Title: Single-cell dissection of a collective behaviour in honeybees

Ian M. Traniello[1,2†*], Syed Abbas Bukhari[2], Payam Dibaeinia[3], Guillermo Serrano[4], Arian Avalos[5], Amy Cash Ahmed[2], Alison L. 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

Understanding how genotypic variation results in phenotypic variation, a major challenge in biology, is especially difficult for collective behaviour because collective group phenotypes arise from complex interactions between group members[1]. Honeybees aggressively defend their colony from attacks with highly integrated collective behaviour in which different groups of bees play specific roles, giving rise to distinct colony-level differences in aggression. A previous genome-wide association study of a population of Africanized honeybees (*Apis mellifera scutellata*) from Puerto Rico that recently evolved decreased aggression identified hundreds of genes with single nucleotide polymorphisms (SNPs) that associated with colony-level variation in aggression[2]. Many of these SNPs also showed strong signals of selection for decreased aggression[2,3], but their influence on brain function was unknown. Using brain single-cell (sc) transcriptomics and sc gene regulatory network analysis, we show here that variants of these genes give rise to genetic differences in transcription factor-target gene relationships. These differences involved the activity of several TFs, some that have been previously associated with aggression, like *single stranded-binding protein c31A*, and some that have been associated with tissue morphogenesis but not behaviour, like *apontic*. The activity of these and other TFs was located in specific brain cell populations related to olfaction and vision, the two sensory modalities that bees use in colony defence. They also implicate metabolism of serotonin, a neurochemical already known to influence honeybee aggression, but not from a genetic perspective. Surprisingly, genetic differences were more pronounced in the brains of forager bees than in similarly aged but more aggressive soldier bees, pointing to an evolutionary change in division of labour for colony defence. Our results demonstrate how group genetics can shape a collective phenotype by modulating individual brain gene regulatory network architecture.

Main

In honeybee societies, individual workers are integrated into collective efforts via division of labour[4,5]. One such example is colony defence following territorial intrusion, in which highly aggressive soldier bees mount a rapid stinging attack while similarly aged nestmates ignore the disturbance and instead forage for nectar and pollen. Colony-level variation in aggression is well known in honeybees, and both inherited and environmental determinants have been identified[2]. This is consistent with Darwin's assertion that natural selection acts at the colony level in the insect societies[6,7], but it is not known at the genomic level how colony-level selection shapes the phenotypes of individual colony members. We used specific variants of genes that associated with variation in colony aggression ("colony aggression genes") to explore this problem.

We tested the prediction that variation in collective behaviour manifests in the neurogenomic architecture of high-aggression soldiers and low-aggression foragers by using the recent evolution of decreased aggression in the otherwise highly aggressive Africanized honeybee (AHB)[3]. We compared whole-brain transcriptomic profiles of 81 soldiers and 82 foragers from nine Puerto Rican honeybee colonies, using the same individuals previously subjected to a whole-genome analysis[2]. We coupled this analysis with scRNA-Sequencing to profile the transcriptomes of ~24,000 cells from 40 soldier and forager brains 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. Neither DEG list overlapped significantly with the colony aggression genes (Fig. 1A; fig. S2C; Data S2-S3).

We next used expression quantitative trait loci (eQTL) mapping to search for specific genes linking variation in group genetics to variation in individual brain function[8]. 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[2] (Data S4). Like for colony aggression genes, we did not find a significant overlap with the soldier-forager DEGs at bulk or sc resolution (Fig. 1B; fig S2C). We therefore asked whether genetic differences in colony aggression manifest at the level of gene regulatory networks (GRNs).

Recent theory predicts that the genetic architecture of polygenic traits, including most behaviours, is best understood at the level of the GRN, which models regulatory relationships between transcription factors (TFs) and their target genes (TGs)[9,10]. This prediction is consistent with an emerging appreciation that behaviourally related GRNs are an important level of organisation in the brain[11]. Developmental GRNs have proven to be useful in the study of evolutionary changes in morphological traits[12], and variation in DNA sequence has been associated with subtle changes in GRN activity underlying disease states[13,14]. However, GRN analysis has not yet been employed to study how changes in the brain affect collective behaviour.

We modeled GRNs from the sc data and determined the regulatory dissimilarity score, a measure of network difference that integrates across all cell-type-specific TF-TG relationships[15]. There was greater GRN dissimilarity between soldiers and foragers from more aggressive colonies relative to less aggressive colonies (Fig. 1C-E; fig. S3). This finding indicates that the collective environment is related to brain gene regulatory architecture for individual colony members, an example of an indirect genetic effect[1] acting on the brain. These genetic differences in regulatory architecture 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 honeybee aggression via *in situ* hybridization[16] and the latter have been associated with behavioural plasticity in honeybees[17] and ants[18], but never before from a genetic perspective.

Having established a relationship between genetic differences in colony aggression and differences in GRN structure, we tested for an influence of the 30 eGenes on the brain gene regulatory architecture underlying the soldier and forager behavioural states. We developed a new computational method, similar to ref. [19] and tested 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 distributed across different brain cell populations. 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 the colony aggression genes[2]. 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 reveals a mechanism by which group genetics can shape brain and behaviour by influencing the functional integration of GRNs.

The eGene subnetworks featured four highly connected TFs: *Single stranded-binding protein c31A* (*Ssb-c31a*), *apontic* (*apt*), *ecdysoneless* (*ecd*), and *CG31460* (Fig. 3A; Data S5). *Ssb-c31a*, which has been associated with population-level variation in honeybee aggression[20], was most prominent in astrocyte-like glia, already identified above, and perineurial glia. The TF *apt*, which has been associated with tissue morphogenesis in *D. melanogaster*[21] but has not yet been implicated in social behaviour, was most prominent in the inner-compact KC, a small population of MB neurons flanking the above-mentioned non-compact KC which also have been implicated in honeybee aggression[16], and in the antennodorsal projection neurons of the antennal lobe, which communicate olfactory sensory information to the KC. Overall, 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).

Also identified in the eGene subnetwork analysis was the TG *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[22], the precursor to the neurochemical serotonin. Serotonin is involved in the regulation of aggression in diverse species of vertebrates and invertebrates[23], including honeybees[24]. The strongest soldier-forager differences in TF-TG relationships for *TrpRS-m* were in cell types that contain optic lobe markers[25] 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 *Drosophila melanogaster*[26]. eGene subnetwork analyses indicate that inherited changes in honeybee aggression are based, in part, on changes in the structure and function of physiological pathways that respond to or process aggression-triggering olfactory and visual stimuli.

The evolution of decreased colony aggression in the Puerto Rican population involved a soft selective sweep[3], so it is likely that colony-level selection had multiple influences on the brains of colony members. Surprisingly, selection apparently affected the brains of foragers more than soldiers, based on the following four results: 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. 3B); 3) serotonin-related genes were significantly upregulated in the brains of foragers compared to soldiers (fig. S1A; Data S1); and 4) expression of the serotonin receptor *5HT2α* more strongly correlated with expression of *TrypRS-m* in foragers than in soldiers, and this correlation was most pronounced in foragers from highly aggressive colonies (fig. 3C-D). These results make sense in light of how division of labour works in honeybee colonies. Unlike in ant and termite societies, there are no morphological differences among honeybee workers[27], so variation in individual aggression is in degree and not kind. Honeybee colony aggression and foraging intensity are inversely correlated because foragers join in defending their colony if triggering stimuli exceed their threshold of response[28]. These results suggest that 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.

How the behaviour of individuals is integrated into a smoothly functioning group is one of the central issues in the study of collective behaviour[27]. Our study shows how genomics can be used to analyze a collective behaviour that has undergone evolutionary change by using genotype-phenotype mapping that focuses on GRN architecture. This genotype-phenotype mapping was not

174 possible with analyses based only on gene expression. The approach used here, which uses
175 scGRNs to integrate population genetics and systems biology, provides tools to better understand
176 the evolutionary tuning of a collective behaviour.
177

178

Figures:

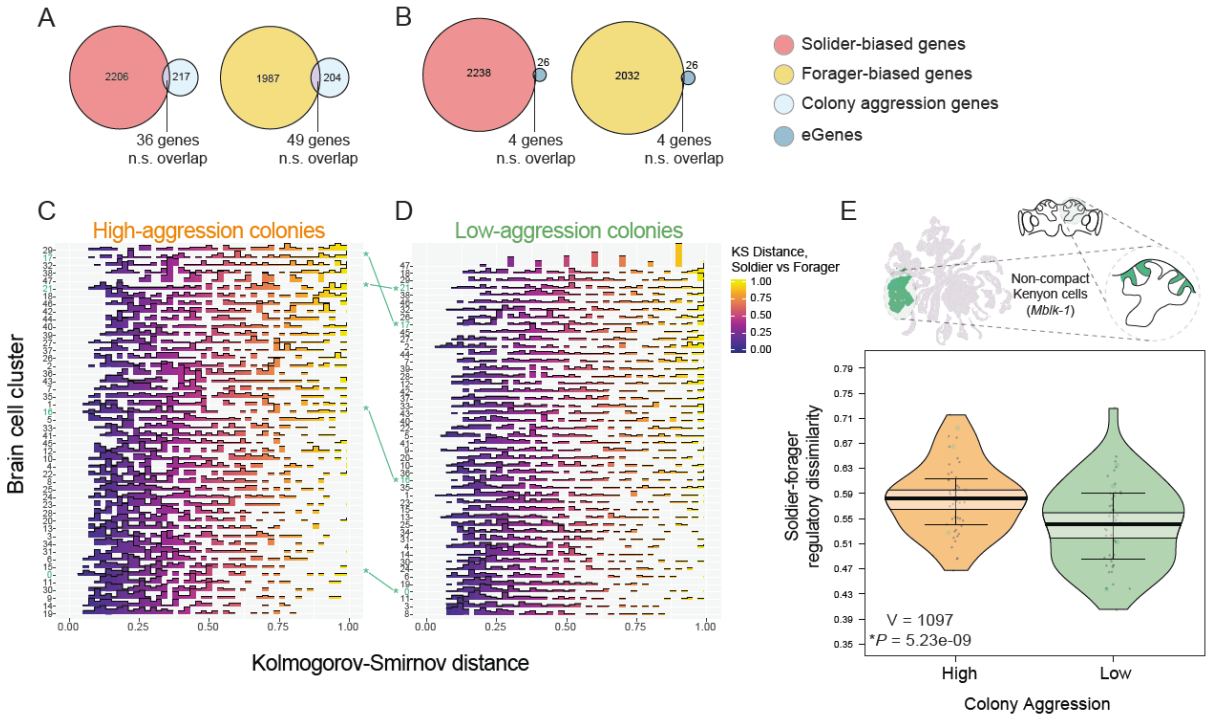


Fig. 1. Dissimilarity of brain gene regulatory network (GRN) architecture between soldiers and foragers is greater in high-aggression honeybee colonies. (A) No significant enrichment of genes containing colony aggression-associated SNPs ("colony aggression genes") or (B) of eGenes in soldier or forager-biased 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.

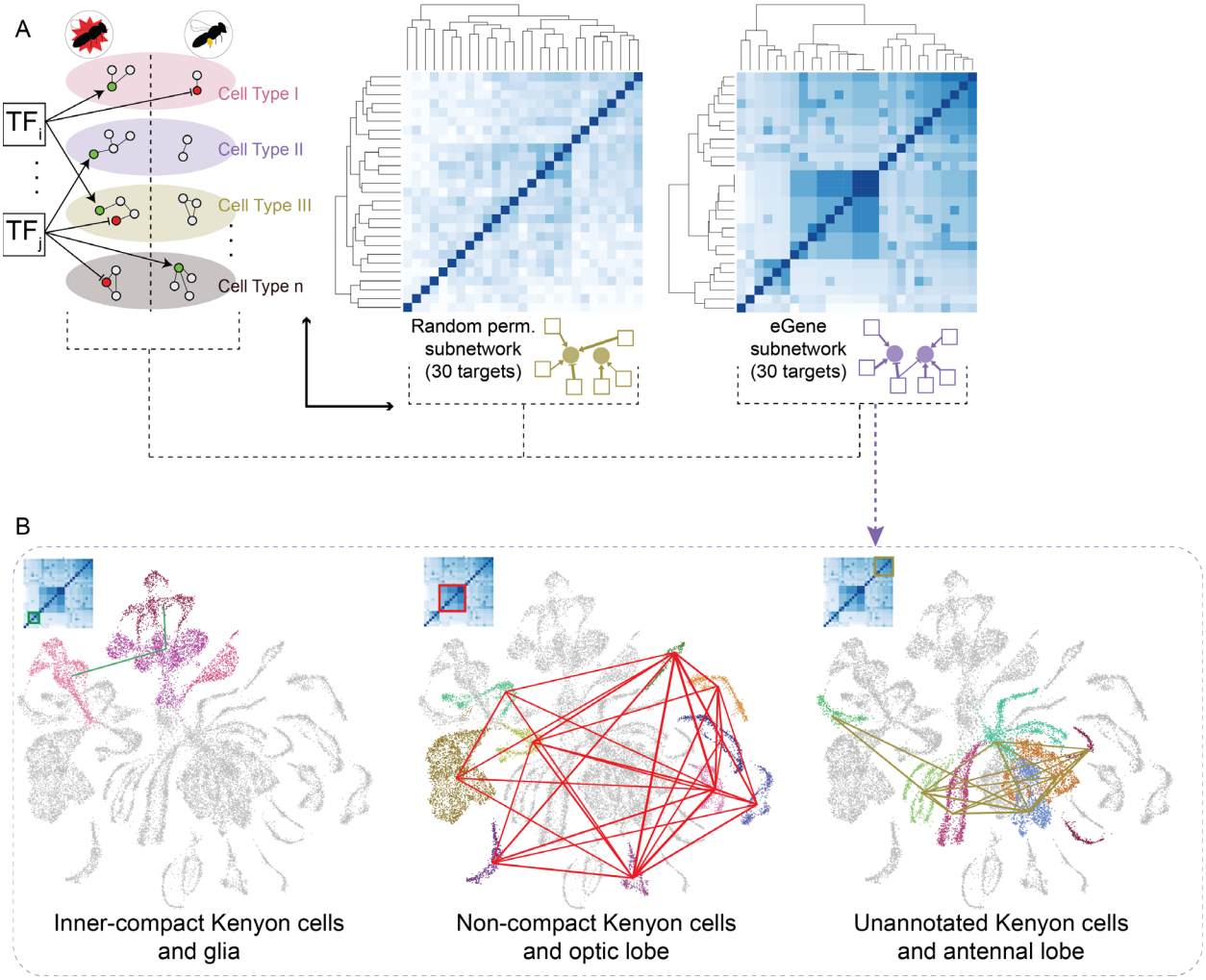
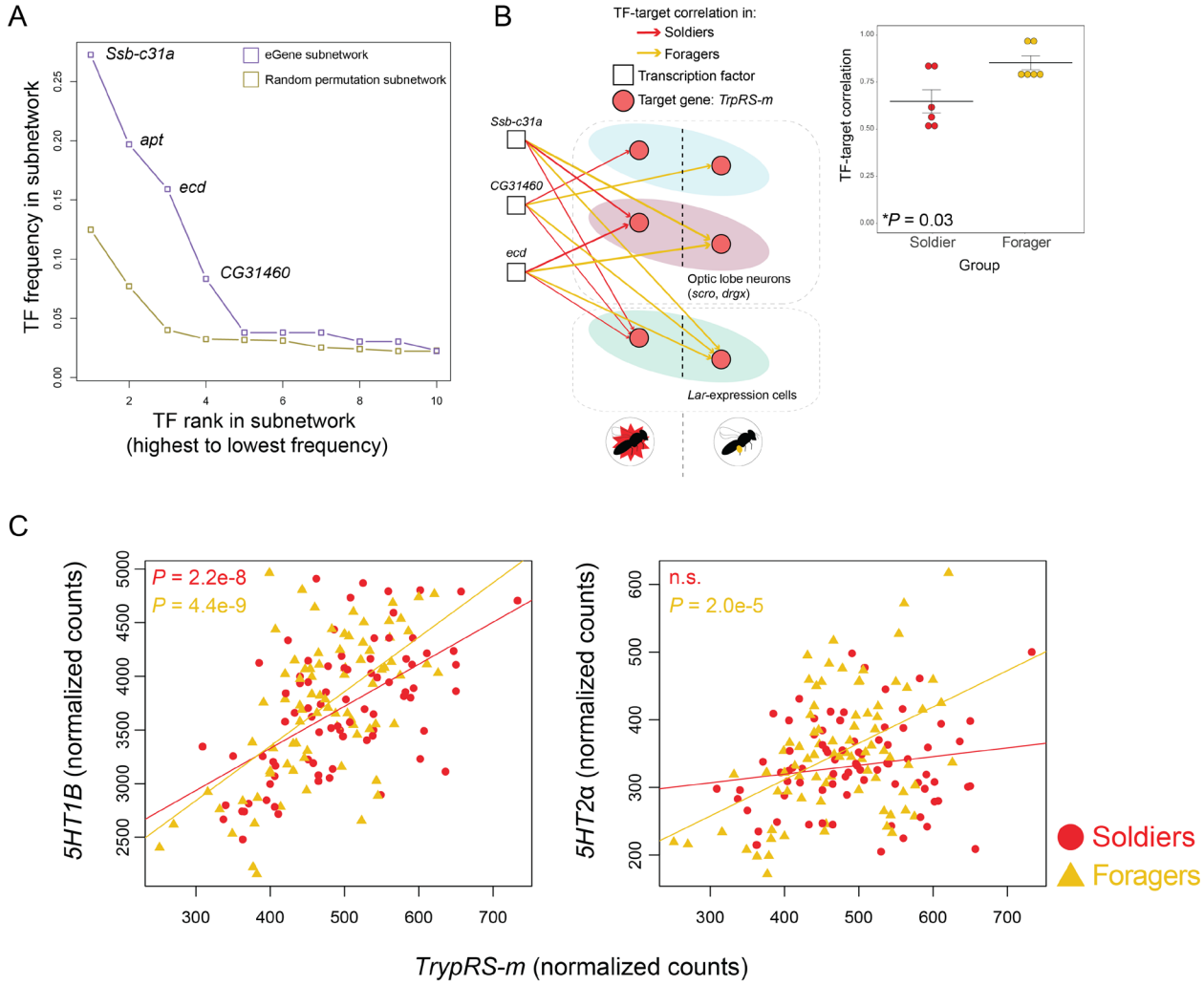


Fig. 2. Evolutionary changes in honeybee colony aggression are associated with changes in brain cell gene regulatory network (GRN) architecture between soldiers and foragers. (A) Subnetworks composed of either eGenes or random subsets of genes (negative control) as target genes (TGs) were extracted from an inferred scGRN (Methods). Heatmaps comparing these subnetworks across cell 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 Transcription Factor-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 cell clusters. For visualization purposes, only edges with similarity ≥ 0.60 are shown.


A
Ssb-c31a
eGene subnetwork
Random permutation subnetwork
apt
ecd
CG31460
TF frequency in subnetwork
TF rank in subnetwork
(highest to lowest frequency)
B
TF-target correlation in:
Soldiers
Foragers
Transcription factor
Target gene: TrpRS-m
Ssb-c31a
CG31460
ecd
Optic lobe neurons
(scro, drgx)
Lar-expression cells
TF-target correlation
*P = 0.03
Soldier
Forager
Group
C
P = 2.2e-8
P = 4.4e-9
5HT1B (normalized counts)
n.s.
P = 2.0e-5
5HT2α (normalized counts)
Soldiers
Foragers
TrypRS-m (normalized counts)

D

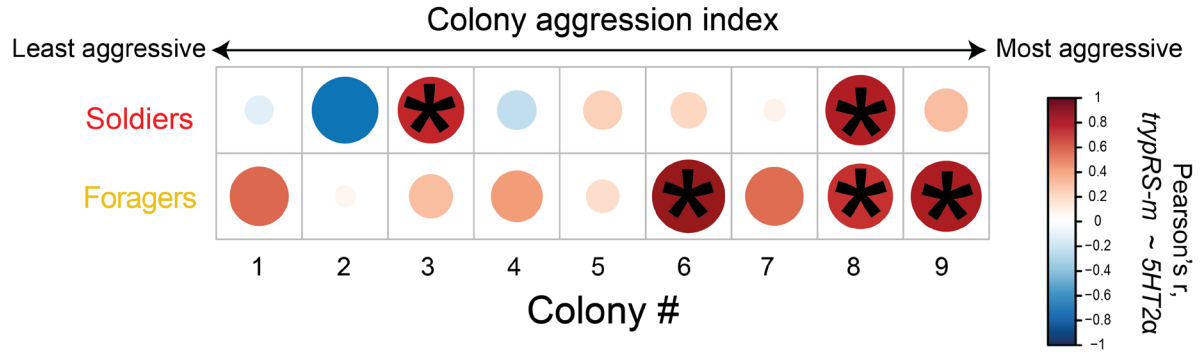

Colony aggression index
Least aggressive
Most aggressive
Soldiers
Foragers
1
2
3
4
5
6
7
8
9
Colony #
Pearson's r,
trypRS-m ~ 5HT2α

Figure 3. Serotonin metabolism as a representative example of the relationship between colony genetics, regulatory plasticity, and individual behavioural variation. (A) Analysis of TF centrality, as measured by frequency of a given TF, revealed four TFs overrepresented in eGene subnetworks relative to control subnetworks. (B) Strong regulatory plasticity of the TG *TrpRS-m* between soldiers in foragers in three cell clusters identified as part of the honeybee 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. (C) 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 behaviour-specific context: while a significant correlation was detected in both soldiers and foragers for *5HT1B*, it was only detected in foragers for *5HT2α*. (D). 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.

Acknowledgements: We thank A. Hernandez and C. Wright at the Roy J. Carver Biotechnology Center (UIUC) for sequencing services, C. Desplan for useful discussion, and M.B. 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. [2] 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 availability: Whole-tissue and single-cell RNA-Sequencing data will be made accessible following acceptance for publication.

Supplementary Materials for

Single-cell dissection of a collective behaviour

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 PDF file includes:

Materials and Methods
Figs. S1 to S5
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 behavioural 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 honeybee (gAHB) soldier and forager collections

Collection strategies for gAHB soldiers and foragers used for the previously published whole genome sequences are described in ref. [2]. 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[29,30]. 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[2] 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 behavioural ranking system[29,31], 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. Behaviours "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 ref. [2].

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[2], 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 Technologies, Santa Clara, CA). RNA-Sequencing (RNA-Seq) libraries were prepared on two plates with the TruSeq Stranded mRNAseq Sample Prep Kit, and libraries were quantitated via qPCR. We performed sequencing on one lane of a NovaSeq 6000 with the NovaSeq S4 reagent kit for 151 cycles from each end of the fragments. Reads were 150 nucleotides in length. We

352 demultiplexed FastQ files with the bcl2fastq Conversion Software v2.20 (all Illumina unless
353 otherwise stated).
354
355 *European honeybee (EHB) soldier and forager collections*
356
357 For collections of EHB soldiers and foragers in Urbana, IL, we used the same approach outlined
358 in ref. [2], as detailed above, which describes the gAHB samples referenced above and the bulk brain
359 RNA-Seq analysis used in this study. All EHB samples were used for single-cell RNA-Seq
360 (scRNA-Seq).
361
362 All colonies were maintained according to standard beekeeping practices at the Bee Research
363 Facility at UIUC. We relocated four standard Langstroth hives containing a colony of ~30,000
364 workers headed by a naturally mated queen to a common yard to control for environmental factors
365 like floral resource access and exposure to ambient conditions. Hives were equal in size, containing
366 two stacked hive boxes with identical amounts of honeycomb and were left undisturbed for at least
367 10 days after relocation and prior to phenotyping. We collected soldiers and foragers from one
368 colony per day across four successive days in August 2019 (n = 4 soldier and 4 forager samples).
369 On each collection day, we began at 10:00AM by performing the colony disturbance / stinging
370 assay as described above. However, instead of being flash-frozen, samples were rapidly
371 transported indoors and processed for scRNA-Seq.
372
373 *Single-cell RNA-Sequencing (scRNA-Seq)*
374
375 Soldier and forager brains were dissected in ice-cold Dulbecco's PBS (DPBS, cat. no. #14190144,
376 Thermo Fisher Scientific) in a sterile, RNase-free space. We then followed the "short dissociation
377 protocol" from[25], which allowed us to go from sample collection in the field to cells in suspension
378 within minutes. Briefly, we cleaned away glands, trachea, and any connective tissue and cut each
379 freshly dissected brain into six pieces, leaving two optic lobes and the central brain divided into
380 four quadrants. We were careful to only cut through neuropil, thus leaving neurons and neuron-
381 dense regions intact. Brain pieces were transferred to a sterile centrifuge tube containing 500 ul of
382 DPBS on ice. After this was repeated for five brains per soldier or forager sample, we allowed all
383 tissue to sink to the bottom of the tube and replaced DPBS with 200 ul of a diluted Trypsin-EDTA
384 solution (0.05%, cat. no. 25300054, Invitrogen, Carlsbad, CA) and transferred the centrifuge tube
385 to a thermomixer. Samples were mixed at 1000 rpm and 25°C for 15 min with gentle manual
386 trituration with an ordinary pipette tip every 5 min. We then removed residual tissue from the
387 bottom of the tube with a wide-bore pipette tip and centrifuged cells at 400 rcf and 4°C. We
388 removed the supernatant and washed the cells twice in 1 ml ice-cold DPS with a centrifugation
389 step in between washes. After the second wash, we performed a final centrifugation step and
390 resuspended cells in 400 ul DPBS with 0.04% UltraPure™ Bovine Serum Albumin (AM2616,
391 Thermo Fisher Scientific). We then filtered cells with a 40 Flowmi™ Tip Strainer (H136800040,
392 Bel-Art, Wayne, NJ). Cell suspensions were stored on ice and immediately transported to the Roy
393 J. Carver Biotechnology Center at UIUC, where they were subjected to a total cell count and %
394 viability assay with AOPI staining on a Nexcelom Cellometer Spectrum Image Cytometry System
395 (Nexcelom Biosciences, Lawrence, MA USA). Viability was consistently > 97% for each sample.
396

397 Libraries were constructed with the Chromium Single-Cell V3 kits (10x Genomics, Pleasanton,
398 CA) and the library pool was amplified via qPCR and loaded on one lane of a NovaSeq 6000 S4
399 flowcell (Illumina, San Diego, CA), where it was sequenced for 28 nt from one end and 151 nt
400 from the other end. FastQ files were generated and demultiplexed with the CellRanger v3.0.2
401 "mkfastq" command. The pool of libraries produced over 5 billion reads, and excellent quality was
402 observed for each sample via fastQC.
403
404 Within a few days after collections, we inspected all four colonies and confirmed the presence of
405 a laying queen, meaning each colony was queenright at the time of sample collection.
406
407 **Bioinformatics**
408
409 *Bulk RNA-Seq*
410
411 We removed 12 samples from analysis due to contamination with a common honeybee virus[17,32]
412 and two samples with an abnormally low number of sequenced reads, leaving us with a final
413 sample size of 163 individuals (81 soldiers and 82 foragers) collected from nine colonies. We
414 identified and removed a batch effect from library preparation plate using Combat-Seq[33] before
415 performing differential gene expression analysis with edgeR[34]. After filtering to remove transcripts
416 that had fewer than one count-per-million in at least the number of samples equal to the smallest
417 group size, we performed TMM normalization on a gene list universe of 9185 genes for pairwise
418 soldier versus forager comparisons and 9827 for an analysis of genes correlated with colony
419 aggression in soldiers or foragers, the latter being larger on account of smaller group sizes and
420 therefore a more relaxed filtering. We used a generalized linear model (GLM) to compare gene
421 expression between soldiers and foragers using colony as a blocking factor. We next independently
422 analyzed soldiers and foragers to identify genes positively or negatively associated with aggression
423 by regressing individual-level gene expression on each individual's colony aggression value,
424 derived from the MDS dimensional vector from[2], as a continuous predictor of gene expression. To
425 correct for multiple tests, we performed the Benjamini-Hochberg (BH) correction [35] and filtered
426 differentially expressed gene (DEG) lists based on a false discovery rate (FDR) threshold of 0.05.
427 For Gene Ontology (GO) enrichment analysis[36], we first converted soldier- or forager-biased
428 honeybee genes to *Drosophila melanogaster* orthologues using a one-to-one reciprocal best hit
429 BLAST (RBH-BLAST). Enrichment of terms associated with Biological Processes was performed
430 with GOrilla [37] and visualized with REVIGO[38].
431
432 Gene list overlap analysis was performed via hypergeometric tests with BH-FDR correction for
433 multiple tests using the GeneOverlap package[39]. This approach was also used to compare bulk and
434 single-cell data. We directly compared gene lists associated with behavioural state, colony
435 aggression level, or cell cluster (described below) with a list of 253 genes that contained
436 aggression-associated single-nucleotide polymorphisms (SNPs); 60 of these genes also showed
437 significant signatures of selection, meaning their frequency was found to be increased in the gAHB
438 population. We refer to these 253 genes as "colony aggression genes," and the complete list can
439 be found in Dataset S1 of ref. [2].
440 *Expression quantitative trait loci (eQTL) mapping*
441

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 ref. [2]. 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[40] 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.

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 ref. [2]. 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[41–43]. 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[44].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[44] 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*[25], jumping ants[18], and a previous honeybee study[17], 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.

We detected DEGs for each cluster using the MAST algorithm with default settings[45].We performed two DEG analyses to determine 1) genes that defined each cluster (independent of behavioural state), 2) genes that differentiated soldiers and foragers in each cluster. Cluster-

487 defining genes were used to annotate the 48 detected clusters, and cluster-specific soldier-forager
488 DEGs were used for overlap analyses with colony aggression genes and eGenes.
489
490 We annotated many of the 48 detected cell clusters using established insect neuronal and glial
491 biomarkers[25,46,47]. Utilizing available insect scRNA-seq brain atlases of gene expression[18,25,47–49],
492 including a previous honeybee brain scRNA-Seq analysis[17], we annotated cell clusters as
493 populations of cells that represented the mushroom bodies (MB), a higher-order multimodal
494 sensory integration center important for learning and memory, navigation, and social behaviour,
495 which formed a clearly segregated set of clusters. Inner-compact Kenyon cells (KCs) were
496 identified by co-expression of *E74* and *ecdysone receptor* while non-compact KCs were identified
497 by co-expression of *Mblk-1* and *dlg5*[50]. Other KC populations detected via expression of
498 *mushroom body-expressed* could not be further annotated, which is expected considering limited
499 knowledge of cellular diversity in the MB. We could also identify cells in the lateral (*ventral veins*
500 *lacking*) and antennodorsal (*abnormal chemosensory jump 6*) projection neurons of the antennal
501 lobe [49]. Numerous cell populations constituting the optic lobe were also identified and annotated
502 at superficial levels, as more sequencing will be necessary to further identify all the subpopulations
503 of this extremely heterogenous brain region. Still, we identified populations that were specifically
504 enriched for *Lim3*, *dorsal root ganglia homeobox,* and co-expression of *homeobox protein*
505 *homothorax* and *brain-specific homeobox*, markers that span the lobula, medulla, and lamina of
506 the optic lobe[25]. We also found that roughly 88% of cells expressed the neuronal marker *fne* while
507 an average of ~11% of cells expressed insect glial markers *repo* and *borderless*, a neuron:glia ratio
508 consistent with previous reports from other insects [18,25]. Within the glial populations, we identified
509 most known insect glia subtypes[51] including astrocyte-like glia, perineurial glia, subperineurial
510 glia, and cortex glia, as well as a small population of hemocytes (*hml*) that clustered close to the
511 glial populations[25,52].
512
513 *Gene regulatory network analyses: regulatory dissimilarity as a function of variation in colony*
514 *aggression*
515
516 To infer single-cell gene regulatory networks (GRNs), we used SimiC, which has been previously
517 shown to perform well with honeybee brain scRNA-Seq data and is better optimized for comparing
518 GRNs across behavioural states than other established methods, to which it has been
519 benchmarked[15]. SimiC imputes scRNA-Seq expression data to infer the relationship between a
520 transcription factor (TF) and its predicted TGs. As input, we used a list of 193 honeybee TFs, each
521 with an robust one-to-one ortholog to a biochemically or genetically validated TF in *D.*
522 *melanogaster*, which was generated by intersecting TFs predicted by ASTRIX in three previous
523 honeybee brain transcriptomics studies[20,32,53]. We then used SimiC to output a set of incidence
524 matrices containing TF-by-TG (TF x TG) information representative of soldier and forager
525 behavioural states. The GRN evaluation pipeline was based on a weighted area under the curve
526 ("wAUC") score. For each cell, the wAUC score quantified the relative activity of a given regulon
527 (i.e. a row of the GRN incidence matrix, inferred in the previous step) with respect to the
528 expression of all TGs. For any given TF, if the overall expression of its associated TGs in a cell
529 was high and weights between the TF and its TGs were large, then we concluded that the TF in
530 question has a large influence on the expression profile of a particular cell. For all clusters of cells
531 and TFs, the wAUC was computed independently for each cell yielding a distribution of cell-
532 specific wAUC values corresponding to a behavioural state (soldier or forager). To measure the

separation of distributions, we calculated the Kolmogorov-Smirnov (KS) distance between distributions of wAUC values for each cell within a cluster. This yielded one value for each TF per cluster, allowing us to describe "dissimilarities" in soldier and forager regulatory relationships for each predicted cell cluster. Finally, for each cluster, we aggregated the KS distance of each TF into one unique value by computing the area under the curve of the empirical cumulative distribution function (CDF). This approach allowed us to make general comparisons of all cell clusters sampled from soldiers and foragers as well as examine GRN differences specific to the aggression level of the colony of origin.

Gene regulatory network analyses: architecture of subnetworks associated with colony aggression genes

To infer regulatory subnetworks, we applied the GENIE3[54] pipeline to both our bulk and sc datasets, using the same list of 193 TFs as the previous step. For soldiers and foragers, we inferred separate GRNs by collecting the top $\boldsymbol{N}$ regulators (sorted by GENIE3) having a "$\geq \mathbf{C_1}$" absolute Spearman correlation with the TG's expression. Next, we compared soldier and forager GRNs to obtain interactions that are present in both groups with the same TF-TG correlation directionality; this reduced the likelihood of identifying variation due to under-sampling in either soldier or forager groups. We repeated this analysis for bulk and sc expression data. For the sc analysis, we first imputed the expression matrix using MAGIC[55] with diffusion ("t") = 2 and split the results into 26 matrices, one for each cell cluster (clustering resolution was set 0.2 in Seurat, with all other parameters unchanged, in order to obtain enough cells per cluster for the analysis). For bulk data, 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 extracted TF-TG relationships in which TGs were either the 30 eGenes (described above) or 15 permutations of 30 random eQTL-genes that were also identified in the eQTL analysis but were not the 30 eGenes (control).

We then assessed similarity of GRN architecture across cell types in eGene and control subnetworks. First, we extracted from each cell cluster a binary vector in which each element corresponded to a TF-TG interaction and compared these vectors across all cell clusters, using Jaccard index as a measure of closeness. Matrices containing Jaccard indices from pairwise comparisons made across cell clusters for eGene or control subnetworks in soldiers and foragers were hierarchically clustered using complete-linkage clustering of the Euclidean distance, and dendrograms for the eGene and control subnetworks were cut at a height of 1.2. After cutting, we specifically explored connectivity of "blocks" (i.e. groups containing more than three cell clusters sharing similar GRN subnetwork architecture based on the above parameters); we identified three blocks from the eGene subnetwork and no blocks from the control subnetwork.

The strength of connectivity (i.e. similarity of subnetwork architecture across cell clusters) was assessed by assigning a score to each block in the eGene subnetwork representing the smallest pairwise connectivity score observed between clusters within a block. This score was then computed in each of the 15 permutation subnetworks using the same cell clusters identified in each 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.

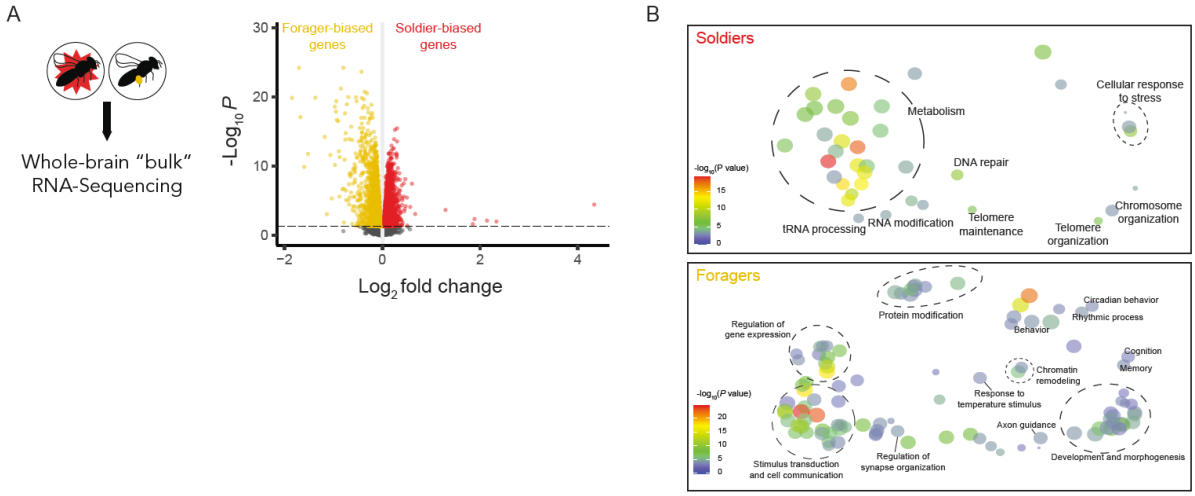


Fig. S1. Whole-tissue brain transcriptomics of soldiers and foragers. (A) Whole-brain RNA-Sequencing of the same soldier and forager bees from [2] revealed thousands of differentially expressed genes (DEGs) associated with behavioural state, termed "soldier-" or "forager-biased" based on fold-change 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.

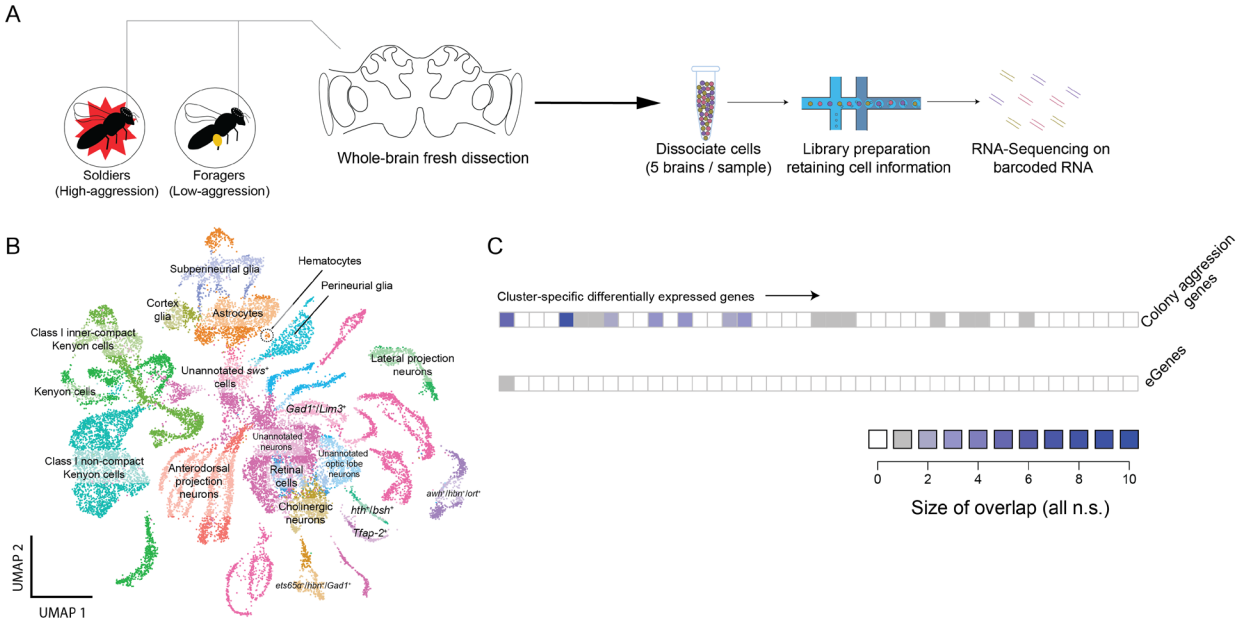


Fig. S2. Single-cell brain transcriptomics of honeybee soldiers and foragers. (A) Workflow for transcriptomic profiling of ~24,000 cells from the brains of field-collected soldiers and foragers and (B) bioinformatic clustering of cells in reduced dimensionality (UMAP) space. Clusters were manually annotated 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 [2] 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.

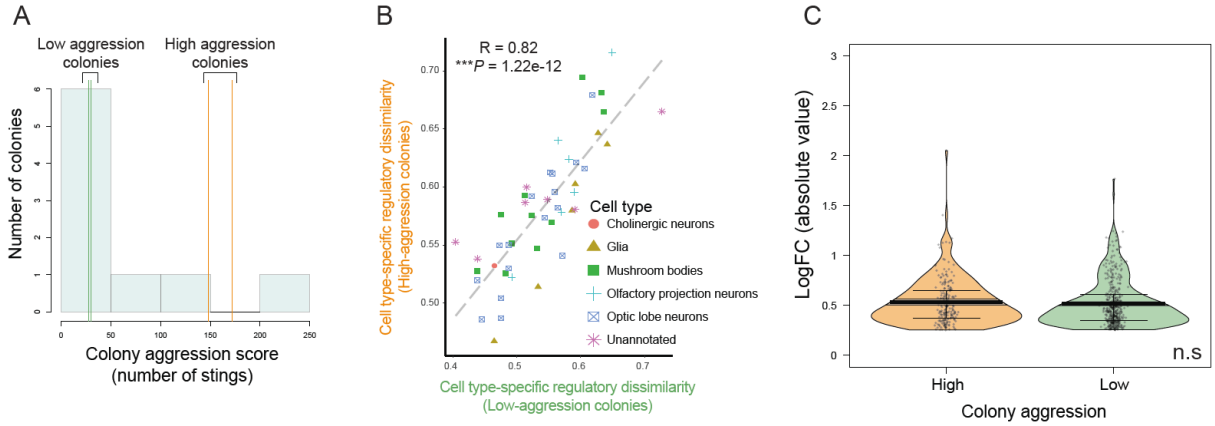
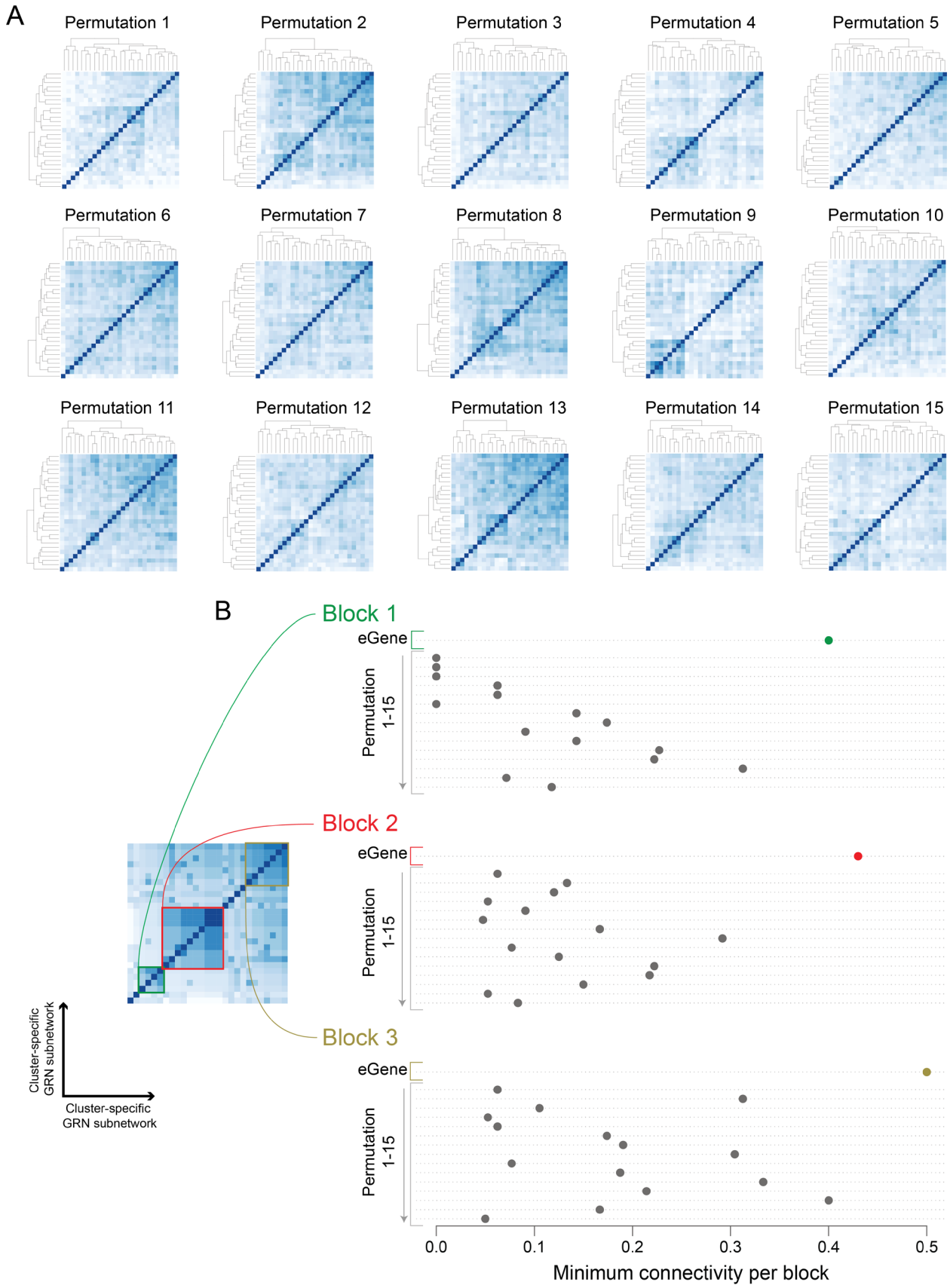
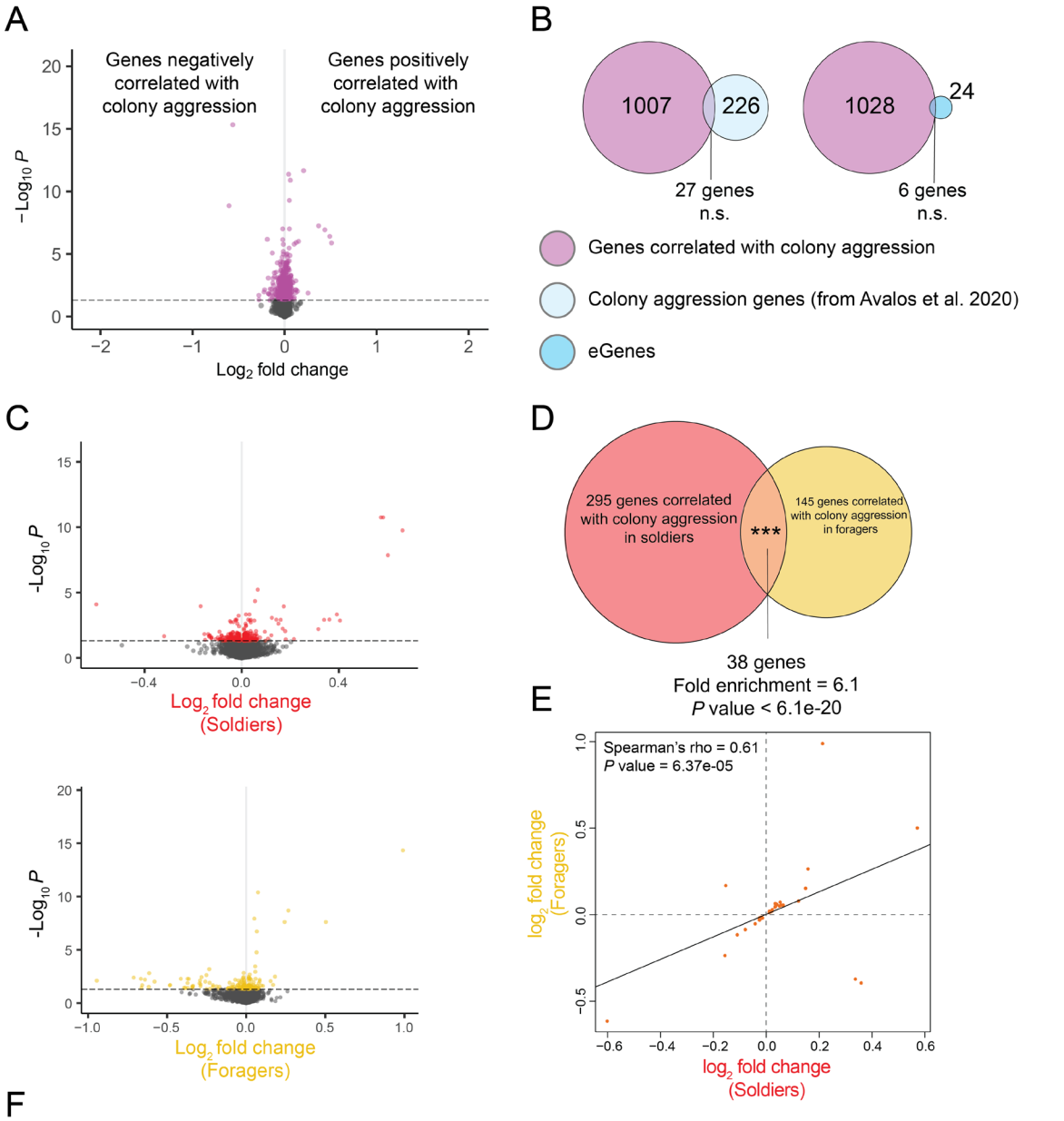


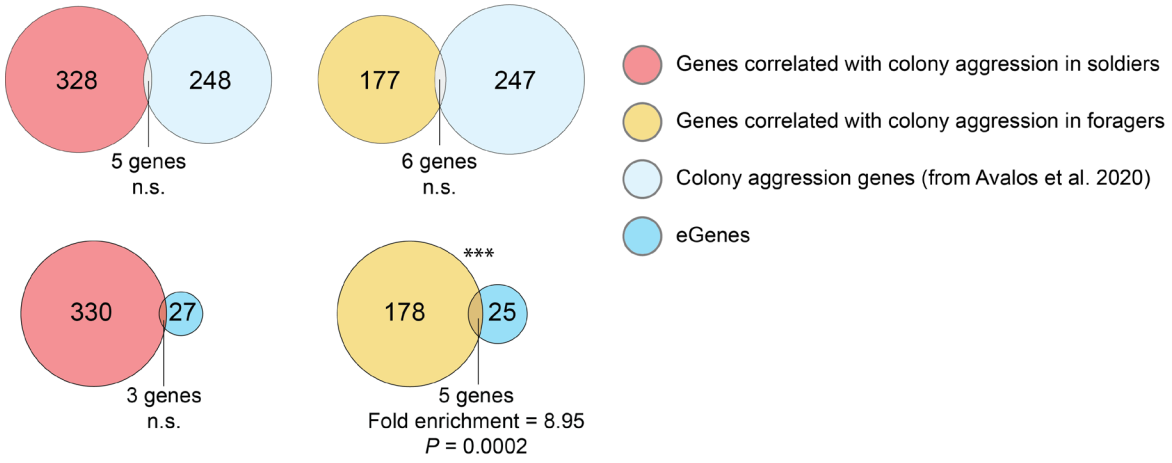
Fig. S3. Single-cell RNA-Sequencing reveals that regulatory dissimilarity, but not differential expression, changes across soldiers and foragers as a function of colony aggression. (A) We divided our scRNA-Seq comparisons of soldiers and foragers by colony aggression level, considering colonies in the top 15% or bottom 25% of the aggression distribution reported in [2] 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

709

710
711 **Fig. S4. Random permutation subnetworks are less connected than eGene subnetwork** (A).
712 Heatmaps as described in Fig. 2 for each of the 15 random permutation subnetworks. (B). Using
713 the three "blocks" of sc clusters identified in Fig. 2, dot charts compare smallest ("minimum")
714 pairwise connectivity for eGene or 15 control subnetworks. For each block, the minimum
715 connectivity was much higher in the eGene subnetwork compared to the 15 controls.
716


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

717

Fig. S5. Molecular correlates of colony aggression identified in soldier and forager whole-tissue brain transcriptomics. (A) We regressed brain gene expression on colony-level aggression for each individual soldier and forager in a given colony, identifying 1037 genes correlated with colony aggression. (B) None of these genes were found to significantly overlap with colony aggression genes from [2] or eGenes identified in this study. (C) Going deeper, we performed the same analysis independently in soldiers and foragers (fold-change on volcano plots as described in [A]) and (D) identified a significantly overlapping set of genes that were (E) highly concordant in terms of differential expression associated with colony aggression. (F) Performing an overlap analysis now resolved to the level of soldiers and foragers identified a significant overlap between genes correlated with colony aggression in foragers and eGenes; no other comparison was significant (hypergeometric test, $P > 0.1$).

764 **References**

765 1. Farine, D. R., Montiglio, P. O. & Spiegel, O. From individuals to groups and back: The
766 evolutionary implications of group phenotypic composition. *Trends Ecol. Evol.* **30**, 609–
767 621 (2015).
768 2. Avalos, A. *et al.* Genomic regions influencing aggressive behavior in honey bees are
769 defined by colony allele frequencies. *Proc. Natl. Acad. Sci.* 201922927 (2020)
770 doi:10.1073/pnas.1922927117.
771 3. Avalos, A. *et al.* A soft selective sweep during rapid evolution of gentle behaviour in an
772 Africanized honeybee. *Nat. Commun.* **8**, 1–9 (2017).
773 4. Wilson, E. O. & Hölldobbler, B. *The superorganism: the beauty, elegance, and*
774 *strangeness of insect societies*. (WW Norton & Company, 2009).
775 5. Robinson, G. E. Regulation of division of labor in insect societies. *Annu. Rev. Entomol.*
776 **37**, 637–665 (1992).
777 6. Amdam, G. V., Norberg, K., Fondrk, M. K. & Page, R. E. Reproductive ground plan may
778 mediate colony-level selection effects on individual foraging behavior in honey bees.
779 *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11350–11355 (2004).
780 7. Darwin 1809-1882, C. *On the origin of species by means of natural selection, or*
781 *preservation of favoured races in the struggle for life*. (London : John Murray, 1859,
782 1859).
783 8. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes
784 generating variation in gene expression. *Nature Reviews Genetics* vol. 22 203–215 (2021).
785 9. Fagny, M. & Austerlitz, F. Polygenic adaptation: Integrating population genetics and gene
786 regulatory networks. *Trends Genet.* **37**, 631–638 (2021).
787 10. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: From
788 polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
789 11. Sinha, S. *et al.* Behavior-related gene regulatory networks: A new level of organization in
790 the brain. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23270–23279 (2020).
791 12. Erwin, D. H. & Davidson, E. H. The evolution of hierarchical gene regulatory networks.
792 *Nat. Rev. Genet. 2009 102* **10**, 141–148 (2009).
793 13. Gao, L. *et al.* Identifying noncoding risk variants using disease-relevant gene regulatory
794 networks. *Nat. Commun. 2018 91* **9**, 1–12 (2018).
795 14. Gustafsson, M. *et al.* A validated gene regulatory network and GWAS identifies early
796 regulators of T cell-associated diseases. *Sci. Transl. Med.* **7**, (2015).
797 15. Peng, J. *et al.* SimiC enables the inference of complex gene regulatory dynamics across
798 cell phenotypes. *Commun. Biol. 2022 51* **5**, 1–19 (2022).
799 16. Traniello, I. M., Chen, Z., Bagchi, V. A. & Robinson, G. E. Valence of social information
800 is encoded in different subpopulations of mushroom body Kenyon cells in the honeybee
801 brain. *Proc. R. Soc. B Biol. Sci.* **286**, 20190901 (2019).
802 17. Traniello, I. M. *et al.* Meta-analysis of honey bee neurogenomic response links Deformed
803 wing virus type A to precocious behavioral maturation. *Sci. Rep.* **10**, 1–12 (2020).
804 18. Sheng, L. *et al.* Social reprogramming in ants induces longevity-associated glia
805 remodeling. *Sci. Adv.* **6**, eaba9869 (2020).
806 19. Fagny, M. *et al.* Exploring regulation in tissues with eQTL networks. *Proc. Natl. Acad.*
807 *Sci. U. S. A.* **114**, E7841–E7850 (2017).
808 20. Chandrasekaran, S. *et al.* Behavior-specific changes in transcriptional modules lead to
809 distinct and predictable neurogenomic states. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18020–5

810 (2011).
811 21. Wang, X. F. *et al.* Apontic directly activates hedgehog and cyclin E for proper organ
812 growth and patterning. *Sci. Reports 2017 71* **7**, 1–13 (2017).
813 22. Boccuto, L. *et al.* Decreased tryptophan metabolism in patients with autism spectrum
814 disorders. **4**, 1–10 (2013).
815 23. de Boer, S. F., Olivier, B., Veening, J. & Koolhaas, J. M. The neurobiology of offensive
816 aggression: Revealing a modular view. *Physiol. Behav.* **146**, 111–127 (2015).
817 24. Hunt, G. J. Flight and fight: A comparative view of the neurophysiology and genetics of
818 honey bee defensive behavior. *J. Insect Physiol.* **53**, 399–410 (2007).
819 25. Davie, K. *et al.* A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell* **174**,
820 982-998.e20 (2018).
821 26. Clandinin, T. R. *et al.* Drosophila LAR regulates R1-R6 and R7 target specificity in the
822 visual system. *Neuron* **32**, 237–248 (2001).
823 27. Wilson, E. O. *The insect societies. The insect societies.* (Harvard University Press, 1971).
824 28. Rivera-Marchand, B., Giray, T. & Guzmán-Novoa, E. The cost of defense in social
825 insects: insights from the honey bee. *Entomol. Exp. Appl.* **129**, 1–10 (2008).
826 29. Hunt, G. J., Guzmán-Novoa, E., Fondrk, M. K. & Page, R. E. Quantitative trait loci for
827 honey bee stinging behavior and body size. *Genetics* **148**, 1203–13 (1998).
828 30. Breed, M. D., Robinson, G. E. & Page, R. E. Division of labor during honey bee colony
829 defense. *Behav. Ecol. Sociobiol.* **27**, 395–401 (1990).
830 31. Avalos, A., Rodríguez-Cruz, Y. & Giray, T. Individual responsiveness to shock and
831 colony-level aggression in honey bees: evidence for a genetic component. *Behav. Ecol.*
832 *Sociobiol.* **68**, 761–771 (2014).
833 32. Shpigler, H. Y. *et al.* Behavioral, transcriptomic and epigenetic responses to social
834 challenge in honey bees. *Genes, Brain Behav.* **16**, 579–591 (2017).
835 33. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for
836 RNA-seq count data. *NAR Genomics Bioinforma.* **2**, (2020).
837 34. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for
838 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40
839 (2010).
840 35. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and
841 powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300
842 (1995).
843 36. Gene Ontology Consortium, T. G. O. Gene Ontology Consortium: Going forward. *Nucleic*
844 *Acids Res.* **43**, D1049-56 (2015).
845 37. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: A tool for discovery
846 and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 1–7
847 (2009).
848 38. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long
849 lists of Gene Ontology terms. *PLoS One* **6**, e21800 (2011).
850 39. Shen, L. GeneOverlap: An R package to test and visualize gene overlaps. http://shenlab-
851 sinai.github.io/shenlab-sinai/ (2014).
852 40. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.
853 *Bioinformatics* **28**, 1353–1358 (2012).
854 41. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome*
855 *Biol. 2016 171* **17**, 1–15 (2016).

42. L, G., O, K. & G, K. Mitochondria: master regulators of danger signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 780–788 (2012).
43. Detmer, S. A. & Chan, D. C. Functions and dysfunctions of mitochondrial dynamics. *Nat. Rev. Mol. Cell Biol. 2007 811* **8**, 870–879 (2007).
44. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21 (2019).
45. Finak, G. *et al.* MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
46. Simola, D. F. *et al.* Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science* **351**, (2016).
47. Croset, V., Treiber, C. D. & Waddell, S. Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *Elife* **7**, e34550 (2018).
48. Konstantinides, N. *et al.* Phenotypic convergence: Distinct transcription factors regulate common terminal features. *Cell* **174**, 622-635.e13 (2018).
49. Li, H. *et al.* Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* **171**, 1206-1220.e22 (2017).
50. Suenami, S., Oya, S., Kohno, H. & Kubo, T. 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).
51. Kretzschmar, D. & Pflugfelder, G. O. Glia in development, function, and neurodegeneration of the adult insect brain. *Brain Res. Bull.* **57**, 121–131 (2002).
52. Shah, A. K., Kreibich, C. D., Amdam, G. V. & Münch, D. Metabolic enzymes in glial cells of the honeybee brain and their associations with aging, starvation and food response. *PLoS One* **13**, e0198322 (2018).
53. Shpigler, H. Y. *et al.* Honey bee neurogenomic responses to affiliative and agonistic social interactions. *Genes, Brain Behav.* **18**, e12509 (2018).
54. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, e12776 (2010).
55. Dijk, D. van *et al.* Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716-729.e27 (2018).