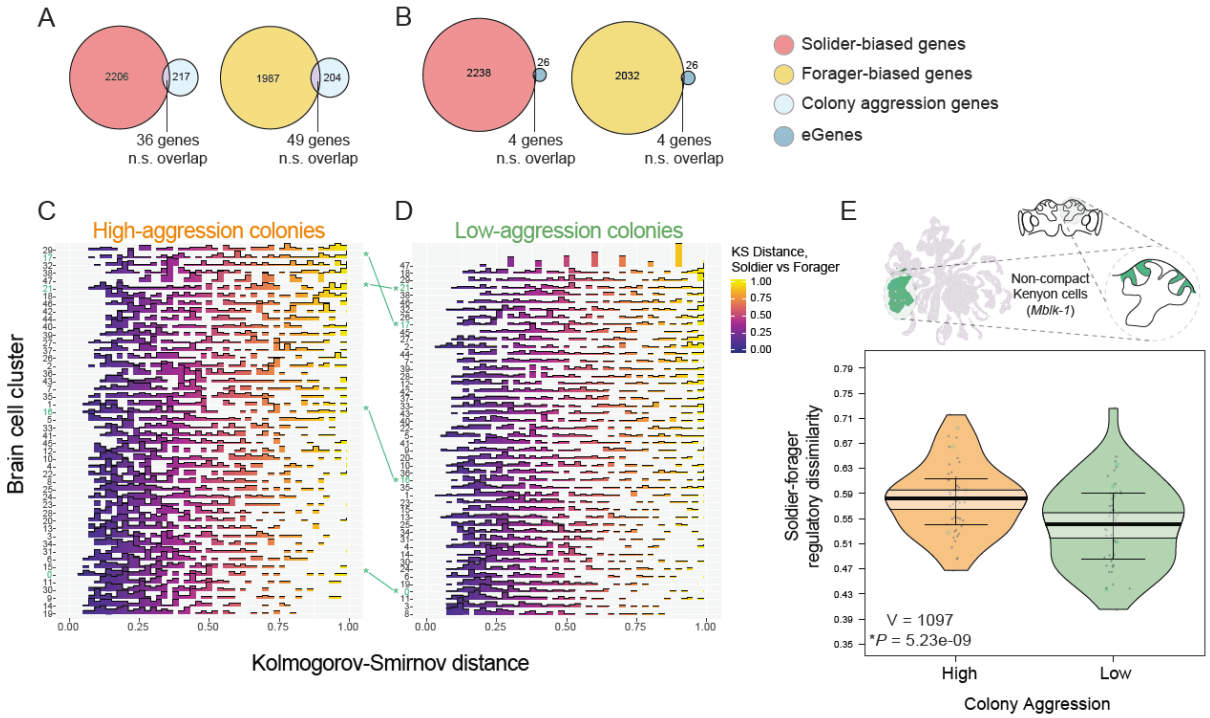


Fig. 1. Dissimilarity of brain gene regulatory network (GRN) architecture between soldiers and foragers is greater in high-aggression colonies. (A) No significant enrichment of genes containing colony aggression-associated SNPs ("colony aggression genes") or (B) of eGenes in soldier or forager-biased differentially expressed genes (proportional Venn diagrams; hypergeometric test, $P > 0.10$). (C-D) By contrast, genetic differences in colony aggression are manifest in brain single-cell GRNs. Distribution of Kolmogorov-Smirnov (KS) regulatory dissimilarity (RD) distances, representing variation in brain GRN architecture between high- (soldiers) and low-aggression (foragers) individuals, showing greater RD between soldiers and foragers from high- compared to low-aggression colonies. Each row represents one of 48 cell clusters from four soldier and four forager scRNA-Seq samples from two high- and two low-aggression colonies. Numbers highlighted in green with asterisks: lines show relative position of specific brain cell clusters as a function of RD in high- and low- aggression colonies. (E) *Above*. UMAP plot and brain diagram provide heuristic example of brain cell populations with greater soldier-forager RD associated with greater colony aggression, highlighting the non-compact Kenyon cells (ncKCs, identified via expression of *Mblk-1*). *Below*. Aggregated data from panels C and D showing significantly greater RD in bees from high- versus low-aggression colonies (Wilcoxon signed-rank exact test). Solid line in center of plots = mean; shaded bands = 95% confidence interval; whiskers = 1st and 3rd quartiles; and raw data shown as points with surrounding smoothed density curve. Each point is derived from a row in corresponding panels C and D. Larger green points represent cell clusters belonging to the ncKCs.

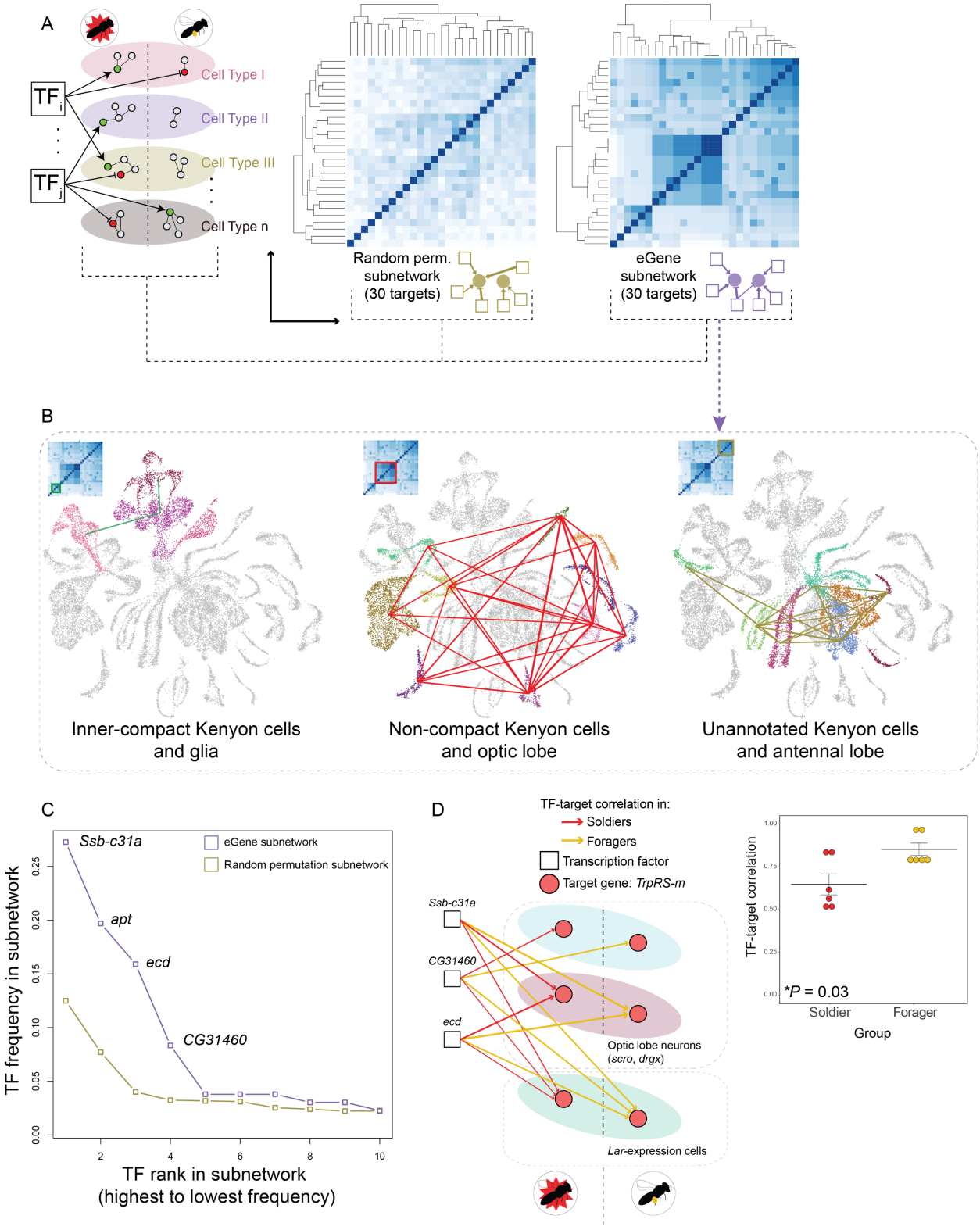


Fig. 2. Evolutionary changes in honey bee colony aggression are associated with changes in brain cell gene regulatory network (GRN) architecture between soldiers and foragers. (A)

Subnetworks composed of either eGenes or random subsets of genes (negative control) as TGs were extracted from an inferred scGRN (Methods). Heatmaps comparing these subnetworks across sc clusters reveal negligible similarities for the control subnetworks (representative example from permutation analysis shown here) but highly similar architectures for the eGene subnetworks. Darker spots represent more similar sets of TF-TG relationships across cell clusters, as determined by Jaccard similarity index. (B) Results from (A) superimposed on UMAP plots in which line thickness is proportional to the similarity of subnetwork architecture between two connected sc clusters. For visualization purposes, only edges with similarity ≥ 0.60 are shown. (C) Analysis of TF centrality, as measured by frequency of a given TF, revealed four TFs overrepresented in eGene subnetworks relative to control subnetworks. (D) Strong regulatory plasticity of the TG *TrpRS-m* between soldiers in foragers in three cell clusters identified as part of the honey bee optic lobe; line thickness corresponds to strength of TF-TG relationship. Inset shows that the strengths of these regulatory relationships are significantly higher in foragers compared to soldiers (Wilcoxon). Sample sizes as in Fig. 1.

Acknowledgements: We thank A. Hernandez and C. Wright at the Roy J. Carver Biotechnology Center (UIUC) for sequencing services, and Marla Sokolowski, members of the Gene Networks in Neural and Developmental Plasticity theme (IGB), and the Robinson lab for valuable comments and feedback on the manuscript.

Funding:
Illinois Sociogenomics Initiative (GER). Genome sequencing from ref. (*7*) funded by a Lundbeckfonden grant (G. Zhang).

Author contributions:
Conceptualization: IMT, AA, GER
Methodology: IMT, SAB, PD, GS, ACA
Investigation: IMT, ACA, AS
Visualization: IMT, PD, GS
Funding acquisition: GER
Project administration: GER
Supervision: MH, SDZ, SS, JC, GER
Writing – original draft: IMT, GER
Writing – review & editing: IMT, SAB, PD, GS, AA, ACA, MH, SS, SDZ, JC, GER

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Whole-tissue and single-cell RNA-Sequencing with relevant code and analysis pipeline will be made publicly available upon publication.

Supplementary Materials for

Evolution of decreased aggression in honey bees targeted brain cell gene regulatory networks

Ian M. Traniello, Syed Abbas Bukhari, Payam Dibaeinia, Guillermo Serrano, Arian Avalos, Amy Cash Ahmed, Alison Sankey, Mikel Hernaez, Saurabh Sinha, Sihai Dave Zhao, Julian Catchen, Gene E. Robinson

Correspondence to: it4770@princeton.edu (IMT); generobi@illinois.edu (GER)

This section includes:

Materials and Methods
Figs. S1 to S6
Captions for Data S1 to S7

Other Supplementary Materials for this manuscript include the following:

Data S1 to S7

Data S1: Cluster-defining differentially expressed genes and cell cluster annotations

Data S2: Soldier- and forager-biased genes differentially expressed in pairwise comparison

Data S3: Cluster-specific differentially expressed genes (soldiers compared to foragers)

Data S4: Results from eQTL analysis

Data S5: TF-target interactions in eGene GRN subnetwork

Data S6: Fifteen TF-target interactions that strongly covary with behavioral state

Data S7: Genes with expression levels that correlate with colony aggression (in all individuals and resolved to soldiers or foragers)

Materials and Methods

"Gentle" African honey bee (gAHB) soldier and forager collections

Collection strategies for gAHB soldiers and foragers used for the previously published whole genome sequences are described in (*7*). Briefly, nine colonies from eight locations were sampled across Puerto Rico in April 2017, and all colonies were maintained in standard Langstroth hives containing a single hive box with 16,000 – 20,000 bees and one naturally mated queen. Colonies were transported to the Gurabo Agricultural Field Station at the University of Puerto Rico, spaced > 1 m apart, and allowed to rest undisturbed for 2 weeks to acclimate to the new conditions. Queen presence was determined before and after the phenotyping.

A previously described assay for collecting soldiers and foragers was deployed for individual phenotyping (*23*, *24*). A string of nine 5 x 5 cm suede leather patches was positioned 21 cm from the colony entrance, and the colony was rhythmically struck with a cinder block for 40 times over 2 min. Bees that rushed out to sting the patches over the next 2 min ("soldiers") were collected and flash-frozen in liquid nitrogen, after which the leather patches were removed and placed in a plastic bag for further analysis. The hive entrance was then restricted and bees that left for the next 30 s were dusted with talcum powder. Next, the hive entrance was opened, and the colony was left for 30 min, after which the entrance was sealed again and talcum powder-dusted bees returning to the colony with a pollen load ("foragers") were collected and flash-frozen. This strategy allowed for the selective collection of foragers that were in the colony during the disturbance but did not respond to it and instead initiated foraging. Frozen samples were stored at the University of Puerto Rico, Rio Piedras, before being transferred to the University of Illinois at Urbana-Champaign (UIUC) for DNA (*7*) and RNA processing, the latter of which is presented here.

Stings in the black suede leather patches were counted to measure the intensity of each colony's response. Two weeks later, colonies were scored again with an established behavioral ranking system (*23*, *25*), which assigns a colony rank score on a scale from one to four based on observations of bees (1) running on the comb, (2) hanging from the comb, (3) flying around the hive, and (4) stinging investigators. Behaviors "1" and "2" are markers of general activity, "3" is associated with stronger arousal, and "4" is a measure of colony aggression during a lower level of disturbance than previously administered with the colony disturbance assay. Colonies were assigned a cumulative rank score from 1-9 following criteria from each assay.

A multidimensional scaling (MDS) analysis was then applied to generate a unified colony phenotype from the combined colony measures. The first dimension generated by the MDS approach was highly correlated with colony rank across the two measures, and we used either this dimensional vector or number of stings in the leather patches as a surrogate for colony-level aggressive phenotype, as noted. Full experimental details including colony scores, ranking, and MDS dimensional vectors are available in (*7*).

Bulk RNA-Sequencing on gAHB samples

We used gAHB soldier and forager samples from Puerto Rico that had been previously collected for whole-genome sequencing (*7*), as described above. Heads from individual bees were removed and freeze-dried for 52 min at -80°C before the brain was dissected in a dry ice/ethanol bath. Total brain RNA was then extracted using the RNeasy Mini Kit (QIAGEN, Hilden, Germany) and eluted in 30 ul of RNase-free water. Quality and quantity of RNA was assessed on a NanoDrop 1000 (Thermo Fisher, Waltham, MA), Qubit 2.0 fluorometer (Thermo Fisher), and Bioanalyzer (Agilent

370 Technologies, Santa Clara, CA). RNA-Sequencing (RNA-Seq) libraries were prepared on two
371 plates with the TruSeq Stranded mRNAseq Sample Prep Kit, and libraries were quantitated via
372 qPCR. We performed sequencing on one lane of a NovaSeq 6000 with the NovaSeq S4 reagent
373 kit for 151 cycles from each end of the fragments. Reads were 150 nucleotides in length. We
374 demultiplexed FastQ files with the bcl2fastq Conversion Software v2.20 (all Illumina unless
375 otherwise stated).
376
377 *European honey bee (EHB) soldier and forager collections*
378
379 For collections of EHB soldiers and foragers in Urbana, IL, we used the same approach outlined
380 in (*7*), as detailed above, which describes the gAHB samples referenced above and the bulk brain
381 RNA-Seq analysis used in this study. All EHB samples were used for single-cell RNA-Seq
382 (scRNA-Seq).
383
384 All colonies were maintained according to standard beekeeping practices at the Bee Research
385 Facility at UIUC. We relocated four standard Langstroth hives containing a colony of ~30,000
386 workers headed by a naturally mated queen to a common yard to control for environmental factors
387 like floral resource access and exposure to ambient conditions. Hives were equal in size, containing
388 two stacked hive boxes with identical amounts of honeycomb and were left undisturbed for at least
389 10 days after relocation and prior to phenotyping. We collected soldiers and foragers from one
390 colony per day across four successive days in August 2019 (n = 4 soldier and 4 forager samples).
391 On each collection day, we began at 10:00AM by performing the colony disturbance / stinging
392 assay as described above. However, instead of being flash-frozen, samples were rapidly
393 transported indoors and processed for scRNA-Seq.
394
395 *Single-cell RNA-Sequencing (scRNA-Seq)*
396
397 Soldier and forager brains were dissected in ice-cold Dulbecco's PBS (DPBS, cat. no. #14190144,
398 Thermo Fisher Scientific) in a sterile, RNase-free space. We then followed the "short dissociation
399 protocol" from (*26*), which allowed us to go from sample collection in the field to cells in
400 suspension within minutes. Briefly, we cleaned away glands, trachea, and any connective tissue
401 and cut each freshly dissected brain into six pieces, leaving two optic lobes and the central brain
402 divided into four quadrants. We were careful to only cut through neuropil, thus leaving neurons
403 and neuron-dense regions intact. Brain pieces were transferred to a sterile centrifuge tube
404 containing 500 ul of DPBS on ice. After this was repeated for five brains per soldier or forager
405 sample, we allowed all tissue to sink to the bottom of the tube and replaced DPBS with 200 ul of
406 a diluted Trypsin-EDTA solution (0.05%, cat. no. 25300054, Invitrogen, Carlsbad, CA) and
407 transferred the centrifuge tube to a thermomixer. Samples were mixed at 1000 rpm and 25°C for
408 15 min with gentle manual trituration with an ordinary pipette tip every 5 min. We then removed
409 residual tissue from the bottom of the tube with a wide-bore pipette tip and centrifuged cells at 400
410 rcf and 4°C. We removed the supernatant and washed the cells twice in 1 ml ice-cold DPS with a
411 centrifugation step in between washes. After the second wash, we performed a final centrifugation
412 step and resuspended cells in 400 ul DPBS with 0.04% UltraPure™ Bovine Serum Albumin
413 (AM2616, Thermo Fisher Scientific). We then filtered cells with a 40 Flowmi™ Tip Strainer
414 (H136800040, Bel-Art, Wayne, NJ). Cell suspensions were stored on ice and immediately
415 transported to the Roy J. Carver Biotechnology Center at UIUC, where they were subjected to a

total cell count and % viability assay with AOPI staining on a Nexcelom Cellometer Spectrum Image Cytometry System (Nexcelom Biosciences, Lawrence, MA USA). Viability was consistently > 97% for each sample.

Libraries were constructed with the Chromium Single-Cell V3 kits (10x Genomics, Pleasanton, CA) and the library pool was amplified via qPCR and loaded on one lane of a NovaSeq 6000 S4 flowcell (Illumina, San Diego, CA), where it was sequenced for 28 nt from one end and 151 nt from the other end. FastQ files were generated and demultiplexed with the CellRanger v3.0.2 "mkfastq" command. The pool of libraries produced over 5 billion reads, and excellent quality was observed for each sample via fastQC.

Within a few days after collections, we inspected all four colonies and confirmed the presence of a laying queen, meaning each colony was queenright at the time of sample collection.

Bioinformatics

Bulk RNA-Seq

We removed 12 samples from analysis due to contamination with a common honey bee virus (*9*, *27*) and two samples with an abnormally low number of sequenced reads, leaving us with a final sample size of 163 individuals (81 soldiers and 82 foragers) collected from nine colonies. We identified and removed a batch effect from library preparation plate using Combat-Seq (*28*) before performing differential gene expression analysis with edgeR (*29*). After filtering to remove transcripts that had fewer than one count-per-million in at least the number of samples equal to the smallest group size, we performed TMM normalization on a gene list universe of 9185 genes for pairwise soldier versus forager comparisons and 9827 for an analysis of genes correlated with colony aggression in soldiers or foragers, the latter being larger on account of smaller group sizes and therefore a more relaxed filtering. We used a generalized linear model (GLM) to compare gene expression between soldiers and foragers using colony as a blocking factor. We next independently analyzed soldiers and foragers to identify genes positively or negatively associated with aggression by regressing individual-level gene expression on each individual's colony aggression value, derived from the MDS dimensional vector from (*7*), as a continuous predictor of gene expression. To correct for multiple tests, we performed the Benjamini-Hochberg (BH) correction (*30*) and filtered differentially expressed gene (DEG) lists based on a false discovery rate (FDR) threshold of 0.05. For Gene Ontology (GO) enrichment analysis (*31*), we first converted soldier- or forager-biased honey bee genes to *Drosophila melanogaster* orthologues using a one-to-one reciprocal best hit BLAST (RBH-BLAST). Enrichment of terms associated with Biological Processes was performed with GOrilla (*32*) and visualized with REVIGO (*33*).
Gene list overlap analysis was performed via hypergeometric tests with BH-FDR correction for multiple tests using the GeneOverlap package (*34*). This approach was also used to compare bulk and single-cell data. We directly compared gene lists associated with behavioral state, colony aggression level, or cell cluster (described below) with a list of 253 genes that contained aggression-associated single-nucleotide polymorphisms (SNPs); 60 of these genes also showed significant signatures of selection, meaning their frequency was found to be increased in the gAHB population. We refer to these 253 genes as "colony aggression genes," and the complete list can be found in Dataset S1 of (*7*).

Expression quantitative trait loci (eQTL) mapping

To test for an association between genotype and gene expression, we accessed the genotype calls made on the same individuals as used in this study from (*7*). eQTL-SNPs (eSNPs) with variance greater than zero were retained for the eQTL analysis, and a distance of 10kb was used to define local SNP-gene pairs. A linear model was fitted in Matrix eQTL (*35*) to scan for genome-wide *cis*-eQTLs for 9185 genes and 3309291 SNPs separately for soldiers and foragers. Individual relatedness and different colonies were blocked in the linear model and an FDR cutoff of 0.05 was used to select the *cis*-eQTLs. Code for this analysis can be found at https://github.com/bukhariabbas/Trannielloetal_eQTLs.

This analysis identified 3935 and 3473 eQTL-genes in soldiers and foragers, respectively. To identify variants most likely to be associated with both group- and individual-level aggression, we short-listed the eQTL-genes by intersecting eQTL variants in both soldiers and foragers with the variants associated with colony aggression in (*7*). This resulted in 30 eGenes which were used for downstream analysis. We tested for overlap enrichment in the bulk and sc transcriptomics experiments, and the 30 eGenes were also used as TGs in the gene regulatory network analyses described below.

Single-cell RNA-Sequencing

scRNA-Seq data were aligned, filtered, and quantitated using the CellRanger "count" command. To assess cell quality, we considered the expression of both mtDNA as well as mitochondrially localized proteins, as previous studies have suggested both may be associated with cell death or dysfunction (*36–38*). We detected exceptionally low levels of each (combined averages: ~2 ± 1%) which, when taken with our viability analysis, indicates that the vast majority of sequenced cells were live and yielded high quality data.

Downstream analyses were performed with Seurat (*39*).We required each gene to be detected in at least 3 cells and each cell to contain at least 200 genes. Data were log normalized via the *NormalizeData* command with default settings, and the mean-variance relationship was controlled for by running *FindVariableFeatures* with default parameters. To correct for batch effects based on day of sequencing, we used the "IntegrateData" function (*39*) which simultaneously controlled for variation due to colony of origin, as each of the four sequencing days was dedicated to a different colony. We then constructed a Shared Nearest Neighbor graph and applied the Louvain community detection algorithm to cluster cells before visualizing data with Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP). UMAP plots were generated using the top 50 dimensions from a principal component analysis and "min.dist" set to 0.75 to evenly distribute cells for visualization. Clustering resolution was set to the default value of 0.8. These parameters were selected based on previous scRNA-Seq data in *D. melanogaster* (*26*), jumping ants (*10*), and a previous honey bee study (*9*), with consideration for the high degree of cellular homogeneity in insect whole-brain samples as well as the large recovery of cells in our experiment. This yielded a total of 48 cell clusters, which were used for cell type-specific soldier-forager differential gene expression testing as well as the regulatory dissimilarity analysis, presented below.

508
509 We detected DEGs for each cluster using the MAST algorithm with default settings (*40*).We
510 performed two DEG analyses to determine 1) genes that defined each cluster (independent of
511 behavioral state), 2) genes that differentiated soldiers and foragers in each cluster. Cluster-defining
512 genes were used to annotate the 48 detected clusters, and cluster-specific soldier-forager DEGs
513 were used for overlap analyses with colony aggression genes and eGenes.
514
515 We annotated the majority of the 48 detected cell clusters using established insect neuronal and
516 glial biomarkers (*26*, *41*, *42*). Utilizing available insect scRNA-seq brain atlases of gene expression
517 (*10*, *26*, *42–44*), including a previous honey bee brain scRNA-Seq analysis (*9*), we annotated cell
518 clusters as populations of cells that included: neurons that make up the Pm (*Lim3*), T4/T5 (*dorsal*
519 *root ganglia homeobox*), and Mi1 (co-expression of *homeobox protein homothorax* and *brain-*
520 *specific homeobox*) neurons, which span the lobula, medulla and lamina of the optic lobe (*26*);
521 lateral (*venrtal veins lacking*) and antennodorsal (*abnormal chemosensory jump 6*) projection
522 neurons of the antennal lobe (*44*), as well as other cell types. The mushroom bodies (MB), a higher-
523 order multimodal sensory integration center important for learning and memory, navigation, and
524 social behavior, formed a clearly segregated set of clusters. Inner-compact Kenyon cells (KCs)
525 were identified by co-expression of *E74* and *ecdysone receptor*. Non-compact KCs were identified
526 by co-expression of *Mblk-1* and *dlg5* (*45*). Other KC populations detected via expression of
527 *mushroom body-expressed* could not be further annotated, which is expected considering limited
528 knowledge of cellular diversity in the MB. We also found that roughly 88% of cells expressed the
529 neuronal marker *fne* while an average of ~11% of cells expressed insect glial markers *repo* and
530 *borderless*, a neuron:glia ratio consistent with previous reports from other insects (*10*, *26*). Within
531 the glial populations, we identified most known insect glia subtypes (*46*) including astrocyte-like
532 glia, perineurial glia, subperineurial glia, and cortex glia, as well as a small population of
533 hemocytes (*hml*) that clustered close to the glial populations (*26*, *47*).
534
535 *Gene regulatory network analyses: regulatory dissimilarity as a function of variation in colony*
536 *aggression*
537
538 To infer single-cell gene regulatory networks (GRNs), we used SimiC, which has been previously
539 shown to perform well with honey bee brain scRNA-Seq data and is better optimized for
540 comparing GRNs across behavioral states than other established methods, to which it has been
541 benchmarked (*48*). SimiC imputes scRNA-Seq expression data to infer the relationship between a
542 transcription factor (TF) and its predicted TGs. As input, we used a list of 193 honey bee TFs, each
543 with an ortholog to a biochemically or genetically validated TF in *D. melanogaster*, which was
544 generated by intersecting TFs predicted by ASTRIX in three previous honey bee brain
545 transcriptomics studies (*4*, *27*, *49*). We then used SimiC to output a set of incidence matrices
546 containing TF-by-TG (TF x TG) information representative of soldier and forager behavioral
547 states. The GRN evaluation pipeline was based on a weighted area under the curve ("wAUC")
548 score. For each cell, the wAUC score quantified the relative activity of a given regulon (i.e. a row
549 of the GRN incidence matrix, inferred in the previous step) with respect to the expression of all
550 TGs. For any given TF, if the overall expression of its associated TGs in a cell was high and
551 weights between the TF and its TGs were large, then we concluded that the TF in question has a
552 large influence on the expression profile of a particular cell. For all clusters of cells and TFs, the
553 wAUC was computed independently for each cell yielding a distribution of cell-specific wAUC

554 values corresponding to a behavioral state (soldier or forager). To measure the separation of
555 distributions, we calculated the Kolmogorov-Smirnov (KS) distance between distributions of
556 wAUC values for each cell within a cluster. This yielded one value for each TF per cluster,
557 allowing us to describe "dissimilarities" in soldier and forager regulatory relationships for each
558 predicted cell cluster. Finally, for each cluster, we aggregated the KS distance of each TF into one
559 unique value by computing the area under the curve of the empirical cumulative distribution
560 function (CDF). This approach allowed us to make general comparisons of all cell clusters sampled
561 from soldiers and foragers as well as examine GRN differences specific to the aggression level of
562 the colony of origin. Code for this analysis can be found at
563 https://github.com/jianhao2016/SimiC/blob/master/Tutorial/SimiC_post.R.
564
565 *Gene regulatory network analyses: architecture of subnetworks associated with colony aggression*
566 *genes*
567
568 To infer regulatory subnetworks, we applied the GENIE3 (*50*) pipeline to both our bulk and sc
569 datasets, using the same list of 193 TFs as the previous step. For soldiers and foragers, we inferred
570 separate GRNs by collecting the top ***N*** regulators (sorted by GENIE3) having a "≥ $\mathbf{C_1}$" absolute
571 Spearman correlation with the TG's expression. Next, we compared soldier and forager GRNs to
572 obtain interactions that are present in both groups with the same TF-TG correlation directionality;
573 this reduced the likelihood of identifying variation due to under-sampling in either soldier or
574 forager groups. We repeated this analysis for bulk and sc expression data. For the sc analysis, we
575 first imputed the expression matrix using MAGIC (*51*) with diffusion ("t") = 2 and split the results
576 into 26 matrices, one for each cell cluster (clustering resolution was set 0.2 in Seurat, with all other
577 parameters unchanged, in order to obtain enough cells per cluster for the analysis). For bulk data,
578 we used $\boldsymbol{N} = 10$, $\mathbf{C_1} = 0.5$, and for sc analysis we used $\boldsymbol{N} = 20$, $\mathbf{C_1} = 0.25$. Next, we specifically
579 extracted TF-TG relationships in which TGs were either the 30 eGenes (described above) or 15
580 permutations of 30 random eQTL-genes that were also identified in the eQTL analysis but were
581 not the 30 eGenes (control).
582
583 We then assessed similarity of GRN architecture across cell types in eGene and control
584 subnetworks. First, we extracted from each cell cluster a binary vector in which each element
585 corresponded to a TF-TG interaction and compared these vectors across all cell clusters, using
586 Jaccard index as a measure of closeness. Matrices containing Jaccard indices from pairwise
587 comparisons made across cell clusters for eGene or control subnetworks in soldiers and foragers
588 were hierarchically clustered using complete-linkage clustering of the Euclidean distance, and
589 dendrograms for the eGene and control subnetworks were cut at a height of 1.2. After cutting, we
590 specifically explored connectivity of "blocks" (i.e. groups containing more than three cell clusters
591 sharing similar GRN subnetwork architecture based on the above parameters); we identified three
592 blocks from the eGene subnetwork and no blocks from the control subnetwork.
593
594 The strength of connectivity (i.e. similarity of subnetwork architecture across cell clusters) was
595 assessed by assigning a score to each block in the eGene subnetwork representing the smallest
596 pairwise connectivity score observed between clusters within a block. This score was then
597 computed in each of the 15 permutation subnetworks using the same cell clusters identified in each
598 respective block in the eGene subnetwork. The minimum connectivity scores identified in the

eGene subnetworks were considerably higher than scores generated in each of the permutation networks.

TF rank was computed by dividing each TF's appearance in a given subnetwork by the total number of interactions in the subnetwork. TF-TG correlations that differed ≥ 0.1 between soldiers and foragers represented ~10% of all interactions in the eGene subnetwork and were therefore considered to be the strongest differences.

Supplementary Figures

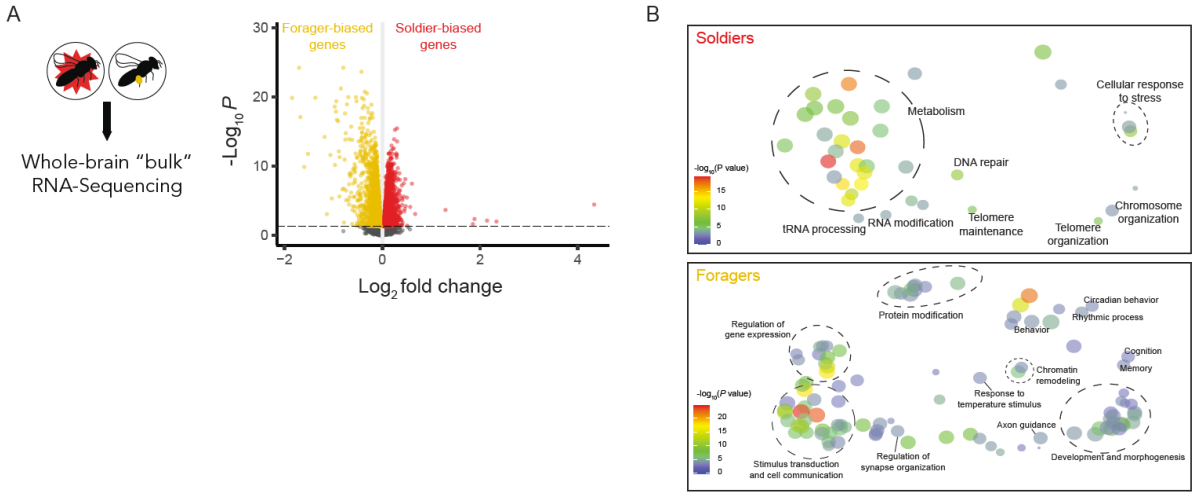


Fig. S1. Whole-tissue brain transcriptomics of soldiers and foragers. (A) Whole-brain RNA-Sequencing of the same soldier and forager bees from (*7*) revealed thousands of differentially expressed genes (DEGs) associated with behavioral state, termed "soldier-" or "forager-biased" based on fold-change following a pairwise comparison between groups. Red or yellow dots represent genes upregulated in soldiers or foragers, respectively, that passed a false discovery rate (FDR)-corrected *P* value threshold of 0.05; grey dots represent tested genes that did not pass this threshold. A total of 4278 DEGs were identified, with 2242 and 2036 that were soldier- or forager-biased, respectively (Data S1). (B) REVIGO plots show Biological Process terms identified by a Gene Ontology (GO) enrichment analysis associated with soldier- or forager-biased gene lists. Circles are organized in semantic space such that more similar terms in the GO hierarchy are positioned more closely together. Circle diameter is inversely correlated with the specificity of each GO term, such that smaller circles represent more specific terms.

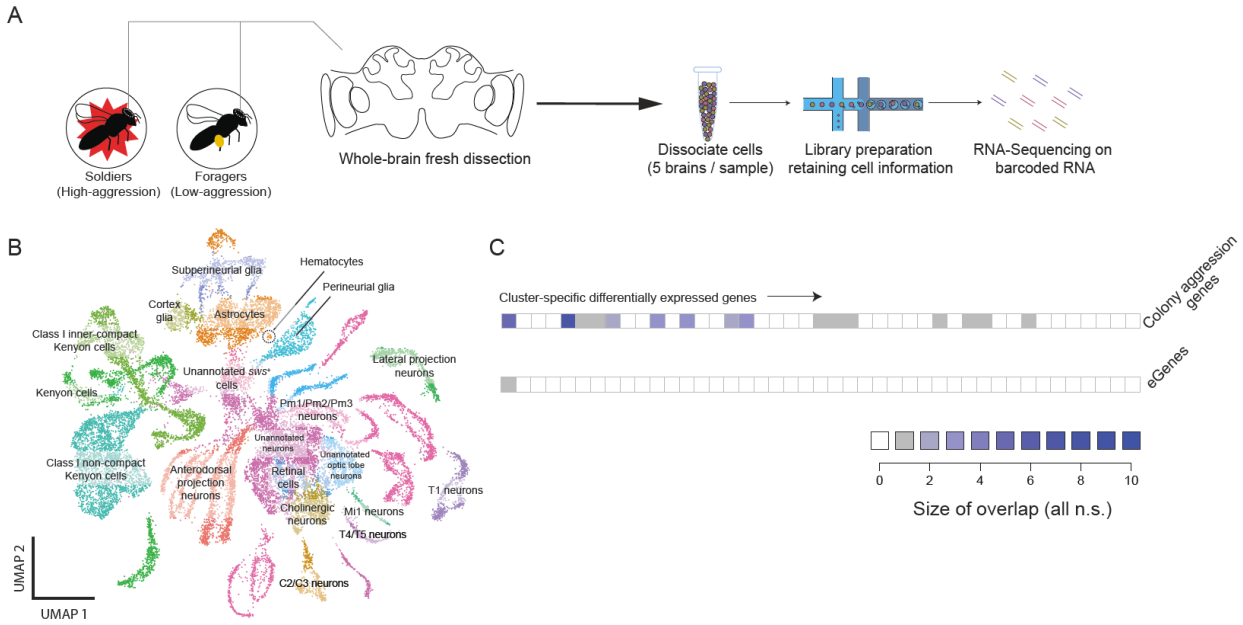


Fig. S2. Single-cell brain transcriptomics of honey bee soldiers and foragers. (A) Workflow for transcriptomic profiling of ~24,000 cells from the brains of field-collected soldiers and foragers and (B) bioinformatic clustering of cells in reduced dimensionality (UMAP) space. Clusters were manually annotated using cell type-specific markers as described in Methods. (C) Cluster-specific differentially expressed genes between soldiers and foragers did not significantly overlap with colony aggression genes from (*7*) or eGenes (hypergeometric test, FDR > 0.10 for all comparisons). Color represents size of overlap for each cluster; clusters with no soldier-forager DEGs not shown.

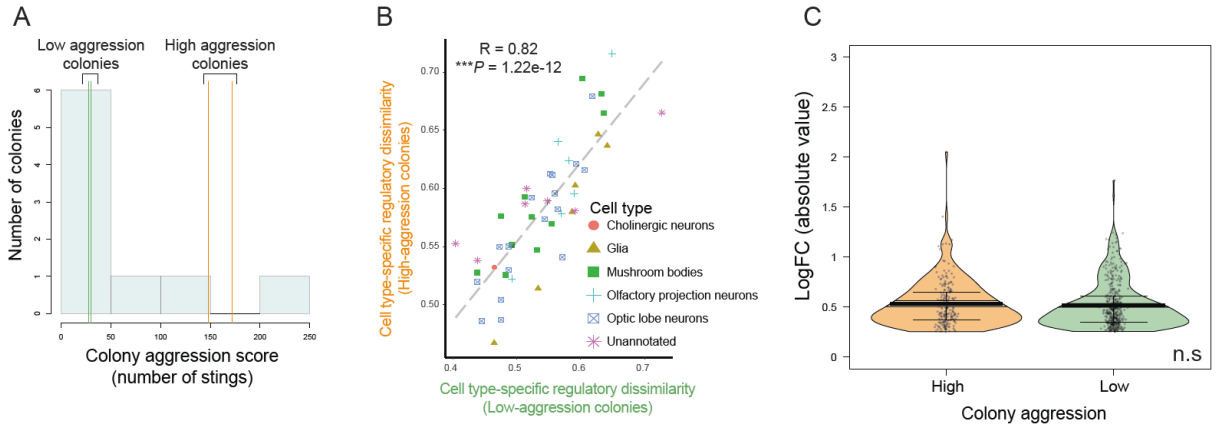
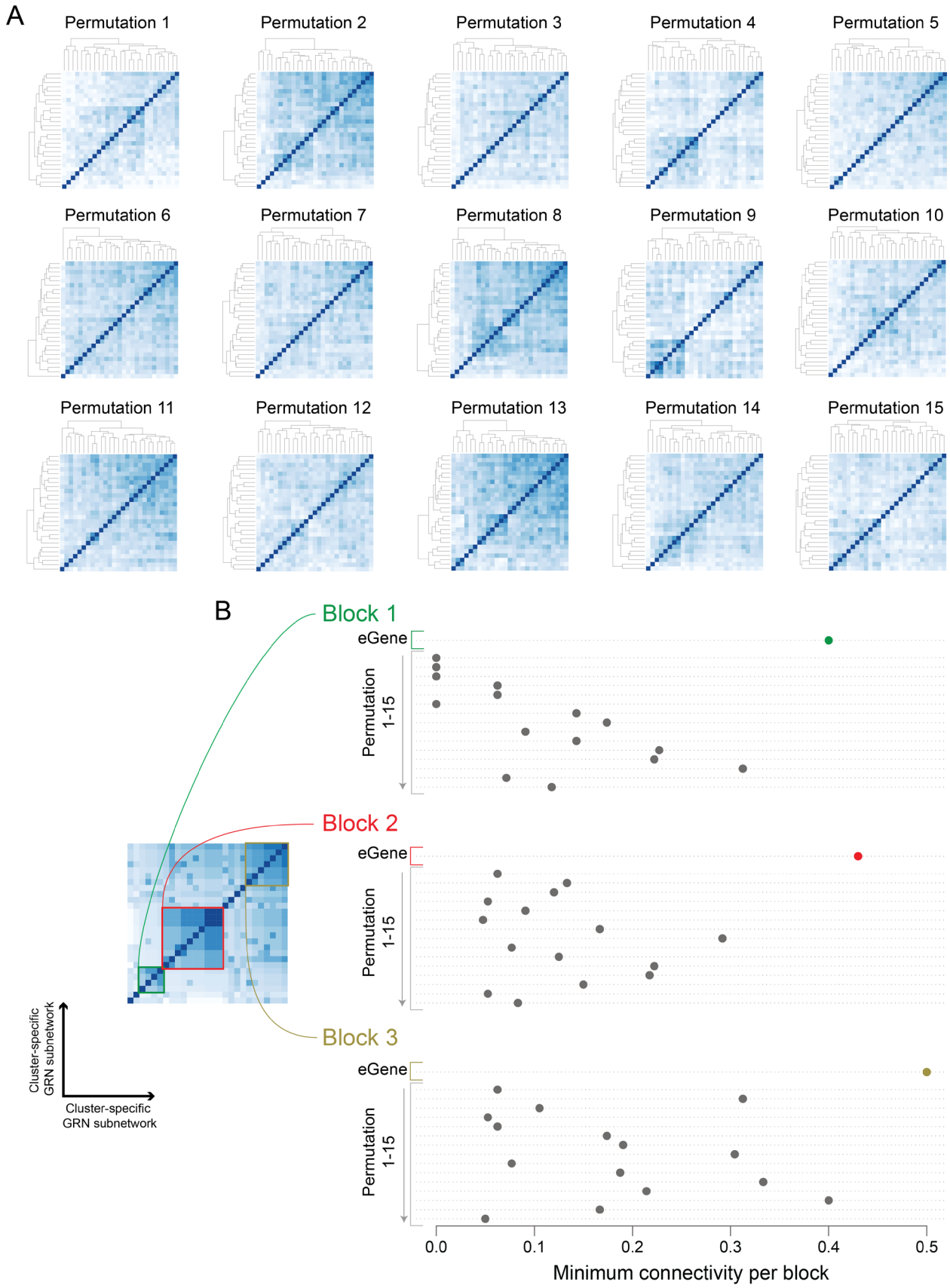
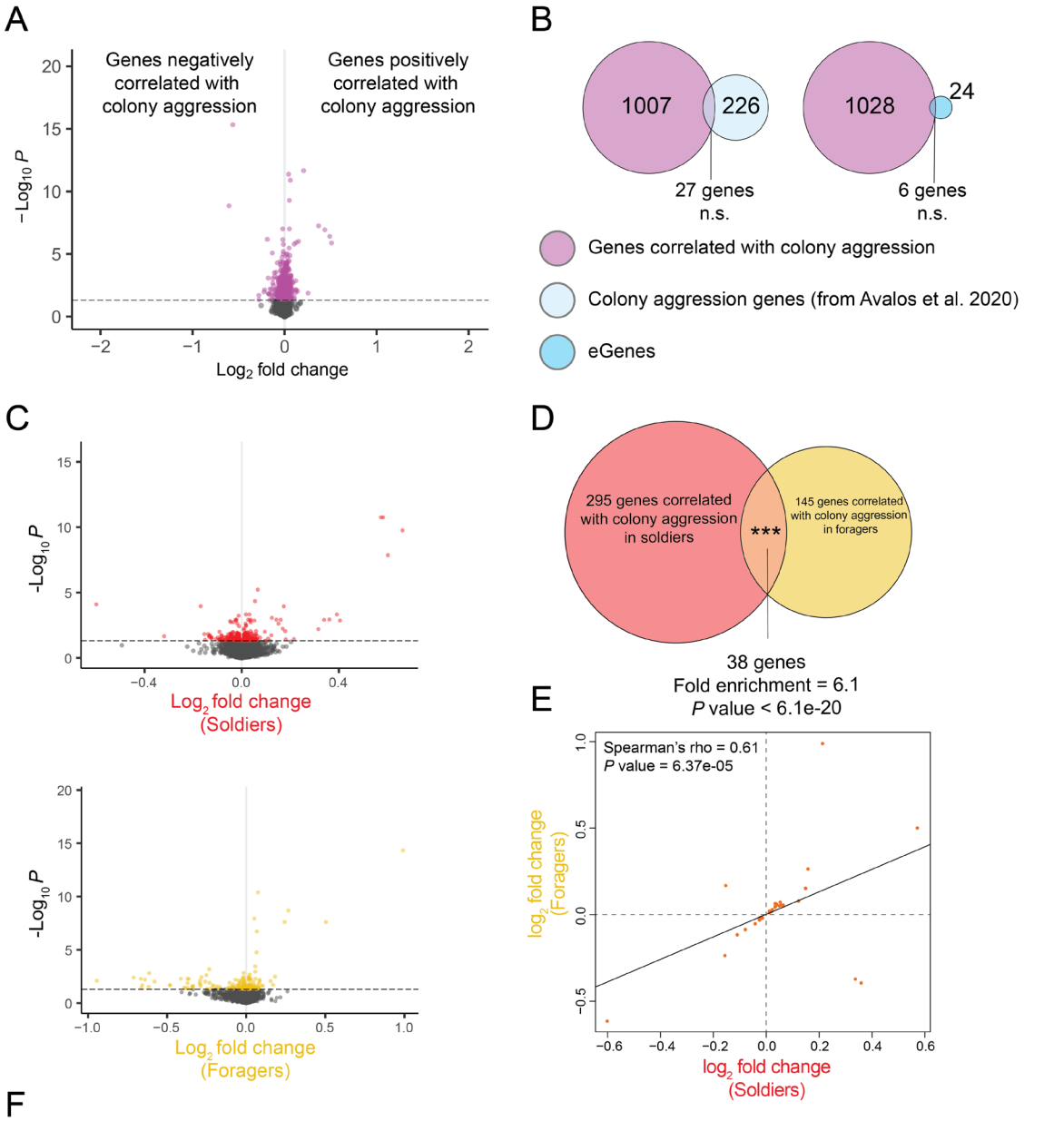


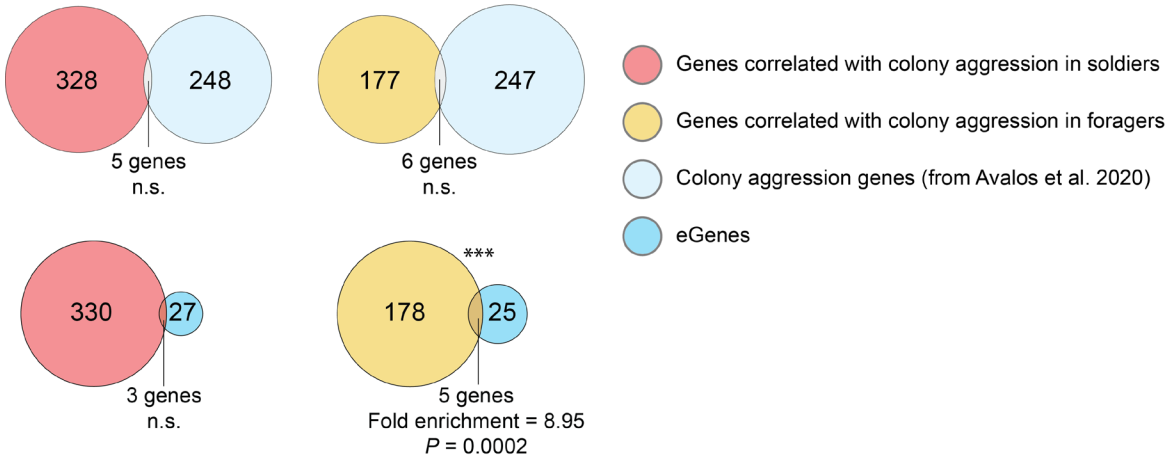
Fig. S3. Single-cell RNA-Sequencing reveals that regulatory dissimilarity, but not differential expression, changes across soldiers and foragers as a function of colony aggression. (A) We divided our scRNA-Seq comparisons of soldiers and foragers by colony aggression level, considering colonies in the top 15% or bottom 25% of the aggression distribution reported in (*7*) to be "high-" (orange vertical lines) and "low-aggression" (green vertical lines) colonies, respectively, with soldier and forager samples from two colonies in each aggression category. X-axis refers to number of stings delivered by resident aggressors to small leather patches placed in front of the entrance during colony disturbance (Methods). (B) The rank-order of cell type-specific regulatory dissimilarity (RD) scores, as described in Fig. 1E, was preserved across high- and low-aggression colonies (Pearson's product-moment correlation). (C) Unlike for RD, as shown in Fig. 2A, there was no difference in the fold-change magnitude of cell type-specific differentially expressed genes between soldiers and foragers from high- and low-aggression colonies (Wilcoxon rank-sum test with continuity correction, $P > 0.10$); gene regulatory relationships, rather than expression alone, may therefore be more appropriate for linking group and individual at the molecular level.


A
Permutation 1
Permutation 2
Permutation 3
Permutation 4
Permutation 5
Permutation 6
Permutation 7
Permutation 8
Permutation 9
Permutation 10
Permutation 11
Permutation 12
Permutation 13
Permutation 14
Permutation 15
B
Block 1
Block 2
Block 3
eGene
Permutation 1-15
Cluster-specific GRN subnetwork
Cluster-specific GRN subnetwork
0.0
0.1
0.2
0.3
0.4
0.5
Minimum connectivity per block

742

743
744 **Fig. S4. Random permutation subnetworks are less connected than eGene subnetwork** (A).
745 Heatmaps as described in Fig. 2 for each of the 15 random permutation subnetworks. (B). Using
746 the three "blocks" of sc clusters identified in Fig. 2, dot charts compare smallest ("minimum")
747 pairwise connectivity for eGene or 15 control subnetworks. For each block, the minimum
748 connectivity was much higher in the eGene subnetwork compared to the 15 controls.
749


A
Genes negatively correlated with colony aggression
Genes positively correlated with colony aggression
−Log10 P
Log2 fold change
B
1007
226
27 genes n.s.
1028
24
6 genes n.s.
Genes correlated with colony aggression
Colony aggression genes (from Avalos et al. 2020)
eGenes
C
-Log10 P
Log2 fold change (Soldiers)
Log2 fold change (Foragers)
D
295 genes correlated with colony aggression in soldiers
***
145 genes correlated with colony aggression in foragers
38 genes
Fold enrichment = 6.1
P value < 6.1e-20
E
Spearman's rho = 0.61
P value = 6.37e-05
log2 fold change (Foragers)
log2 fold change (Soldiers)


F
328
248
5 genes n.s.
177
247
6 genes n.s.
330
27
3 genes n.s.
***
178
25
5 genes
Fold enrichment = 8.95
P = 0.0002
Genes correlated with colony aggression in soldiers
Genes correlated with colony aggression in foragers
Colony aggression genes (from Avalos et al. 2020)
eGenes

750

751
752 **Fig. S5. Molecular correlates of colony aggression identified in soldier and forager whole-**
753 **tissue brain transcriptomics.** (A) We regressed brain gene expression on colony-level aggression
754 for each individual soldier and forager in a given colony, identifying 1037 genes correlated with
755 colony aggression. (B) None of these genes were found to significantly overlap with colony
756 aggression genes from (*7*) or eGenes identified in this study. (C) Going deeper, we performed the
757 same analysis independently in soldiers and foragers (fold-change on volcano plots as described
758 in [A]) and (D) identified a significantly overlapping set of genes that were (E) highly concordant
759 in terms of differential expression associated with colony aggression. (F) Performing an overlap
760 analysis now resolved to the level of soldiers and foragers identified a significant overlap between
761 genes correlated with colony aggression in foragers and eGenes; no other comparison was
762 significant (hypergeometric test, $P > 0.1$).
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778

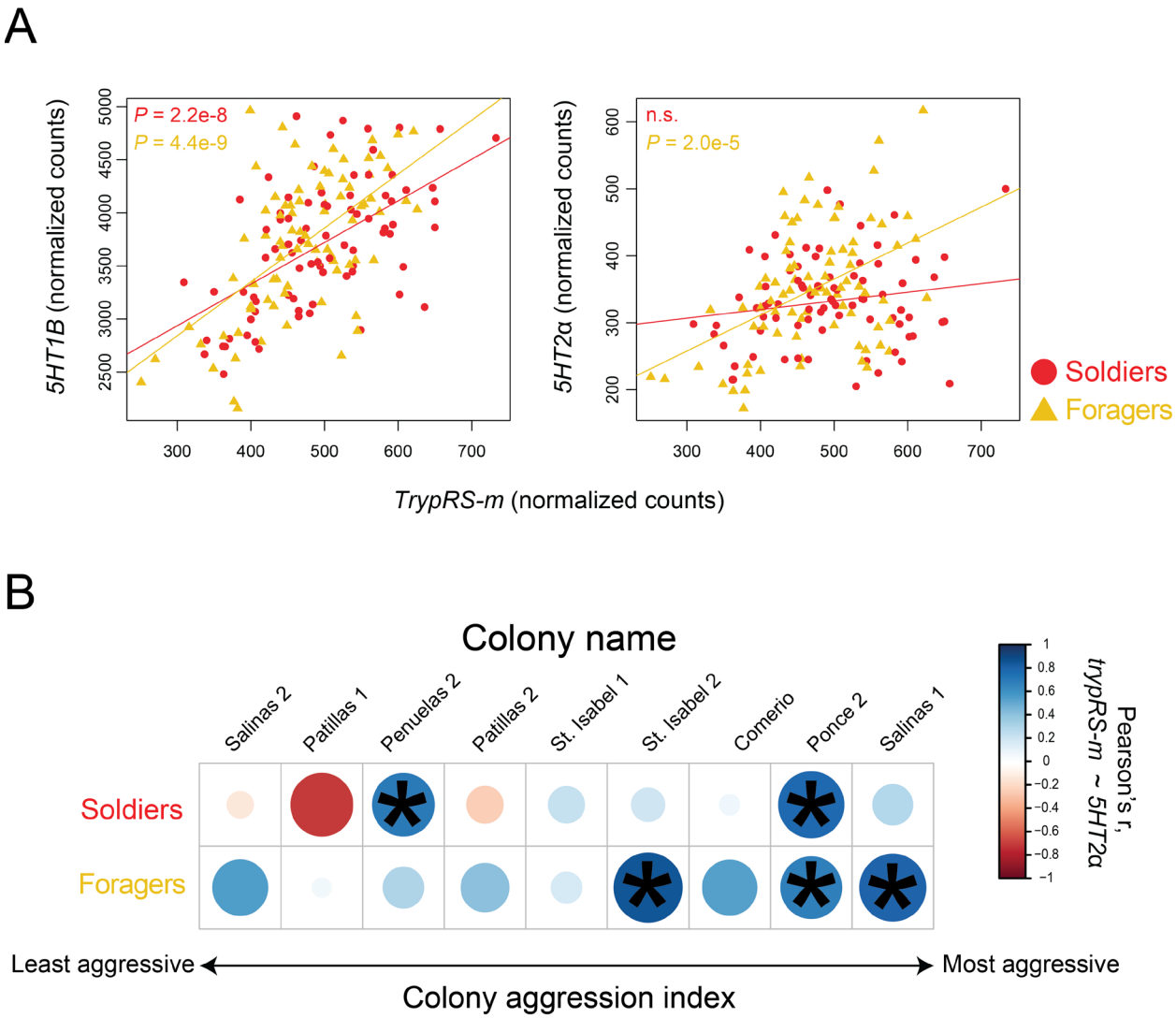


Fig. S6. Brain expression of the serotonin receptor gene *5HT2α* is correlated with *TrypRS-m* expression in foragers, but not soldiers. (A) Normalized expression of the serotonin receptor genes *5HT1B* (left) and *5HT2α* (right), both of which were differentially expressed in foragers compared to soldiers, were correlated with expression of the tryptophan metabolism gene *TrypRS-m* in behavior-specific context: while a significant correlation was detected in both soldiers and foragers for *5HT1B*, it was only detected in foragers for *5HT2α*. (B). When broken down by colony, the correlation between *5HT2α* and *TrypRS-m* was stronger in foragers than soldiers for the most aggressive colonies. Circle size corresponds to absolute value of correlation, and only significant (Pearson's correlation, $*P < 0.10$) positive associations are shown.

796 **References**

797 1. M. Fagny, F. Austerlitz, Polygenic adaptation: Integrating population genetics and gene
798 regulatory networks. *Trends Genet.* **37**, 631–638 (2021).
799 2. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From
800 polygenic to omnigenic. *Cell*. **169**, 1177–1186 (2017).
801 3. S. Sinha, B. M. Jones, I. M. Traniello, S. A. Bukhari, M. S. Halfon, H. A. Hofmann, S.
802 Huang, P. S. Katz, J. Keagy, V. J. Lynch, M. B. Sokolowski, L. J. Stubbs, S. Tabe-
803 Bordbar, M. F. Wolfner, G. E. Robinson, Behavior-related gene regulatory networks: A
804 new level of organization in the brain. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23270–23279
805 (2020).
806 4. S. Chandrasekaran, S. A. Ament, J. A. Eddy, S. L. Rodriguez-Zas, B. R. Schatz, N. D.
807 Price, G. E. Robinson, Behavior-specific changes in transcriptional modules lead to
808 distinct and predictable neurogenomic states. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18020–5
809 (2011).
810 5. K. L. Hoke, E. Adkins-Regan, A. H. Bass, A. R. McCune, M. F. Wolfner, Co-opting evo-
811 devo concepts for new insights into mechanisms of behavioural diversity. *J. Exp. Biol.*
812 **222**, jeb190058 (2019).
813 6. A. Avalos, H. Pan, C. Li, J. P. Acevedo-Gonzalez, G. Rendon, C. J. Fields, P. J. Brown, T.
814 Giray, G. E. Robinson, M. E. Hudson, G. Zhang, A soft selective sweep during rapid
815 evolution of gentle behaviour in an Africanized honeybee. *Nat. Commun.* **8**, 1–9 (2017).
816 7. A. Avalos, M. Fang, H. Pan, A. Ramirez Lluch, A. E. Lipka, S. D. Zhao, T. Giray, G. E.
817 Robinson, G. Zhang, M. E. Hudson, Genomic regions influencing aggressive behavior in
818 honey bees are defined by colony allele frequencies. *Proc. Natl. Acad. Sci.*, 201922927
819 (2020).
820 8. I. M. Traniello, Z. Chen, V. A. Bagchi, G. E. Robinson, Valence of social information is
821 encoded in different subpopulations of mushroom body Kenyon cells in the honeybee
822 brain. *Proc. R. Soc. B Biol. Sci.* **286**, 20190901 (2019).
823 9. I. M. Traniello, S. A. Bukhari, J. Kevill, A. C. Ahmed, A. R. Hamilton, N. L. Naeger, D.
824 C. Schroeder, G. E. Robinson, Meta-analysis of honey bee neurogenomic response links
825 Deformed wing virus type A to precocious behavioral maturation. *Sci. Rep.* **10**, 1–12
826 (2020).
827 10. L. Sheng, E. J. Shields, J. Gospocic, K. M. Glastad, P. Ratchasanmuang, S. L. Berger, A.
828 Raj, S. Little, R. Bonasio, Social reprogramming in ants induces longevity-associated glia
829 remodeling. *Sci. Adv.* **6**, eaba9869 (2020).
830 11. M. S. Hill, P. Vande Zande, P. J. Wittkopp, Molecular and evolutionary processes
831 generating variation in gene expression. *Nat. Rev. Genet.* **22** (2021), pp. 203–215.
832 12. B. A. Harpur, S. M. Kadri, R. O. Orsi, C. W. Whitfield, A. Zayed, Defense response in
833 Brazilian honey bees (*Apis mellifera scutellata* × spp.) is underpinned by complex patterns
834 of admixture. *Genome Biol. Evol.* **12**, 1367–1377 (2020).
835 13. M. Fagny, J. N. Paulson, M. L. Kuijjer, A. R. Sonawane, C. Y. Chen, C. M. Lopes-
836 Ramos, K. Glass, J. Quackenbush, J. Platig, Exploring regulation in tissues with eQTL
837 networks. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7841–E7850 (2017).
838 14. L. Boccuto, C.-F. F. Chen, A. R. Pittman, C. D. Skinner, H. J. McCartney, K. Jones, B. R.
839 Bochner, R. E. Stevenson, C. E. Schwartz, Decreased tryptophan metabolism in patients
840 with autism spectrum disorders. **4**, 1–10 (2013).
841 15. C. C. Rittschof, H. J. Vekaria, J. H. Palmer, P. G. Sullivan, Biogenic amines and activity

levels alter the neural energetic response to aggressive social cues in the honey bee *Apis mellifera*. *J. Neurosci. Res.* **97**, 991–1003 (2019).

16. S. F. de Boer, B. Olivier, J. Veening, J. M. Koolhaas, The neurobiology of offensive aggression: Revealing a modular view. *Physiol. Behav.* **146**, 111–127 (2015).
17. G. J. Hunt, Flight and fight: A comparative view of the neurophysiology and genetics of honey bee defensive behavior. *J. Insect Physiol.* **53**, 399–410 (2007).
18. M. Nouvian, S. Mandal, C. Jamme, C. Claudianos, P. D'Ettorre, J. Reinhard, A. B. Barron, M. Giurfa, Cooperative defence operates by social modulation of biogenic amine levels in the honey bee brain. *Proc. R. Soc. B Biol. Sci.* **285**, 20172653 (2018).
19. T. R. Clandinin, C. H. Lee, T. Herman, R. C. Lee, A. Y. Yang, S. Ovasapyan, S. L. Zipursky, Drosophila LAR regulates R1-R6 and R7 target specificity in the visual system. *Neuron*. **32**, 237–248 (2001).
20. B. Rivera-Marchand, T. Giray, E. Guzmán-Novoa, The cost of defense in social insects: insights from the honey bee. *Entomol. Exp. Appl.* **129**, 1–10 (2008).
21. J. N. Weber, B. K. Peterson, H. E. Hoekstra, Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nat. 2013 4937432*. **493**, 402–405 (2013).
22. A. Sih, Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. *Anim. Behav.* **85**, 1077–1088 (2013).
23. G. J. Hunt, E. Guzmán-Novoa, M. K. Fondrk, R. E. Page, Quantitative trait loci for honey bee stinging behavior and body size. *Genetics*. **148**, 1203–13 (1998).
24. M. D. Breed, G. E. Robinson, R. E. Page, Division of labor during honey bee colony defense. *Behav. Ecol. Sociobiol.* **27**, 395–401 (1990).
25. A. Avalos, Y. Rodríguez-Cruz, T. Giray, Individual responsiveness to shock and colony-level aggression in honey bees: evidence for a genetic component. *Behav. Ecol. Sociobiol.* **68**, 761–771 (2014).
26. K. Davie, J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Q. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo Gonzá lez-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken, S. Aerts, A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell*. **174**, 982-998.e20 (2018).
27. H. Y. Shpigler, M. C. Saul, E. E. Murdoch, A. C. Cash-Ahmed, C. H. Seward, L. Sloofman, S. Chandrasekaran, S. Sinha, L. J. Stubbs, G. E. Robinson, Behavioral, transcriptomic and epigenetic responses to social challenge in honey bees. *Genes, Brain Behav.* **16**, 579–591 (2017).
28. Y. Zhang, G. Parmigiani, W. E. Johnson, ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinforma.* **2** (2020), doi:10.1093/NARGAB/LQAA078.
29. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–40 (2010).
30. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
31. T. G. O. Gene Ontology Consortium, Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049-56 (2015).
32. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. **10**, 1–7 (2009).

33. F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS One*. **6**, e21800 (2011).
34. L. Shen, GeneOverlap: An R package to test and visualize gene overlaps (2014), (available at http://shenlab-sinai.github.io/shenlab-sinai/).
35. A. A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. **28**, 1353–1358 (2012).
36. T. Ilicic, J. K. Kim, A. A. Kolodziejczyk, F. O. Bagger, D. J. McCarthy, J. C. Marioni, S. A. Teichmann, Classification of low quality cells from single-cell RNA-seq data. *Genome Biol. 2016 171*. **17**, 1–15 (2016).
37. G. L, K. O, K. G, Mitochondria: master regulators of danger signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 780–788 (2012).
38. S. A. Detmer, D. C. Chan, Functions and dysfunctions of mitochondrial dynamics. *Nat. Rev. Mol. Cell Biol. 2007 811*. **8**, 870–879 (2007).
39. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell*. **177**, 1888-1902.e21 (2019).
40. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
41. D. F. Simola, R. J. Graham, C. M. Brady, B. L. Enzmann, C. Desplan, A. Ray, L. J. Zwiebel, R. Bonasio, D. Reinberg, J. Liebig, S. L. Berger, Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science*. **351** (2016), doi:10.1126/science.aac6633.
42. V. Croset, C. D. Treiber, S. Waddell, Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *Elife*. **7**, e34550 (2018).
43. N. Konstantinides, K. Kapuralin, C. Fadil, L. Barboza, R. Satija, C. Desplan, Phenotypic convergence: Distinct transcription factors regulate common terminal features. *Cell*. **174**, 622-635.e13 (2018).
44. H. Li, F. Horns, B. Wu, Q. Xie, J. Li, T. Li, D. J. Luginbuhl, S. R. Quake, L. Luo, Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell*. **171**, 1206-1220.e22 (2017).
45. S. Suenami, S. Oya, H. Kohno, T. Kubo, Kenyon cell subtypes/populations in the honeybee mushroom bodies: possible function based on their gene expression profiles, differentiation, possible evolution, and application of genome editing. *Front. Psychol.* **9**, 1717 (2018).
46. D. Kretzschmar, G. O. Pflugfelder, Glia in development, function, and neurodegeneration of the adult insect brain. *Brain Res. Bull.* **57**, 121–131 (2002).
47. A. K. Shah, C. D. Kreibich, G. V. Amdam, D. Münch, Metabolic enzymes in glial cells of the honeybee brain and their associations with aging, starvation and food response. *PLoS One*. **13**, e0198322 (2018).
48. J. Peng, G. Serrano, I. M. Traniello, M. E. Calleja-Cervantes, U. V. Chembazhi, S. Bangru, T. Ezponda, J. R. Rodriguez-Madoz, A. Kalsotra, F. Prosper, I. Ochoa, M. Hernaez, A single-cell gene regulatory network inference method for identifying complex regulatory dynamics across cell phenotypes. *bioRxiv* (2021).
49. H. Y. Shpigler, M. C. Saul, E. E. Murdoch, F. Corona, A. C. Cash-Ahmed, C. H. Seward,

934 S. Chandrasekaran, L. J. Stubbs, G. E. Robinson, Honey bee neurogenomic responses to
935 affiliative and agonistic social interactions. *Genes, Brain Behav.* **18**, e12509 (2018).
936 50. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from
937 expression data using tree-based methods. *PLoS One*. **5**, e12776 (2010).
938 51. D. van Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. J. Carr, C. Burdziak, K. R.
939 Moon, C. L. Chaffer, D. Pattabiraman, B. Bierie, L. Mazutis, G. Wolf, S. Krishnaswamy,
940 D. Pe'er, Recovering gene interactions from single-cell data using data diffusion. *Cell*.
941 **174**, 716-729.e27 (2018).
942