**Sociality emerges from solitary behaviours and reproductive plasticity in the orchid bee *Euglossa dilemma*.**

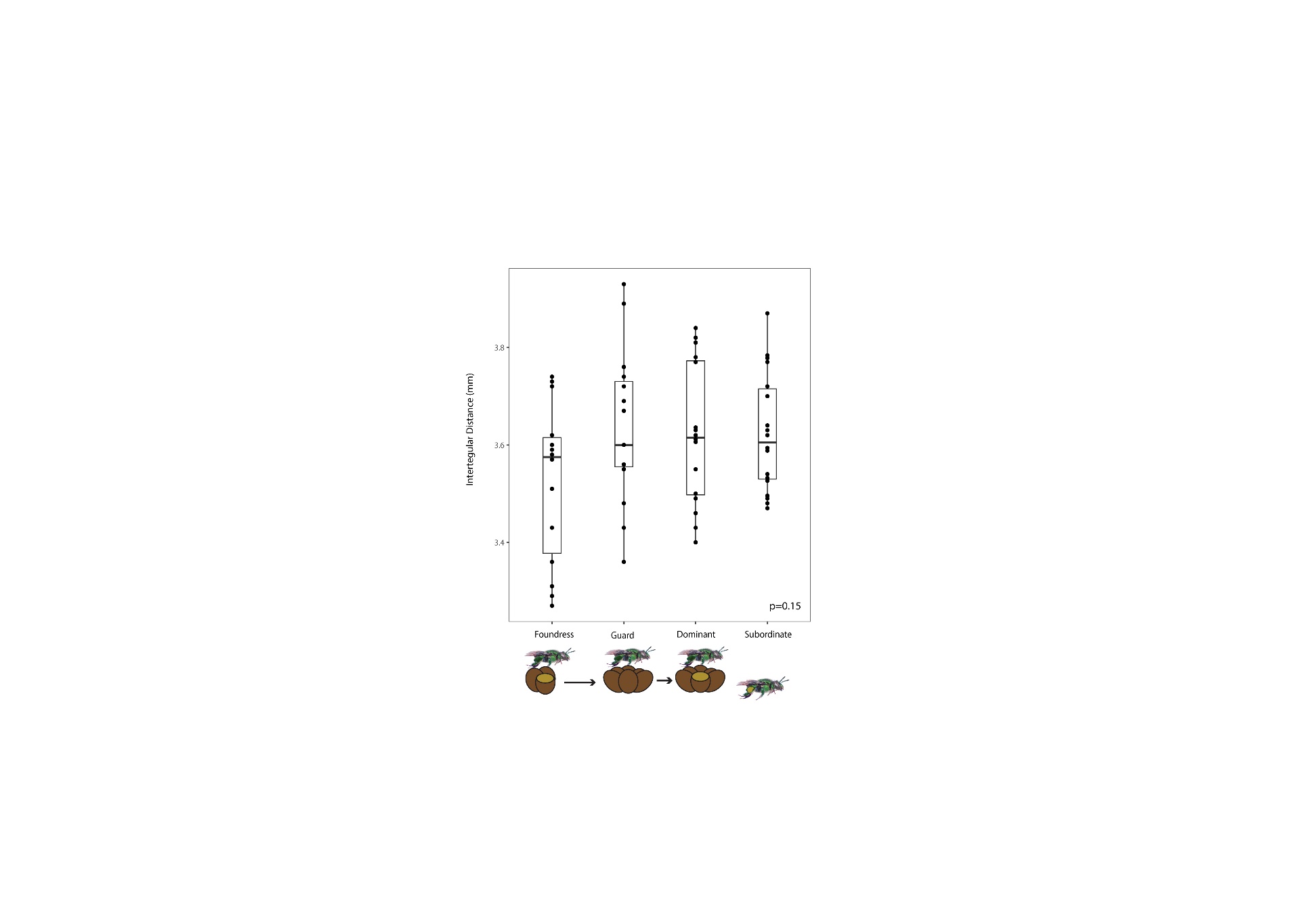
**Nicholas W. Saleh & Santiago R. Ramírez.**

**Proceedings b: 10.1098/rspb20190588**

**Supplementary File 1: Additional tables, figures, and methodology**

**Table S1**. Number of DEGs between all pairwise comparisons in the brains and the ovaries. The first number shows genes upregulated in the first member of a pairwise comparison, the number in parentheses is the number downregulated relative to the first member of that comparison.

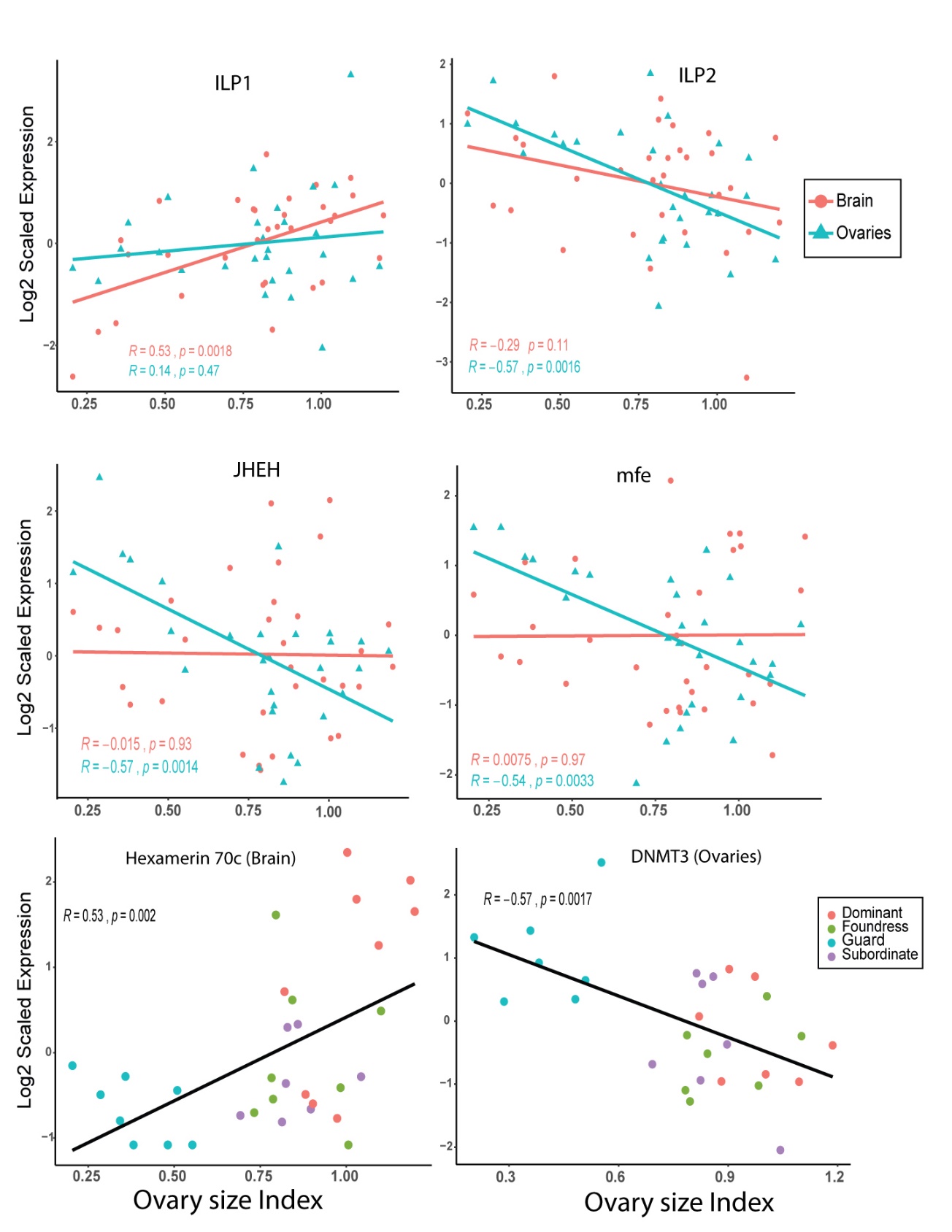
|  |  |  |
| --- | --- | --- |
| **Behavioural Comparison** | **Ovaries** | **Brain** |
| Foundress vs Guard | 971 (1242) | 1880 (2080) |
| Dominant vs Guard | 1162 (1099) | 1445 (1390) |
| Dominant vs Subordinate | 4 (6) | 108 (96) |
| Foundress vs Subordinate | 6 (58) | 18 (52) |
| Subordinate vs Guard | 1293 (1360) | 1697 (1728) |
| Dominant vs Foundress | 69 (21) | 604 (538) |



**Figure S1.**Intertegular distance across behaviours. N = 63 (D = 16, S = 18, G = 15, F = 14). Arrows indicate sequential behaviours.

**Table S2**. Toolkit gene expression correlations with ovary size index for both brain and ovary data sets. Log2 scaled expression levels for each gene were regressed with ovary size index. Bold, underlined values show correlations which are significant after correcting for multiple comparisons. Additional information about toolkit genes is found in supplemental file 2. An “Na” listed indicates that the gene is not in the data set for that tissue.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Gene ID | Annotation | Brain  Pearson r | Brain p-value (fdr corrected) | Ovary Pearson r | Ovary p-value  (fdr corrected) |
| Edil\_03080 | EGFR | -0.1 | 0.88 | 0.013 | 0.95 |
| Edil\_06731 | ILP1 | **0.53** | **0.019** | 0.14 | 0.62 |
| Edil\_08899 | ILP2 | -0.29 | 0.61 | **-0.57** | **0.015** |
| Edil\_05912 | INR2 | -0.17 | 0.78 | 0.24 | 0.46 |
| Edil\_02848 | INR1 | -0.12 | 0.88 | 0.16 | 0.61 |
| Edil\_00144 | TOR | -0.14 | 0.86 | -0.091 | 0.69 |
| Edil\_08553 | RPTOR | **0.58** | **0.013** | 0.14 | 0.62 |
| Edil\_03353 | RICTOR | 0.05 | 0.97 | 0.035 | 0.89 |
| Edil\_00482 | Jhepox | 0.096 | 0.88 | 0.38 | 0.20 |
| Edil\_02026 | JHAMT | 0.011 | 0.97 | Na | Na |
| Edil\_05067 | Mfe | 0.0075 | 0.97 | **-0.54** | **0.021** |
| Edil\_00868 | JHE1 | 0.1 | 0.88 | -0.15 | 0.62 |
| Edil\_02814 | JHEH1 | -0.015 | 0.97 | **-0.57** | **0.015** |
| Edil\_10000 | usp | -0.18 | 0.78 | -0.32 | 0.32 |
| Edil\_03259 | met | -0.17 | 0.78 | -0.091 | 0.69 |
| Edil\_05489 | hex70c | **0.53** | **0.019** | 0.37 | 0.20 |
| Edil\_00209 | hex110 | -0.14 | 0.86 | 0.091 | 0.69 |
| Edil\_05964 | DNMT3 | -0.23 | 0.68 | **-0.57** | **0.015** |
| Edil\_09875 | dsx | -0.02 | 0.97 | -0.38 | 0.20 |
| Edil\_05644 | Br-C | -0.31 | 0.60 | 0.21 | 0.50 |
| Edil\_00668 | E74 | -0.27 | 0.61 | 0.29 | 0.40 |
| Edil\_03109 | ECR | -0.24 | 0.63 | 0.24 | 0.46 |
| Edil\_10699 | Vitellogenin | 0.25 | 0.63 | 0.26 | 0.45 |
| Edil\_05241 | Corazonin | 0.21 | 0.7 | Na | Na |
| Edil\_10944 | Yellow b | -0.034 | 0.97 | -0.19 | 0.55 |
| Edil\_10127 | Yellow x2 | -0.034 | 0.97 | -0.25 | 0.46 |
| Edil\_01676 | Yellow g | -0.065 | 0.97 | -0.12 | 0.67 |
| Edil\_01682 | Yellow g2 | -0.03 | 0.91 | -0.21 | 0.50 |



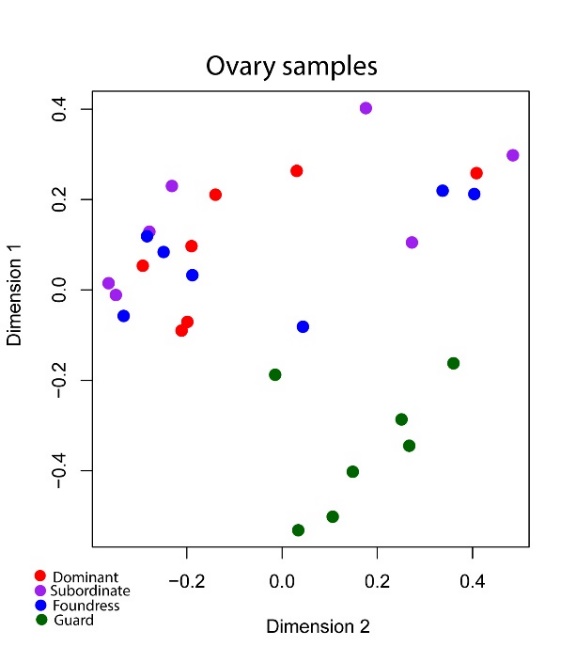
**Figure S2.** Gene expression patterns for a select set of toolkit genes that exhibit correlation with ovary size index. The X-axis on all panels shows ovary size index and the Y-axis shows log2 scaled expression. The top 4 panels show brain and ovary gene expression data plotted together. The bottom two panels show only the brain data (left panel) and only the ovary data (right panel). Pearson r correlation values are shown. P-values are from the original correlations and are not adjusted for multiple comparisons. Adjusted p-values are displayed in table S2. All significant correlations shown in the panel remain significant after correction in table S2.

**Additional discussion of batch correction and differential expression analysis.**

First, we discuss our overall approach with batch correction and then discuss the actions we take in our analysis. A priori, we expected that our four original pilot samples would introduce batch effects because they were prepared first and underwent deeper sequencing. Further, we expected that our collection trips, dissections, RNA extractions, and library preparations could all be possible sources of batch effects. While we attempted to always balance our behavioural groups equally across these batches, failure of several samples at each of these steps resulted in unbalanced batches.

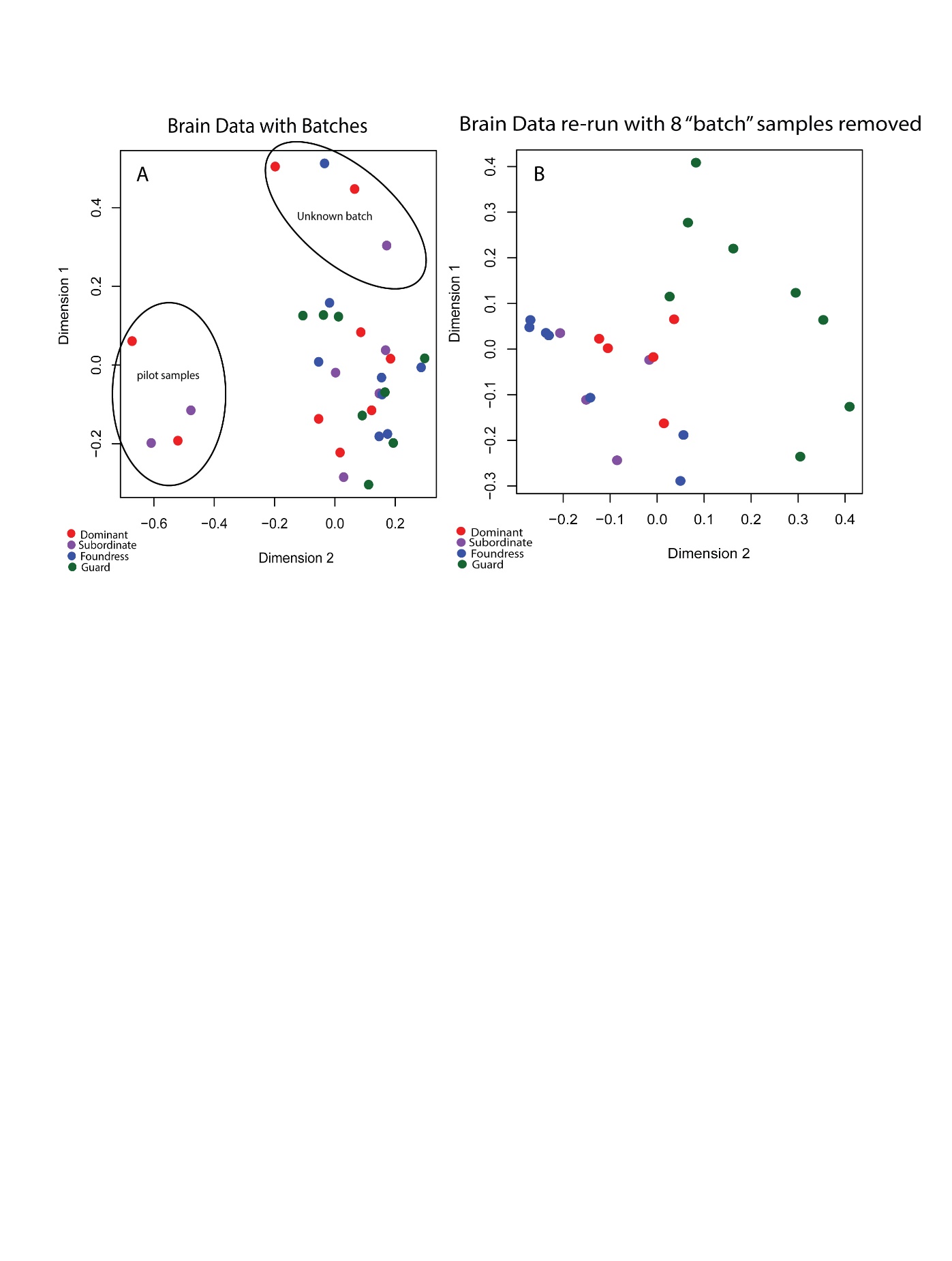
Consequently, our batch correction approach is as follows: (1) Do we see putative batch effects in sample clustering that are of unknown origin or do we observe an unbalanced distribution of samples (and thus have no corresponding balanced variable that could be included in analysis)? (2) if yes, we proceed with SV (surrogate variable) analysis. (3) Do the SVs returned correspond to the possible batches that we identify in sample clustering? (4) If yes, do these SVs also correlate with possible sources of batch effects we expected to see a prioi? (5) If yes, we include SVs in the edgeR model if they correspond to the batches identified in sample clustering and are correlated with known batch effects. This process led us to include two SVs with the brain data and none with the ovary data, which we discuss in further detail below.

Initially, we performed sample clustering using MDS. For the ovaries we found that behavioural group was a primary source of variation in the data set (Fig. S3). Consequently, no additional batch correction information was added to the edgeR model for this data set.



**Figure S3.** MDS plot from edgeR following gene set filtering for the ovaries. Samples cluster primarily due to behaviour (guards vs others).

Brain samples, however, showed two potential batches of individuals that resulted in the largest sources of variation (Fig. S4A). We note that these batches are not balanced across behavioural groups, which is not ideal and requires further investigation to determine the source of the batch effect. Given that most individuals are clustered in one area without clear correlation with behaviour, we first wanted to determine if the variation introduced with the two batches was likely obscuring true behavioural clustering. Consequently, we removed the 8 samples involved in the batches and re-ran sample clustering. When we remove these 8 samples, we see that guard individuals now clearly separate from the other groups, forming the largest source of variation in this smaller data set (Fig. S4B). This suggested that batch effects were obscuring real differences associated with behaviour and that failure to correct for them (if they are the result of technical differences) would result in a large drop in power or a possible distortion of the results of our differential expression analysis.

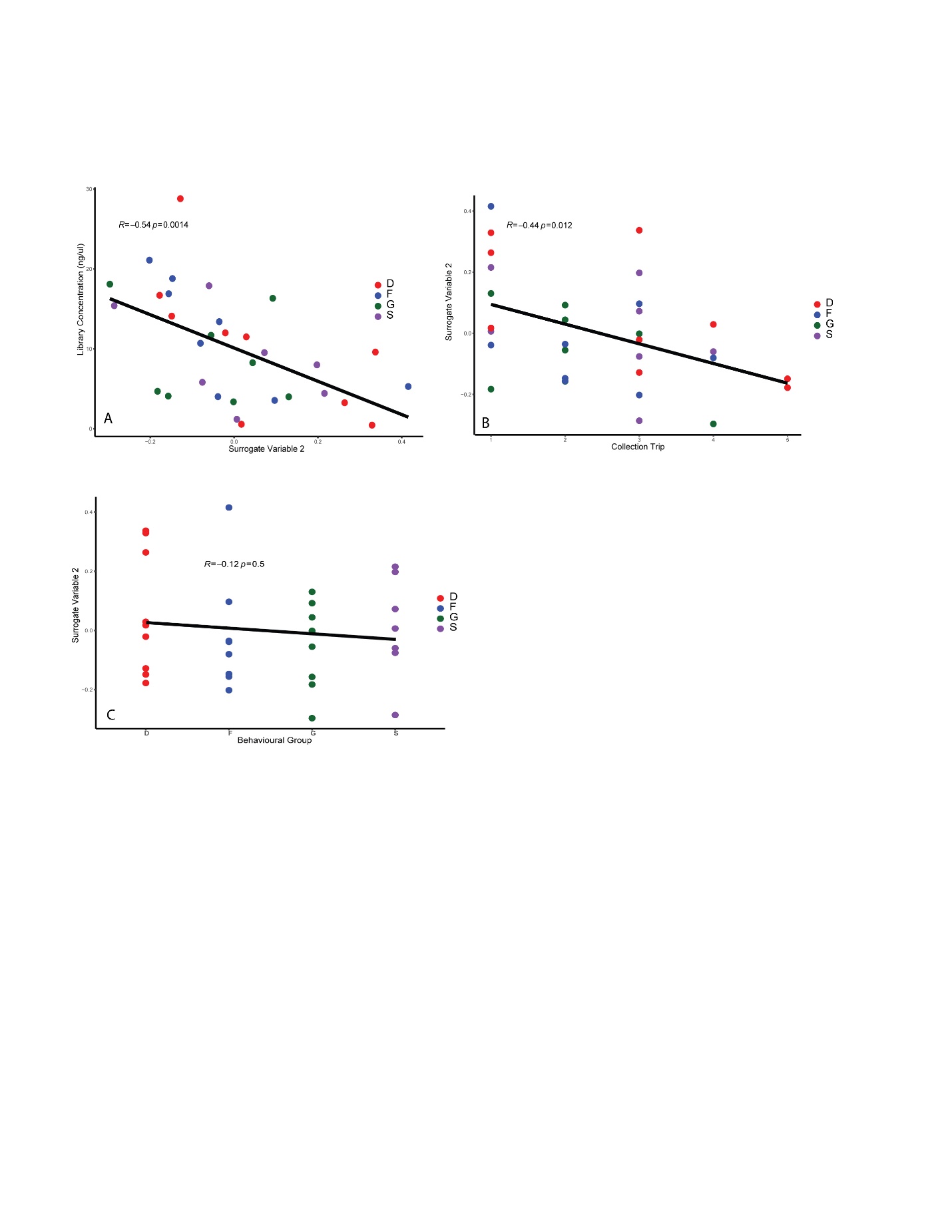


**Figure S4.** MDS plots from edgeR following gene set filtering for the Brain samples. (A) Initial clustering showed three main clusters. The two potential batch effects, one known and one unknown are shown in ellipses. (B) We re-ran the MDS analysis removing the 8 “batch” samples to see if this resulted in behavioural group as the primary source of variation.

Next, we began to examine the sources of variation that might lead to these batches. One batch clearly corresponded to our four initial pilot samples and so we decided to include the corresponding SV in the edgeR model. However, the second putative batch effect was unknown and so we examined the correlation between the SV capturing this axis of variation and our known batches. We found that this SV was significantly correlated with library concentration (r = -0.54, p = 0.0014, Fig. S5A), sampling trip (r = -0.44, p = 0.012, Fig. S5B), and library prep kit (r = -0.36, 0.045). While we realize that this doesn’t preclude a group bias in these measures, we do note that none of these measures show a significant behavioural group difference (ANOVA p-values: p = 0.78, p = 0.59, p = 0.72, respectively). Consequently, though they are not perfectly balanced, they are also not strongly confounded with our behavioural groups of interest. Similarly, the SV is not significantly different among behavioural groups or significantly correlated with them (ANOVA of SV and behavioural group: p = 0.66; correlation shown in Fig. S5C).

We note that these correlations do not capture all the variation in the SV. However, given that the SV is estimated directly from the data, we think it is reasonable that these correlations are suggestive of real technical effects that were generated over the course of sample collection and library preparation, especially since library preparation was done manually in small batches in the lab and our measurements of library concentration likely contain measurement error. Since these sources of variation are known, we included this SV in the edgeR model as well as the SV corresponding to our pilot samples.

We also note that ovary samples underwent the same basic collection and sample preparation techniques, but they do not appear to be as strongly affected as the brain data set. This is not surprising since the brain dissections yielded much less tissue with a higher failure rate in RNA extraction and library preparation than the ovary samples. For example, glycogen needed to be added to the Trizol brain RNA extractions to increase yield, which was not an issue with ovary samples. Consequently, we expect brain samples to be more sensitive to technical effects, as we were working with much less material overall.



**Figure S5**. Correlations between the unknown surrogate variable and (A) library concentration (B) sampling trip and (C) behavioural group. D = dominant, F = foundress, G = guard, S = subordinate.

**Analysis with and without the addition of Surrogate Variables**

While the SVs appear to represent technical variation, we wanted to further investigate the effect that including these SVs had on the results of the differential expression analysis. Therefore, we conducted an additional analysis of the brain data set excluding the SVs from the model and instead included both library concentration and RNA concentration in the model, as these were our only continuous variables that might capture technical variation in an unbiased fashion. Consequently, this analysis could capture aspects of one of the SVs, but not the one associated with the pilot samples. The edgeR output for this analysis for all pairwise comparisons can be found in supplemental file 4 along with the results of the SV model, for direct comparison. Overall, we find that inclusion of the SVs does result in a substantial gain in DEGs across all behaviours (we detect 5,446 DEGs in the brain compared to 3,249 without the SVs). While this is a significant difference in the total number of DEGs, this is not surprising since our pilot samples represent a large source of variation in the dataset (Fig. S4A) and are not accounted for, thus resulting in lower statistical power to detect DEGs. Further, while we find many quantitative differences between these two approaches, the qualitative results are largely similar with and without SVs. We see thousands of genes differentially expressed between guards and other behavioural groups and a smaller subset between the reproductive behavioural groups. Clustering patterns for foraging/non-foraging individuals based on DEGS between dominants and subordinates is also similar with either approach (we find 51 DEGs without SVs), with 30/32 samples falling into predicted clusters.

Finally, we calculated a correlation between log fold change values for DEGs from both models since this could reveal fundamental differences in expression pattern when adding the SVs. Consequently, for each of the 6 pairwise comparisons possible, we looked at the R2 value between the log fold change values for DEGS from the model with SVs and those same genes from the model without the SVs. Our expectation is that high R2 values indicate that the relationship between the expression patterns is similar between the two models, and that the higher number of DEGs with SVs in the model reflects greater statistical power due to an increase in the explained variance in the data. Overall, we find that 5 of 6 comparisons show high R2 values >0.94, with one comparison (Dominants vs Guards), showing a slightly lower correlation R2 = 0.86 (Table S3). Generally, this indicates high correspondence between patterns of expression when using the two models.

To further investigate the reason for a slightly lower R2 value for the Dominant and Guard comparison we also compared the log fold change values from the two model outputs (with and without SVs) to the results of this same comparison (Dominant vs Guard) when run without the 8 batch samples, as seen in figure S4B. Our expectation is that a high correlation between log fold change values with and without the batch samples suggests similar patterns of gene expression that is reflective of true differences. We find that the model with SVs added and the model with the 8 batch samples removed show R2 = 0.92 compared to R2 = 0.87 without the SVs. Consequently, addition of the SVs seems to be as accurate, if not slightly more accurate in capturing the expression patterns underlying sample clustering in figure S4B, when the 8 batch samples are removed.

**Table S3.** Correlations between the log fold change values for DEGs from the edgeR model with SVs and those same genes from the model without the addition of SVs.

|  |  |
| --- | --- |
| **Pairwise Comparison** | **R2 value between the two models** |
| Foundress vs Guard | 0.94 |
| Dominant vs Subordinate | 0.97 |
| Dominant vs Guard | 0.86 |
| Dominant vs Foundress | 0.97 |
| Foundress to Subordinate | 0.97 |
| Subordinate to Guard | 0.94 |