

1 *Chapter 1*

2 **Gene expression responses to diet**
3 **quality and viral infection in *Apis***
4 ***mellifera***

5 **1.1 Introduction**

6 Commerically managed honeybees have undergone unusually large declines in the United
7 States and parts of Europe over the past decade (van Engelsdorp et al. 2009, Kulhanek et al.
8 2017, Laurent et al. 2016), with annual mortality rates exceeding what beekeepers consider
9 sustainable (Caron and Sagili 2011, Bond et al. 2014). More than 70 percent of major
10 global food crops (including fruits, vegatables, and nuts) at least benefit from pollination,
11 and yearly insect pollination services are valued wordwide at \$175 billion (Gallai et al.
12 2009). As honeybees are largely considered to be the leading pollinator of numerous crops,
13 their marked loss has considerable implications regarding agricultural sustainability (Klein
14 et al. 2007).

15 Honeybee declines have been associated with several factors, including pesticide use,
16 parasites, pathogens, habitat loss, and poor nutrition (Potts et al. 2010, Spivak et al. 2011).
17 Researchers generally agree that these stressors do not act in isolation; instead, they appear
18 to influence the large-scale loss of honeybees in interactive fashions as the environment
19 changes (Goulson et al. 2015). Nutrition and viral infection are two broad factors that pose
20 heightened dangers to honeybee health in response to recent environmental changes.

21 Pollen is the main source of nutrition (including proteins, amino acids, lipids, sterols,
22 starch, vitamins, and minerals) in honeybees (Roulston and Buchmann 2000, Stanley and
23 Linskens 1974). At the individual level, pollen supplies most of the nutrients necessary
24 for physiological development (Brodschneider and Crailsheim 2010) and is believed to
25 have considerable impact on longevity (Haydak 1970). At the colony level, pollen enables

26 young workers to produce jelly, which then nourishes larvae, drones, older workers, and the
27 queen (Crailsheim et al. 1992, Crailsheim 1992). Various environmental changes (including
28 urbanization and monoculture crop production) have significantly altered the nutritional
29 profile available to honeybees. In particular, honeybees are confronted with less diverse
30 selections of pollen, which is of concern because mixed-pollen (polyfloral) diets are generally
31 considered healthier than single-pollen (monofloral) diets (Schmidt 1984, Schmidt et al. 1987,
32 Alaux et al. 2010). Indeed, reported colony mortality rates are higher in developed land
33 areas compared to undeveloped land areas (Naug 2009), and beekeepers rank poor nutrition
34 as one of the main reasons for colony losses (Engelsdorp et al. 2008). Understanding how
35 undiversified diets affect honeybee health will be crucial to resolve problems that may arise
36 as agriculture continues to intensify throughout the world (Neumann and Carreck 2010,
37 Engelsdorp and Meixner 2010).

38 Viral infection was a comparatively minor problem in honeybees until the last century when
39 Varroa destructor (an ectoparasitic mite) spread worldwide (Rosenkranz et al. 2010). This
40 mite feeds on honeybee hemolymph (Weinberg and Madel 1985), transmits cocktails of
41 viruses, and supports replication of certain viruses (Shen et al. 2005, Yang and Cox-Foster
42 2007, Yang and Cox-Foster 2005). More than 20 honeybee viruses have been identified (Chen
43 and Siede 2007). One of these viruses that has been linked to honeybee decline is Israeli
44 Acute Paralysis Virus (IAPV). A positive-sense RNA virus of the Dicistroviridae family
45 (Miranda et al. 2010), IAPV causes infected honeybees to display shivering wings, decreased
46 locomotion, muscle spasms, and paralysis, and 80% of caged infected adult honeybees die
47 prematurely (Maori et al. 2009). IAPV has demonstrated higher infectious capacities
48 than other honeybee viruses in certain conditions (Carrillo-Tripp et al. 2016) and is more
49 prevalent in colonies that do not survive the winter (Chen et al. 2014). Its role in the rising
50 phenomenon of “Colony Collapse Disorder” (in which the majority of worker bees disappear
51 from a hive) remains unclear: It has been implicated in some studies (Cox-Foster et al.
52 2007, Hou et al. 2014) but not in other studies (van Engelsdorp et al. 2009, Cornman et al.
53 2012, Miranda et al. 2010). Nonetheless, it seems likely that IAPV reduces colony strength
54 and survival.

55 Although there is growing interest in how viruses and diet quality affect the health and
56 sustainability of honeybees, as well as a recognition that such factors might operate
57 interactively, there are only a small number of experimental studies thus far directed toward
58 elucidating the interactive effects of these two factors in honeybees (DeGrandi-Hoffman and
59 Chen 2015, DeGrandi-Hoffman et al. 2010, Conte et al. 2011). We recently used laboratory
60 cages and nucleus hive experiments to investigate the health effects of these two factors,
61 and our results show a significant interaction between diet quality and virus infection.
62 Specifically, high quality pollen is able to mitigate virus-induced mortality to the level of
63 diverse, polyfloral pollen (Dolezal et al. 2018).

Following up on these phenotypic findings from our previous study, we now aim to understand the corresponding underlying mechanisms by which high quality diets protect bees from virus-induced mortality. For example, it is not known whether the protective effect of good diet is due to direct, specific effects on immune function (resistance), or if it is due to indirect effects of good nutrition on energy availability and vigor (resilience). Transcriptomics is one means to achieve this goal. Transcriptomic analysis can help us identify 1) the genomic scale of transcriptomic response to diet and virus infection, 2) whether these factors interact in an additive or synergistic way on transcriptome function, and 3) the types of pathways affected by diet quality and viral infection. This information, heretofore lacking in the literature, can help us better understand how good nutrition may be able to serve as a "buffer" against other stressors (Dolezal and Toth 2018). As it stands, there are only a small number of published experiments examining gene expression patterns related to diet effects (Alaux et al. 2011) and IAPV infection effects (Galbraith et al. 2015) in honeybees. As far as we know, there are few to no studies investigating honeybee gene expression patterns specifically related to monofloral diets, and few to no studies investigating honeybee gene expression patterns related to the interaction effects of diet in any broad sense and viral inoculation in any broad sense.

In this study, we examine how monofloral diets and viral inoculation influence gene expression patterns in honeybees by focusing on four treatment groups (low quality diet without IAPV exposure, high quality diet without IAPV exposure, low quality diet with IAPV exposure, and high quality diet with IAPV exposure). We conduct RNA-sequencing analysis on a randomly selected subset of the honeybees we used in our previous study (as is further described in our methods section). We then examine pairwise combinations of treatment groups, the main effect of monofloral diet, the main effect of IAPV exposure, and the interactive effect of the two factors on gene expression patterns.

We also compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues (Galbraith et al. 2015). As RNA-sequencing data can be highly noisy, this comparison allowed us to characterize how repeatable and robust our RNA-seq results were in comparison to previous studies. Importantly, we use an in-depth data visualization approach to explore and validate our data, and suggest such an approach can be useful for cross-study comparisons of RNA-sequencing data in the future.

1.2 Methods

Details of the procedures we used to prepare virus inoculum, infect and feed caged honeybees, and quantify IAPV can be reviewed in our previous work (Dolezal et al. 2018). The statistical analysis we used to study the main and interaction effects of the two factors on mortality and IAPV titers is also described in our earlier report (Dolezal et al. 2018).

101 **1.2.1 Design of two-factor experiment**

102 There are several reasons why, in the current study, we focused only on diet quality
103 (monofloral diets) as opposed to diet diversity (monofloral diets versus polyfloral diets).
104 First, when assessing diet diversity, a sugar diet is often used as a control. However,
105 such an experimental design does not reflect real-world conditions for honeybees as they
106 rarely face a total lack of pollen ([Pasquale et al. 2013](#)). Second, in studies that compared
107 honeybee health using monofloral and polyfloral diets at the same time, if the polyfloral
108 diet and one of the high-quality monofloral diets both exhibited similarly beneficial effects,
109 then it was difficult for the authors to assess if the polyfloral diet was better than most
110 of the monofloral diets because of its diversity or because it contained as a subset the
111 high-quality monofloral diet ([Pasquale et al. 2013](#)). Third, colonies used for pollination in
112 agricultural areas (monoculture) face less diversified pollens (according to Brodschneider,
113 2010). Pollinating areas are currently undergoing landscape alteration and agriculture
114 intensification, and bees are increasingly faced with less diversified diets (monoculture)
115 ([Decourtye et al. 2010](#), [Brodschneider and Crailsheim 2010](#)). As a result, there is a need to
116 better understand how monofloral diets affect honeybee health as a step toward mitigating
117 the negative impact of human activity on the honeybee population.

118 Consequently, for our nutrition factor, we examined two monofloral pollen diets, Rockrose
119 (Cistus) and Chestnut (Chestnut). Rockrose pollen is generally considered less nutritious
120 than Chestnut pollen due to its lower levels of protein, amino acids, antioxidants, calcium,
121 and iron ([Pasquale et al. 2013](#), [Dolezal et al. 2018](#)). For our virus factor, one level contained
122 bees that were infected with IAPV and another level contained bees that were not infected
123 with IAPV. This experimental design resulted in four treatment groups (Rockrose pollen
124 without IAPV exposure, Chestnut pollen without IAPV exposure, Rockrose pollen with
125 IAPV exposure, and Chestnut pollen with IAPV exposure) that allowed us to assess main
126 effects and interactive effects between diet quality and IAPV infection in honeybees.

127 **1.2.2 RNA extraction**

128 Fifteen cages per treatment were originally sampled. Six live honeybees from each cage
129 were randomly selected 36 hours post inoculation and placed into tubes. Tubes were kept
130 on dry ice and then transferred into a -80C freezer until processing. Eight cages were
131 randomly selected from the original 15 cages, and 2 honeybees per cage were randomly
132 selected from the original six live honeybees per cage. Whole body RNA from each pool of
133 two honeybees were extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase
134 treatment. Samples were suspended in water to 200-400 ng/ μ l. All samples were then
135 tested on a Bioanalyzer at the DNA core facility to ensure quality (RIN>8).

136 **1.2.3 Gene expression**

137 Samples were sequenced starting on January 14, 2016 at the Iowa State University DNA
138 Facility (Platform: Illumina HiSeq Sequencing; Category: Single End 100 cycle sequencing).
139 A standard Illumina mRNA library was prepared by the DNA facility. Reads were aligned
140 to the BeeBase Version 3.2 genome ([Consortium 2014](#)) from the Hymenoptera Genome
141 Database ([Elsik et al. 2016](#)) using the programs GMAP and GSNAp ([Wu et al. 2016](#)). We
142 tested all six pairwise combinations of treatments for DEGs (pairwise DEGs). We also
143 tested the diet main effect (diet DEGs), virus main effect (virus DEGs), and interaction
144 term for DEGs (interaction DEGs). We then also tested for virus main effect DEGs (virus
145 DEGs) in public data derived from a previous study exploring the gene expression of
146 IAPV virus infection in honeybees ([Galbraith et al. 2015](#)). We tested each DEG analysis
147 using recommended parameters with DESeq2 ([Love et al. 2014](#)), edgeR ([Robinson et al.
148 2010](#)), and LimmaVoom ([Ritchie et al. 2015](#)). In all cases, we used a false discovery rate
149 (FDR) threshold of 0.05 ([Benjamini and Hochberg 1995](#)). Fisher's exact test was used to
150 determine significant overlaps between DEG sets (whether from the same dataset but across
151 different analysis pipelines or from different datasets across the same analysis pipelines).
152 The `eulererr` shiny application was used to construct Venn diagram overlap images ([Larsson
153 2018](#)). In the main section of our paper and in subsequent analyses, we focus on the DEG
154 results from DESeq2 ([Love et al. 2014](#)) as this pipeline was also used in the Galbraith study
155 ([Galbraith et al. 2015](#)).

156 @@@ What percent of reads mapped? @@@ Total number of raw reads @@@ How many
157 lanes @@@ How many samples per lane

158 **1.2.4 Comparison to previous studies on transcriptomic response to viral
159 infection**

160 We also compare the main effect of IAPV exposure in our dataset to that obtained in a
161 previous study conducted by Galbraith and colleagues ([Galbraith et al. 2015](#)) who also
162 addressed honey bee transcriptomic responses to virus infection.

163 While our study examines honeybees from polyandrous colonies, the Galbraith study
164 examined honeybees from single-drone colonies. As a consequence, the honeybees in our
165 study will be on average 25% genetically identical, whereas honeybees from the Galbraith
166 study will be on average 75% genetically identical ([Page and Laidlaw 1988](#)). We should
167 therefore expect that the Galbraith study may generate data with lower signal:to:noise
168 ratios than our data due to the lower genetic variation between its replicates. At the same
169 time, our honeybees will be more likely to display the health benefits gained from increased
170 genotypic variance within colonies, including decreased parasitic load ([Sherman et al. 1988](#)),
171 increased tolerance to environmental changes ([Crozier and Page 1985](#)), and increasead colony
172 performance ([Mattila and Seeley 2007](#), [Tarpay 2003](#)). Given that honeybees are naturally

173 very polyandrous ([Brodschneider et al. 2012](#)), our honeybees may also reflect more realistic
174 environmental and genetic simulations. Taken together, each study provides a different point
175 of value: Our study likely presents less artificial data while the Galbraith data likely presents
176 less messy data. We wish to explore how the gene expression effects of IAPV inoculation
177 compare between these two studies that used such different experimental designs. To
178 achieve this objective, we use visualization techniques to assess the signal:to:noise ratio
179 between these two datasets, and differential gene expression (DEG) analyses to determine
180 any significantly overlapping genes of interest between these two datasets. It is our hope
181 that this aspect of our study may shine light on how experimental designs that control
182 genetic variability to different extents might affect the resulting gene expression data in
183 honeybees.

184 **1.2.5 Visualization**

185 We used visualization tools from the DESeq2 package, visualization tools from our work-in-
186 progress bigPint package, and visual inference techniques to assess the signal:to:noise ratio
187 in the datasets and to assess the suitability of the DEG calls.

188 **1.2.6 Gene Ontology**

189 DEGs were uploaded as a background list to DAVID Bioinformatics Resources 6.7 ([Huang](#)
190 [et al. 2009a](#), [Huang et al. 2009b](#)). The overrepresented gene ontology (GO) terms of DEGs
191 were identified using the BEEBASE_ID identifier. To fine-tune the GO term list, only
192 terms correlating to Biological Processes were considered. The refined GO term list was
193 then imported into REVIGO ([Supek et al. 2011](#)), which uses semantic similarity measures
194 to cluster long lists of GO terms.

195 **1.2.7 Detecting resilience versus resistance**

196 To investigate whether the protective effect of good diet is due to direct, specific effects
197 on immune function (resistance), or if it is due to indirect effects of good nutrition on
198 energy availability and vigor (resilience), we created contrasts of interest (Table 1.10). In
199 particular, we assigned "resistance candidate genes" to be the ones that were upregulated
200 in the Chestnut group within the virus infected bees but not upregulated in the Chestnut
201 group within the non-infected bees. We also assigned "resilience candidate genes" to be
202 the ones that were upregulated in the Chestnut group for both the virus infected bees
203 and non-infected bees. Our interpretation of these genes is that they represent genes that
204 are constitutively activated in bees fed a high quality diet, regardless of whether they are
205 experiencing infection or not. We then determined how many genes fell into these two
206 categories and analyzed their GO terminologies.

207 **1.3 Results**208 **1.3.1 Phenotypic results**

209 We reanalyzed our previously published dataset with a subset more relevant to our RNA-
210 sequencing approaches in the current study that have a more focused question regarding
211 diet quality. We briefly show it again here to inform the RNA-seq comparison because we
212 reduced the number of treatments (from eight to four) from the original published data
213 ([Dolezal et al. 2018](#)).

214 Mortality rates of honeybees 72 hour post-inoculation significantly differed among the
215 treatment groups (mixed model ANOVA across all treatment groups, $df=3, 55; F=10.07$;
216 $p<2.18e-05$). The effect of virus treatment (mixed model ANOVA, $df=1, 55; F=24.343$;
217 $p<7.84e-06$) and diet treatment (mixed model ANOVA, $df=1, 55; F=5.796; p<0.0194$)
218 were significant, but the interaction between the two factors (mixed model ANOVA, $df=1,$
219 $55; F=0.062, p=0.8039$) was not significant. The virus treatment was significant: For a
220 given diet, honeybees exposed to the virus showed significantly higher mortality rate than
221 honeybees not exposed to the virus (Tukey HSD, $p<0.05$). In comparing mortality levels
222 based on pairwise comparisons, we found that bees fed Rockrose pollen had significantly
223 elevated mortality with virus infection compared to uninfected controls. However, bees
224 fed Castanea pollen had no significant difference in mortality between virus infected and
225 control groups. These results suggest the high quality Castanea diet can “rescue” virus
226 induced mortality (Figure 1.1A).

227 IAPV titer volumes of honeybees 72 hour post-inoculation significantly differed among the
228 treatment groups (mixed model ANOVA across all treatment groups, $df=3, 34; F=6.096$;
229 $p<0.00196$). The effect of virus treatment (mixed model ANOVA, $df=1, 34; F=15.686$;
230 $p<0.000362$) was significant, but the diet treatment (mixed model ANOVA, $df=1, 34;$
231 $F=1.898; p>0.05$) and the interaction between the two factors (mixed model ANOVA, $df=1,$
232 $34; F=0.702, p>0.05$) were not significant. Honeybees that were infected with the virus
233 and fed a poor-quality Rockrose diet showed significant increases in IAPV titer volumes
234 compared to honeybees that were not infected with the virus regardless of their diet quality
235 (Tukey HSD, $p<0.05$). Overall, we interpreted this effect to mean that Rockrose pollen could
236 not “rescue” high virus titers resulting from the inoculation treatment, whereas Castanea
237 pollen could (Figure 1.1B).

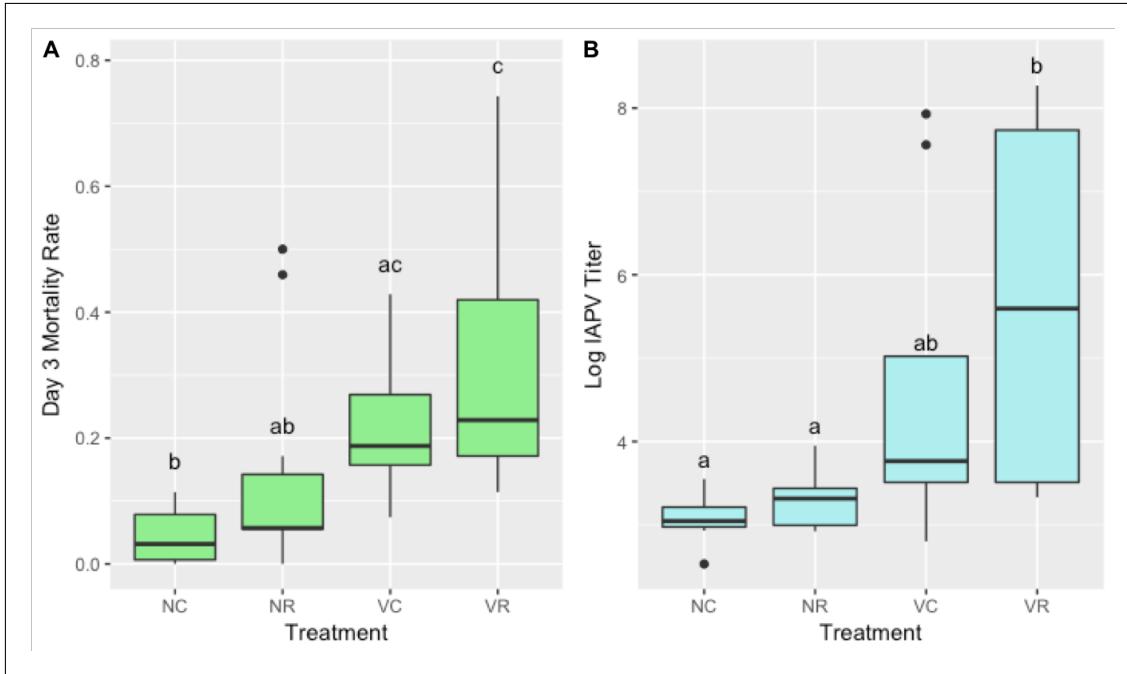


Figure 1.1: Mortality rates (A) and IAPV titers (B) for the four treatment groups. “N” represent non-inoculation, “V” represents viral inoculation, “C” represents Chestnut pollen, and “R” represents Rockrose pollen. The mortality rate data included 59 samples with 15 replicates per treatment group, except for the “NC” group having 14 replicates. The IAPV titer data included 38 samples with 10 replicates per treatment group, except for the “NR” group having 8 replicates. ANOVA values and p-values for the statistical tests are listed in the text of the paper. The letters above the bars represent Tukey honest significant differences with a confidence level of 95%.

238 1.3.2 Main effect DEG results

239 We observed a substantially larger number of DEGs in our diet main effect ($n = 1914$) than
240 in our virus main effect ($n = 43$) (Table 1.1A and B). In the diet factor, there were more
241 Chestnut-upregulated DEGs ($n = 1033$) than Rockrose-upregulated DEGs ($n = 881$). In
242 the virus factor, there were more virus-upregulated DEGs ($n = 38$) than control-upregulated
243 DEGs ($n = 5$). While these reported DEGs numbers are from the DESeq2 package, we saw
244 similar trends for the edgeR and limma package results (Table 1.1A and B).

245 GO analysis of the Chestnut-upregulated DEGs revealed the following enriched categories
246 (Benjamini correction < 0.05): Wnt signaling, hippo signaling, and dorso-ventral axis forma-
247 tion, as well as pathways related to circadian rhythm, mRNA surveillance, insulin resistance,
248 inositol phosphate metabolism, FoxO signaling, ECM-receptor interaction, phototrans-
249 duction, Notch signaling, Jak-STAT signaling, MAPK signaling, and carbon metabolism
250 (Table 1.2). GO analysis of the Rockrose DEGs revealed pathways related to terpenoid
251 backbone biosynthesis, homologous recombination, SNARE interactions in vesicular trans-

252 port, aminoacyl-tRNA biosynthesis, Fanconi anemia, and pyrimidine metabolism (Table
253 1.3).

254 With so few DEGs ($n = 43$) in our virus main effect study, we focused on individual genes
255 and their known functionalities (Table 1.5). Of the 43 virus-related DEGs, only 10 had GO
256 assignments within the DAVID database. These genes had implications in the recognition
257 of pathogen-related lipid products and the cleaving of transcripts from viruses, as well
258 as involvement in ubiquitin and proteosome pathways, transcription pathways, apoptotic
259 pathways, oxidoreductase processes, and several more functions (Table 1.5).

260 No interaction DEGs were observed between the diet and virus factors of the study, in any
261 of the pipelines (DESeq2, edgeR, limma).

262 1.3.3 Pairwise comparison DEG results

263 The number of DEGs across the six treatment pairings between the diet and virus factor
264 ranged from 0 to 941 (Table 1.4). Some of the trends observed in the main effect comparisons
265 persisted: The diet level appeared to have greater influence on the number of DEGs than
266 the virus level. Across every pair comparing the Chestnut and Rockrose levels, regardless
267 of the virus level, the number of Chestnut-upregulated DEGs was higher than the number
268 of Rockrose-upregulated DEGs (Table 1.4 C, D, E, F). For the pairs in which the diet level
269 was controlled, the virus-exposed treatment showed equal to or more DEGs than the control
270 treatment (Table 1.4 A, B). There were no DEGs between the treatment pair controlling
271 for the control level of the virus effect (Table 1.4 A). These trends were observed for all
272 three pipelines used (DESeq2, edgeR, and limma).

273 1.3.4 Comparison with Galbraith study

274 We wished to explore the signal:to:noise ratio between the Galbraith dataset and our
275 dataset. Basic MDS plots were constructed with the DESeq2 analysis pipeline, and we
276 could immediately determine that the Galbraith dataset may better separate the infected
277 and uninfected honeybees better than our dataset (Figure 1.6). We also noted that the
278 first replicate of both treatment groups in the Galbraith data did not cluster as cleanly in
279 the MDS plots. However, through this automatically-generated plot, we can only visualize
280 information at the sample level. Wanting to learn more about the data at the gene level,
281 we continued with additional visualization techniques.

282 We used parallel coordinate lines superimposed onto boxplots to visualize the DEGs
283 associated with virus infection in the two studies. The background boxplot represents
284 the distribution of all genes in the data, and each parallel coordinate line represents one
285 DEG. To reduce overlapping of parallel coordinate lines, we often use hierarchical clustering
286 techniques to separate DEGs into common patterns. See more information about this

287 plotting method and the ideal visual structure of DEGs in our earlier chapter @@@.

288 We see that the 1,019 DEGs from the Galbraith dataset form relatively clean-looking visual
289 displays (Figure 1.2). We do see that the first replicate of the virus group appears somewhat
290 inconsistent with the other virus replicates in Cluster 2, confirming that this trend in the
291 data that we saw in the MDS plot carried through into the DEG results. In contrast, we see
292 that the 43 virus-related DEGs from our dataset do not look as clean in their visual displays
293 (Figure 1.3). The replicates appear somewhat inconsistent in their esimated expression
294 levels and there is not always such a large difference between treatment groups. We see a
295 similar finding when we also examine a larger subset of 1,914 diet-related DEGs from our
296 study.

297 We also used litre plots to examine the structure of individual DEGs: We see that indeed
298 the individual DEGs from our data (Figure 1.9) show less consistent replications and
299 less differences between the treatment groups compared to the individual DEGs from the
300 Galbraith data (Figure 1.8 and Figure 1.7). For the Galbraith data, we examined individual
301 DEGs from the first cluster (Figure 1.8) and second cluster (Figure 1.7) because the second
302 cluster was a bit less ideal due to its inconsistent first replicate of the treatment group.

303 Finally, we looked at scatterplot matrices to assess the DEGs. We created standardized
304 scatterplot matrices for each of the four clusters (Figure 1.2) of the Galbraith data (Figures
305 1.10, 1.11, 1.12, and 1.13). We also created standardized scatterplot matrices for our data.
306 However, as our dataset contained 24 samples, we would need to include 276 scatterplots in
307 our matrix, which would be too numerous to allow for efficient visual assessment of the
308 data. As a result, we created four scatterplot matrices of our data, each with subsets of 6
309 samples to be more comparable to the Galbraith data (Figures 1.14, 1.15, 1.16, and 1.17).
310 We can again confirm through these plots that the DEGs from the Galbraith data appeared
311 more as expected: Deviating more from the $x=y$ line in the treatment scatterplots while
312 staying close to the $x=y$ line in replicate scatterplots.

313 Despite the DEGs from the Galbraith dataset displaying the expected patterns more than
314 those from our dataset, there was significant overlap in the DEGs between the two studies
315 (Figure 1.5).

316 1.3.5 Resilience versus resistance

317 Within our “resilience” gene ontologies, we found functions related to metabolism (such
318 as carbohydrate metabolism, fructose metabolism, and chitin metabolism). However, we
319 also discovered gene ontologies related to RNA polyerase II transcription and immune
320 response (Figure 1.18A). Within our “resistance” gene ontologies, we found functions
321 related to metabolism (such as carbohydrate metabolism, chitin metabolism, and general
322 metabolism). (Figure 1.18B).

323 **1.4 Discussion**

324 1) Diet effect seemed larger by number of DEGs than virus effect

325

326 2) No interaction terms in DEGs

327

328 3) Elaborate more on the GOs that came out of different contrasts

329

330 4) Possible preliminary evidence for resilience effects, whereby immune-related genes are

331 turned on by good diet in preparation to defend against viral attacks

332

333 5) Galbraith data was much cleaner as expected due to 75 percent genetic similarity.

334 However, we had significant overlap in the virus-related DEGs. Possible avenue for future

335 studies to explore as the intersection DEGs between these two studies may represent those

336 that are reliable (well-controlled study) and relevant (study conducted in more realistic

337 environmental settings).

338 **1.5 Appendix**

A	OUR DIET EFFECT	C higher	R higher	Total
DESeq2	1033	881	1914	
EdgeR	889	832	1721	
Limma	851	789	1640	

B	OUR VIRUS EFFECT	V higher	C higher	Total
DESeq2	38	5	43	
EdgeR	17	3	20	
Limma	0	0	0	

C	GALBRAITH VIRUS EFFECT	V higher	C higher	Total
DESeq2	795	224	1019	
EdgeR	580	150	730	
Limma	193	20	213	

Table 1.1: Number of DEGs across three analysis pipelines for (A) the diet effect in our study, (B) the virus main effect in our study, and (C) the virus main effect in the Galbraith study.

CHAPTER 1. GENE EXPRESSION RESPONSES TO DIET QUALITY AND VIRAL
12 INFECTION IN APIS MELLIFERA

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	15	2.20E-03	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, C-terminal-binding protein, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
Dorso-ventral axis formation	8	2.80E-02	<i>CUGBP Elav-like family member 2, ETS-like protein pointed, cytoplasmic polyadenylation element-binding protein 2, encore, epidermal growth factor receptor-like, neurogenic locus Notch protein, protein giant-lease, protein son of sevenless</i>
Hippo signaling pathway	12	3.00E-02	<i>actin, cadherin-related tumor suppressor, casein kinase I-like, cisks large tumor suppressor protein, division abnormally delayed protein, hemicentin-2, protein dachsous, protein expanded-like, stress-activated protein kinase JNK</i>
Circadian rhythm	4	2.40E-01	<i>casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor</i>
mRNA surveillance pathway	10	2.60E-01	<i>cleavage and polyadenylation specificity factor subunit CG7185, eukaryotic peptide chain release factor GTP-binding subunit ERF3A, heterogeneous nuclear ribonucleoprotein 27C, polyadenylate-binding protein 1, regulator of nonsense transcripts 1, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform</i>
Insulin resistance	8	2.80E-01	<i>insulin-like receptor-like (InR-2), long-chain fatty acid transport protein 1, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform, stress-activated protein kinase JNK, tyrosine-protein phosphatase non-receptor type 61F-like</i>
Inositol phosphate metabolism	8	2.90E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II, inositol oxygenate, methylmalonate-semialdehyde dehydrogenase (acylating)-like protein, multiple inositol polyphosphate phosphatase 1-like, myotubularin-related protein 4, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, uncharacterized oxidoreductase YrbE-like</i>
FoxO signaling pathway	9	3.00E-01	<i>casein kinase I-like, epidermal growth factor receptor-like, histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of seven less, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
ECM-receptor interaction	5	3.20E-01	<i>agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like</i>
Phototransduction	6	3.30E-01	<i>1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase</i>
Notch signaling pathway	5	3.80E-01	<i>C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb</i>
Jak-STAT signaling pathway	4	3.90E-01	<i>histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless</i>
MAPK signaling pathway	4	4.40E-01	<i>epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS</i>
Carbon metabolism	12	4.50E-01	<i>2-oxoglutarate dehydrogenase, aminomethyltransferase, fructose-bisphosphate aldolase, glycine dehydrogenase (decarboxylating), L-threonine ammonia-lyase, methylmalonate-semialdehyde dehydrogenase [acylating]-like protein, NADP-dependent malic enzyme, probable aconitate hydratase, PTS-dependent dihydroxyacetone kinase, pyruvate carboxylase, succinate dehydrogenase [ubiquinone] iron-sulfur subunit</i>

Table 1.2: Pathways related to diet main effect Chestnut-upregulated DEGs.

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	15	2.20E-03	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, C-terminal-binding protein, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
Dorso-ventral axis formation	8	2.80E-02	<i>CUGBP Elav-like family member 2, ETS-like protein pointed, cytoplasmic polyadenylation element-binding protein 2, encore, epidermal growth factor receptor-like, neurogenic locus Notch protein, protein giant-lease, protein son of sevenless</i>
Hippo signaling pathway	12	3.00E-02	<i>actin, cadherin-related tumor suppressor, casein kinase I-like, cisks large tumor suppressor protein, division abnormally delayed protein, hemicentin-2, protein dachsous, protein expanded-like, stress-activated protein kinase JNK</i>
Circadian rhythm	4	2.40E-01	<i>casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor</i>
mRNA surveillance pathway	10	2.60E-01	<i>cleavage and polyadenylation specificity factor subunit CG7185, eukaryotic peptide chain release factor GTP-binding subunit ERF3A, heterogeneous nuclear ribonucleoprotein 27C, polyadenylate-binding protein 1, regulator of nonsense transcripts 1, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform</i>
Insulin resistance	8	2.80E-01	<i>insulin-like receptor-like (InR-2), long-chain fatty acid transport protein 1, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform, stress-activated protein kinase JNK, tyrosine-protein phosphatase non-receptor type 61F-like</i>
Inositol phosphate metabolism	8	2.90E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II, inositol oxygenate, methylmalonate-semialdehyde dehydrogenase (acylating)-like protein, multiple inositol polyphosphate phosphatase 1-like, myotubularin-related protein 4, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, uncharacterized oxidoreductase YrbE-like</i>
FoxO signaling pathway	9	3.00E-01	<i>casein kinase I-like, epidermal growth factor receptor-like, histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of seven less, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
ECM-receptor interaction	5	3.20E-01	<i>agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like</i>
Phototransduction	6	3.30E-01	<i>1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase</i>
Notch signaling pathway	5	3.80E-01	<i>C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb</i>
Jak-STAT signaling pathway	4	3.90E-01	<i>histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless</i>
MAPK signaling pathway	4	4.40E-01	<i>epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS</i>
Carbon metabolism	12	4.50E-01	<i>2-oxoglutarate dehydrogenase, aminomethyltransferase, fructose-bisphosphate aldolase, glycine dehydrogenase (decarboxylating), L-threonine ammonia-lyase, methylmalonate-semialdehyde dehydrogenase [acylating]-like protein, NADP-dependent malic enzyme, probable aconitate hydratase, PTS-dependent dihydroxyacetone kinase, pyruvate carboxylase, succinate dehydrogenase [ubiquinone] iron-sulfur subunit</i>

Table 1.3: Pathways related to diet main effect Rockrose-upregulated DEGs.

A	OUR PAIRS (NC, VC)	NC higher	VC higher	Total
DESeq2		0	0	0
EdgeR		0	0	0
Limma		0	0	0

B	OUR PAIRS (NR, VR)	VR higher	NR higher	Total
DESeq2		152	26	178
EdgeR		87	9	96
Limma		0	0	0

C	OUR PAIRS (VC, VR)	VC higher	VR higher	Total
DESeq2		247	129	376
EdgeR		130	59	189
Limma		10	1	11

D	OUR PAIRS (NC, VR)	NC higher	VR higher	Total
DESeq2		496	278	774
EdgeR		320	215	535
Limma		108	47	155

E	OUR PAIRS (VC, NR)	VC higher	NR higher	Total
DESeq2		540	415	955
EdgeR		431	251	682
Limma		140	91	231

F	OUR PAIRS (NC, NR)	NC higher	NR higher	Total
DESeq2		601	340	941
EdgeR		502	295	797
Limma		219	139	358

Table 1.4: Number of DEGs across three analysis pipelines for all six treatment pair combinations between the diet and virus factor.

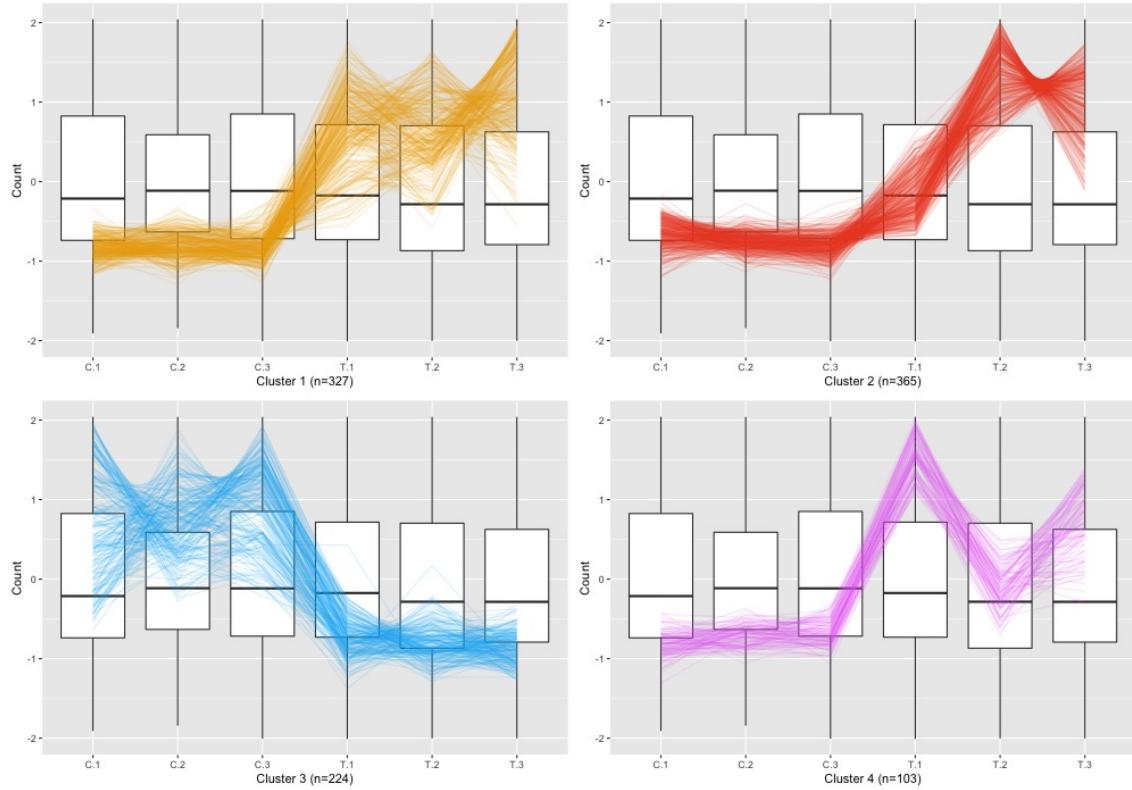


Figure 1.2: Parallel coordinate plots of the 1,019 DEGs after hierarchical clustering of size four between the virus-infected and control groups of the Galbraith study. Here “C” represents control, and “T” represents treatment of virus. Clusters 1, 3, and 4 seem to represent DEGs that were overexpressed in the virus inoculated group, and Cluster 2 seems to represent DEGs that were overexpressed in the control group. In general, the DEGs appeared as expected, but there is rather noticeable deviation of the first replicate from the virus-treated sample (“T.1”) from the other virus-treated replicates in Cluster 2. Cluster 4 also has some inconsistent replicates across the virus-treated replicates.

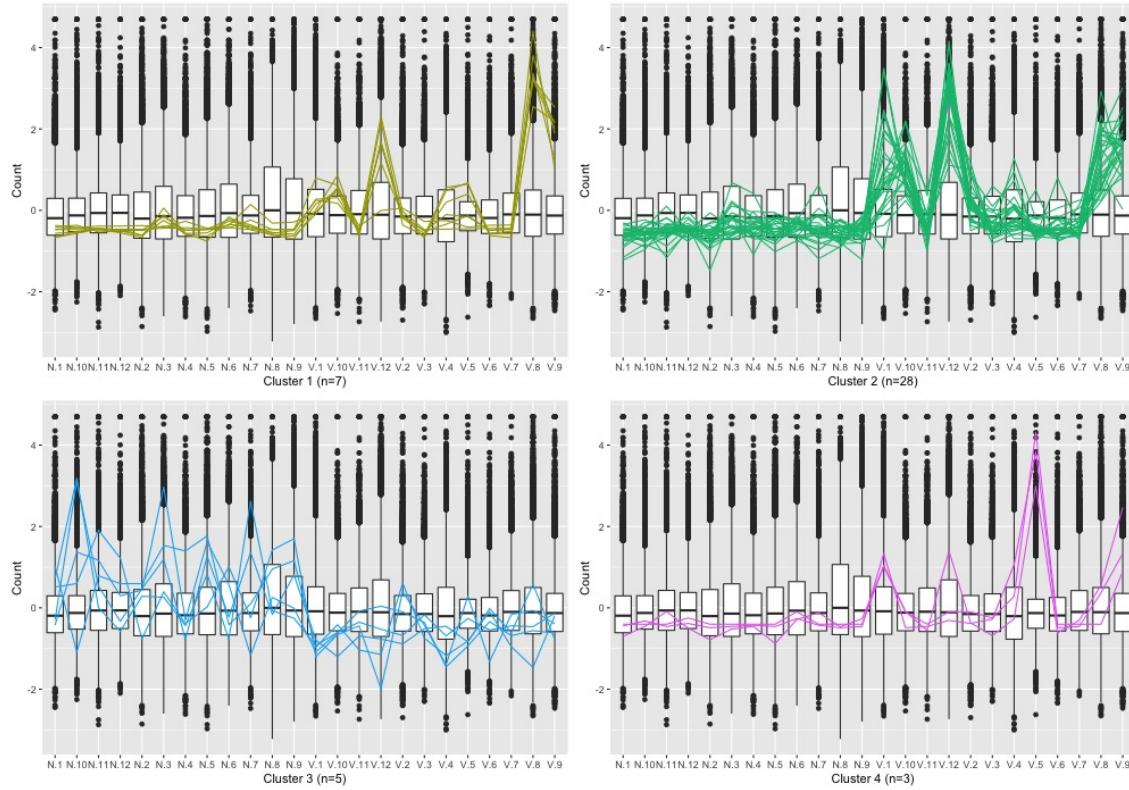


Figure 1.3: Parallel coordinate plots of the 43 DEGs after hierarchical clustering of size four between the virus-infected and control groups of our study. Here “N” represents non-infected control group, and “V” represents treatment of virus. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 1.2.

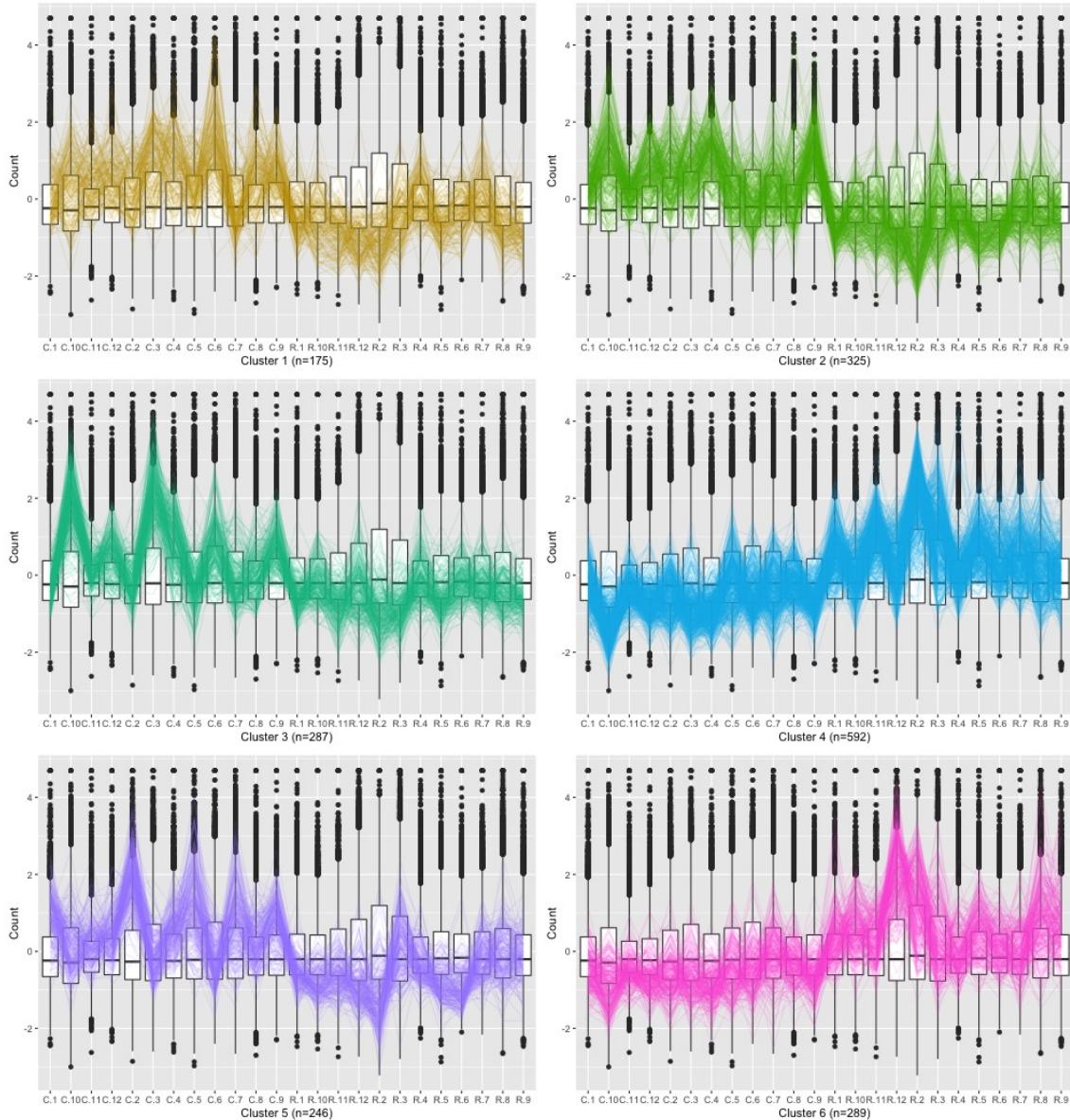


Figure 1.4: Parallel coordinate plots of the 1,914 DEGs after hierarchical clustering of size six between the Chestnut and Rockrose groups of our study. Here “N” represents non-infected control group, and “V” represents treatment of virus. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 1.2.

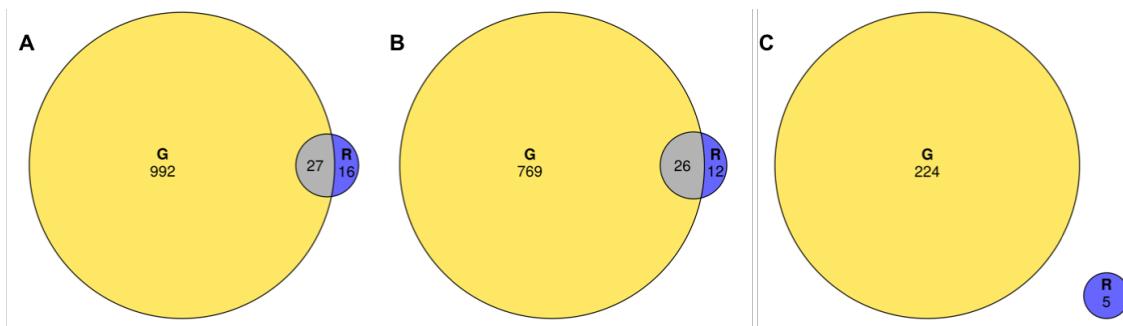


Figure 1.5: Venn diagrams comparing the virus-related DEG overlaps between the Galbraith study (labeled as “G”) and our study (labeled as “R”). From left to right: Total virus-related DEGs (subplot A), virus-upregulated DEGs (subplot B), control-upregulated DEGs (subplot C). Both the total virus-related and virus-upregulated DEGs showed significant overlap between the studies ($p\text{-value} < 2.2\text{e-}16$) as per Fisher’s Exact Test for Count Data. There was one gene that was virus-upregulated in the Galbraith study but control-upregulated in our study.

BeeBase ID	Gene Name	Known functions	Our DEG Group	Galbraith DEG Group
GB41545	MD-2-related lipid-recognition protein-like	<i>Implicated in lipid recognition, particularly in the recognition of pathogen related products</i>	N	-
GB50955	Protein argonaute-2	<i>Interacts with small interfering RNAs to form RNA-induced silencing complexes, which target and cleave transcripts that are mostly from viruses and transposons</i>	V	V
GB48755	UBA-like domain-containing protein 2	<i>Found in diverse proteins involved in ubiquitin/proteasome pathways</i>	V	V
GB47407	Histone H4	<i>Capable of affecting transcription, DNA repair, and DNA replication when post-transcriptionally modified</i>	V	V
GB42313	Leishmanolysin-like peptidase	<i>Encodes a protein involved in cell migration and invasion; implicated in mitotic progression in <i>D. melanogaster</i></i>	V	V
GB50813	Rho guanine nucleotide exchange factor 11	<i>Implicated in regulation of apoptotic processes, cell growth, signal transduction, and transcription</i>	V	V
GB54503	Thioredoxin domain-containing protein	<i>Serves as a general protein disulphide oxidoreductase</i>	N	-
GB53500	Transcriptional regulator Myc-B	<i>Regulator gene that codes for a transcription factor</i>	V	V
GB51305	Tropomyosin-like	<i>Related to protein involved in muscle contraction</i>	N	N
GB50178	Cilia and flagella-associated protein 61-like	<i>Includes components required for wild-type motility and stable assembly of motile cilia</i>	V	V

Table 1.5: Known functions of the mapped subset of 43 DEGs in the virus main effect of our study. Whether the gene was overrepresented in the virus or non-virus group is also indicated for both our study and the Galbraith study. Functionalities were extracted from Flybase, National Center for Biotechnology Information, and The European Bioinformatics Institute databases.

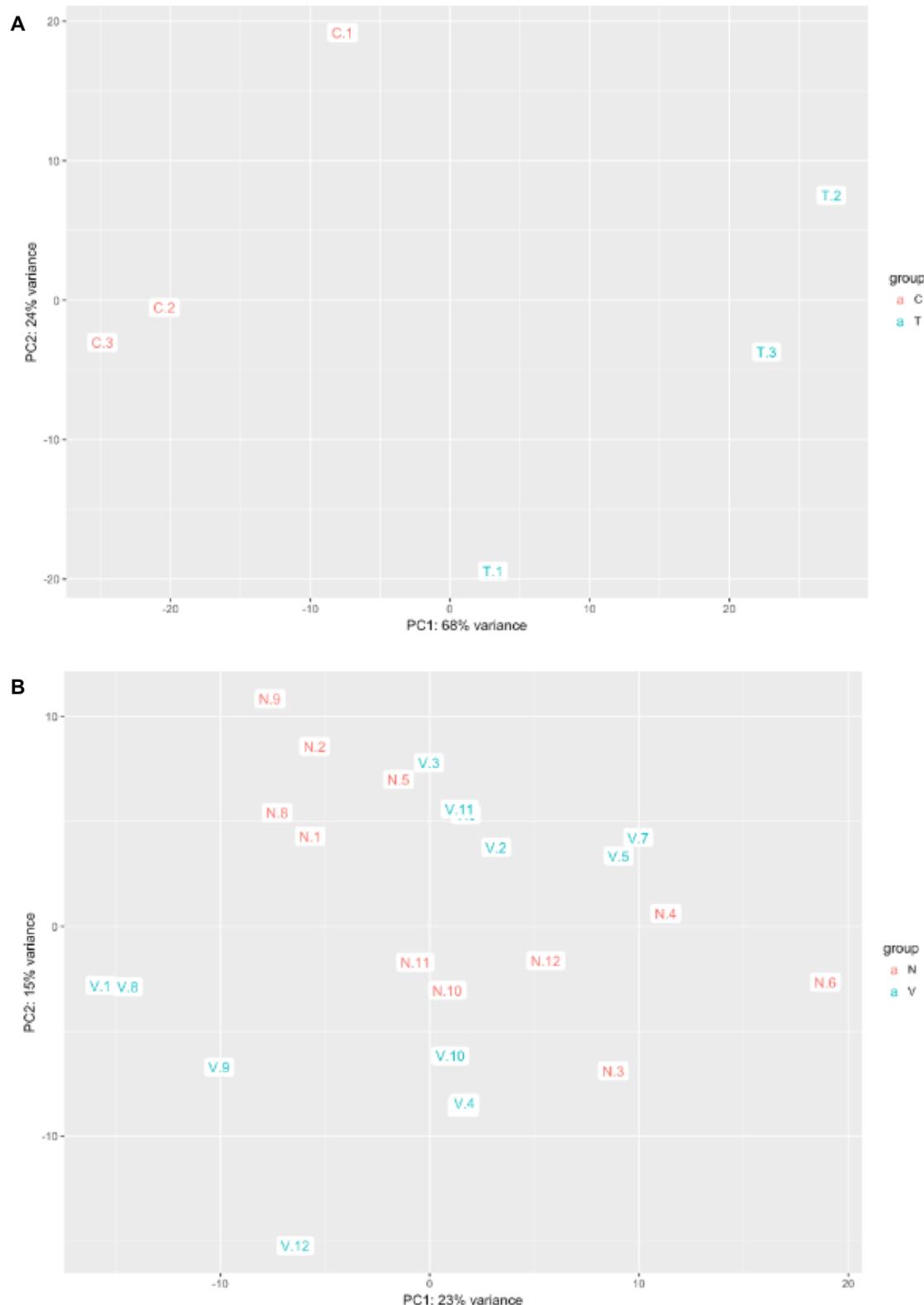


Figure 1.6: MDS plots constructed from DESeq2 package for the Galbraith dataset for non-infected control “C” and virus treated “T” samples (A) and our dataset for the non-infected control “N” and virus treated “V” samples (B). the x-axis represents the principal component with the most variation and the y-axis represents the principal component with the second-most variation.

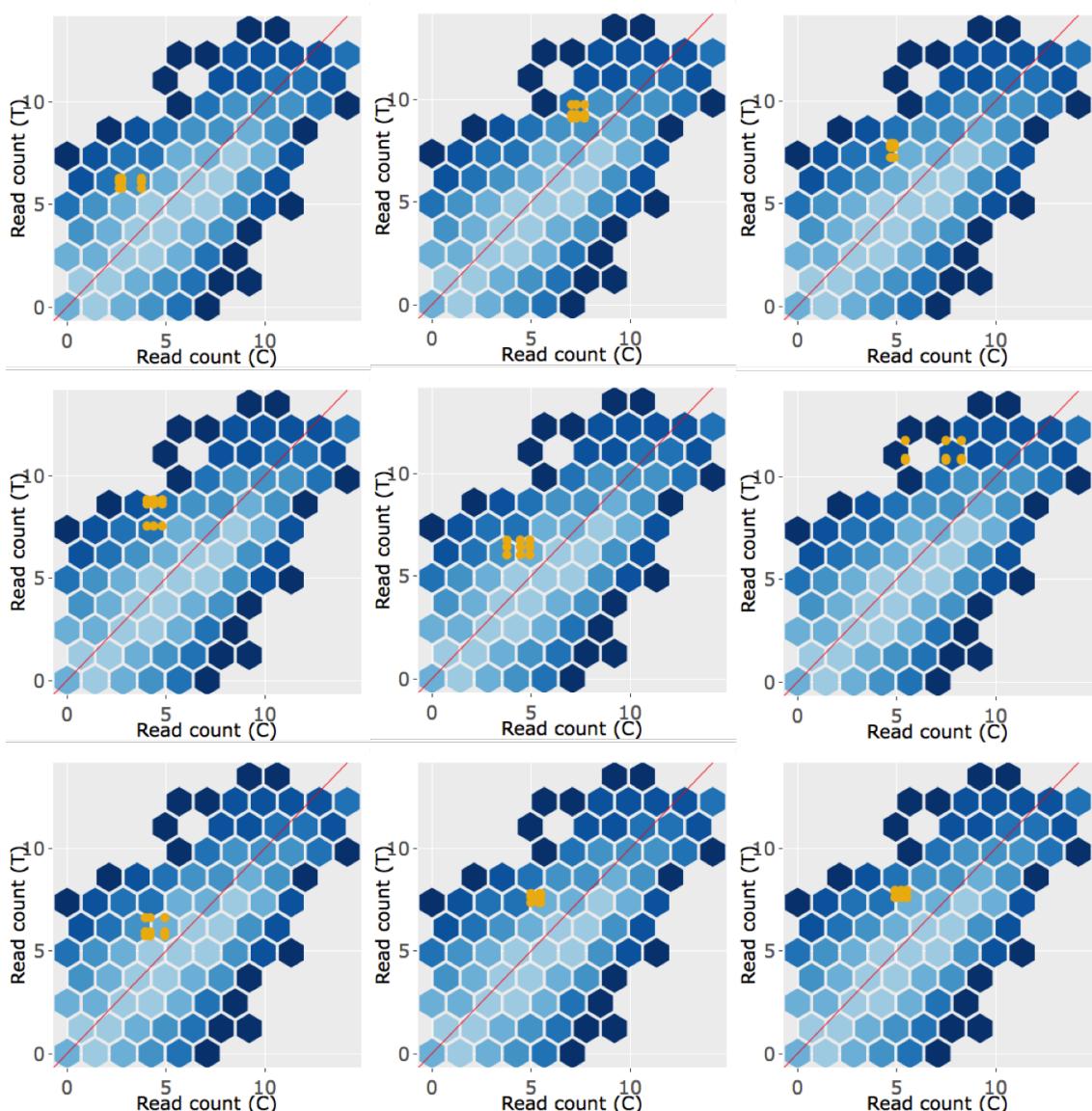


Figure 1.7: Example litre plots of the nine DEGs with the lowest FDR values from Cluster 1 (originally shown in Figure 1.2) of the Galbraith dataset. “C” represents non-infected control samples and “T” represents virus-treated samples. Most of the light orange points (representing the nine combinations of samples between treatment groups for a given DEG) deviate from the $x=y$ line in a cluster as expected.

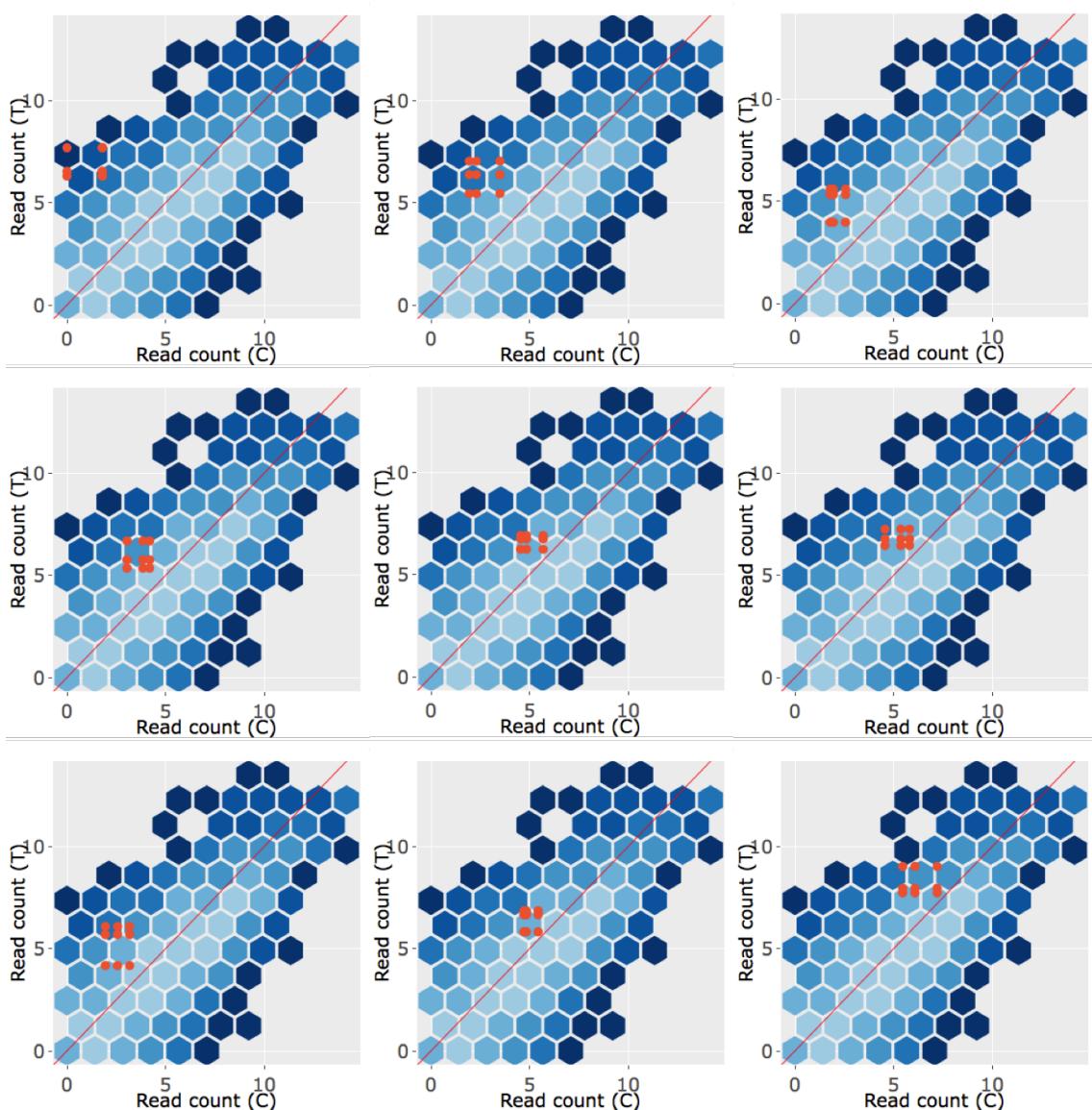


Figure 1.8: Example litre plots of the nine DEGs with the lowest FDR values from Cluster 2 (originally shown in Figure 1.2) of the Galbraith dataset. “C” represents non-infected control samples and “T” represents virus-treated samples. Most of the dark orange points (representing all combinations of samples between treatment groups for a given DEG) deviate from the $x=y$ line in a cluster as expected. However, they are not as tightly clustered together compared to what we saw in the example litre plots of Cluster 1 (shown in Figure 1.7). As a result, what we see in these litre plots reflects what we saw in the parallel coordinate lines of Figure 1.2: The replicate consistency in the Cluster 2 DEGs is not as clean as that in the Cluster 1 DEGs, but is still relatively clean.

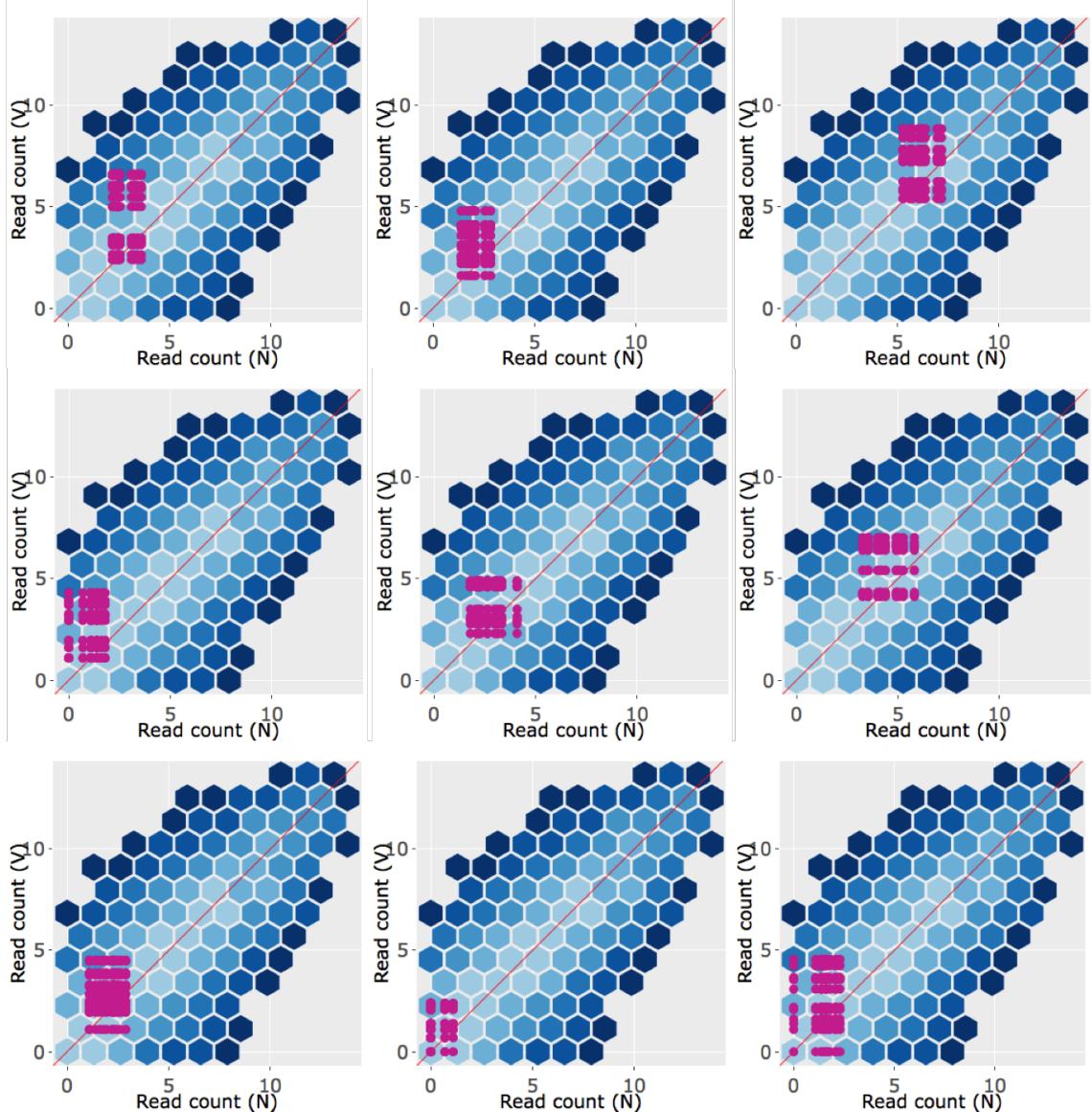


Figure 1.9: Example litre plots of the nine DEGs with the lowest FDR values from the 43 DEGs of our dataset. “N” represents non-infected control samples and “V” represents virus-treated samples. Most of the magenta points (representing the 144 combinations of samples between treatment groups for a given DEG) do not reflect the expected pattern as clearly compared to what we saw in the litre plots of the Galbraith data. They are not as clustered together (representing replicate inconsistency) and they sometimes overlap the $x=y$ line (representing lack of difference between treatment groups). This finding reflects what we saw in the messy looking parallel coordinate lines of Figure 1.2

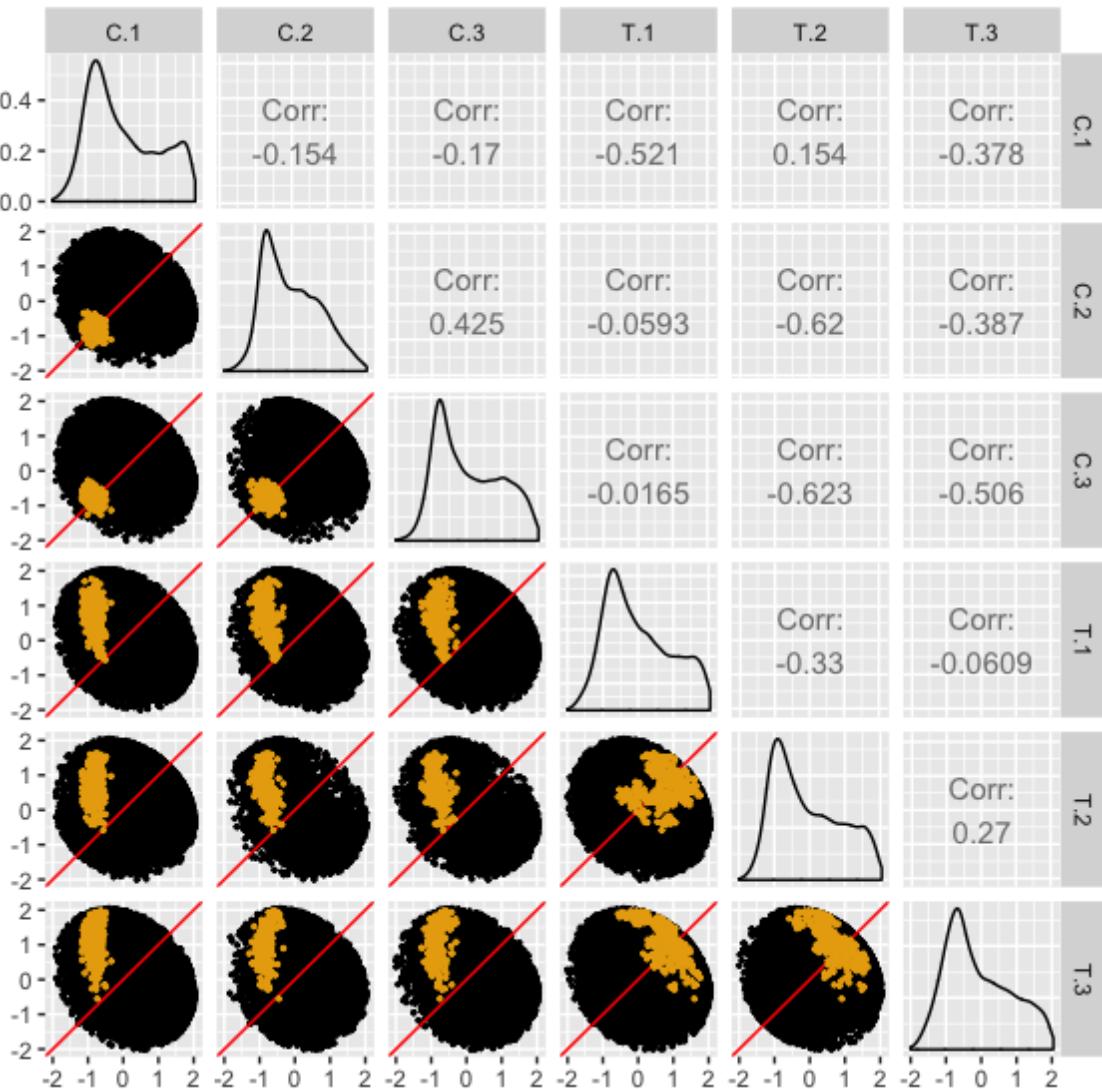


Figure 1.10: The 327 DEGs from the first cluster of the Galbraith dataset (shown in Figure 1.2) superimposed as light orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the $x=y$ line in the treatment scatterplots, but adhering to the $x=y$ line in the replicate scatterplots.

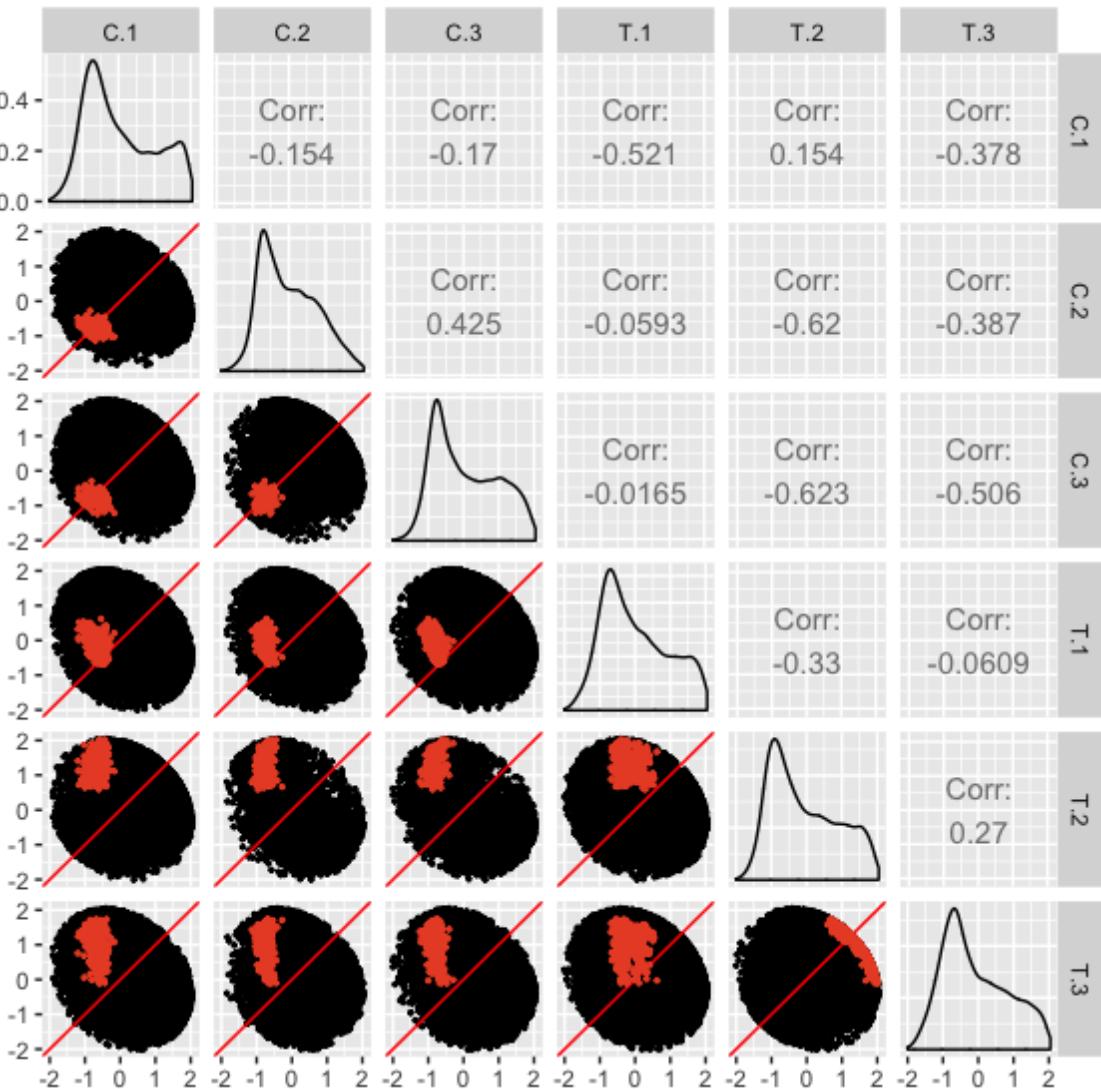


Figure 1.11: The 365 DEGs from the second cluster of the Galbraith dataset (shown in Figure 1.2) superimposed as dark orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the $x=y$ line in the treatment scatterplots, but adhering to the $x=y$ line in the replicate scatterplots. We also see again that the first replicate from the virus-treated sample (“T.1”) may be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots results in the DEGs unexpectedly deviating from the $x=y$ line and its presence in the treatment scatterplots results in the DEGs unexpectedly adhering to the $x=y$ line.

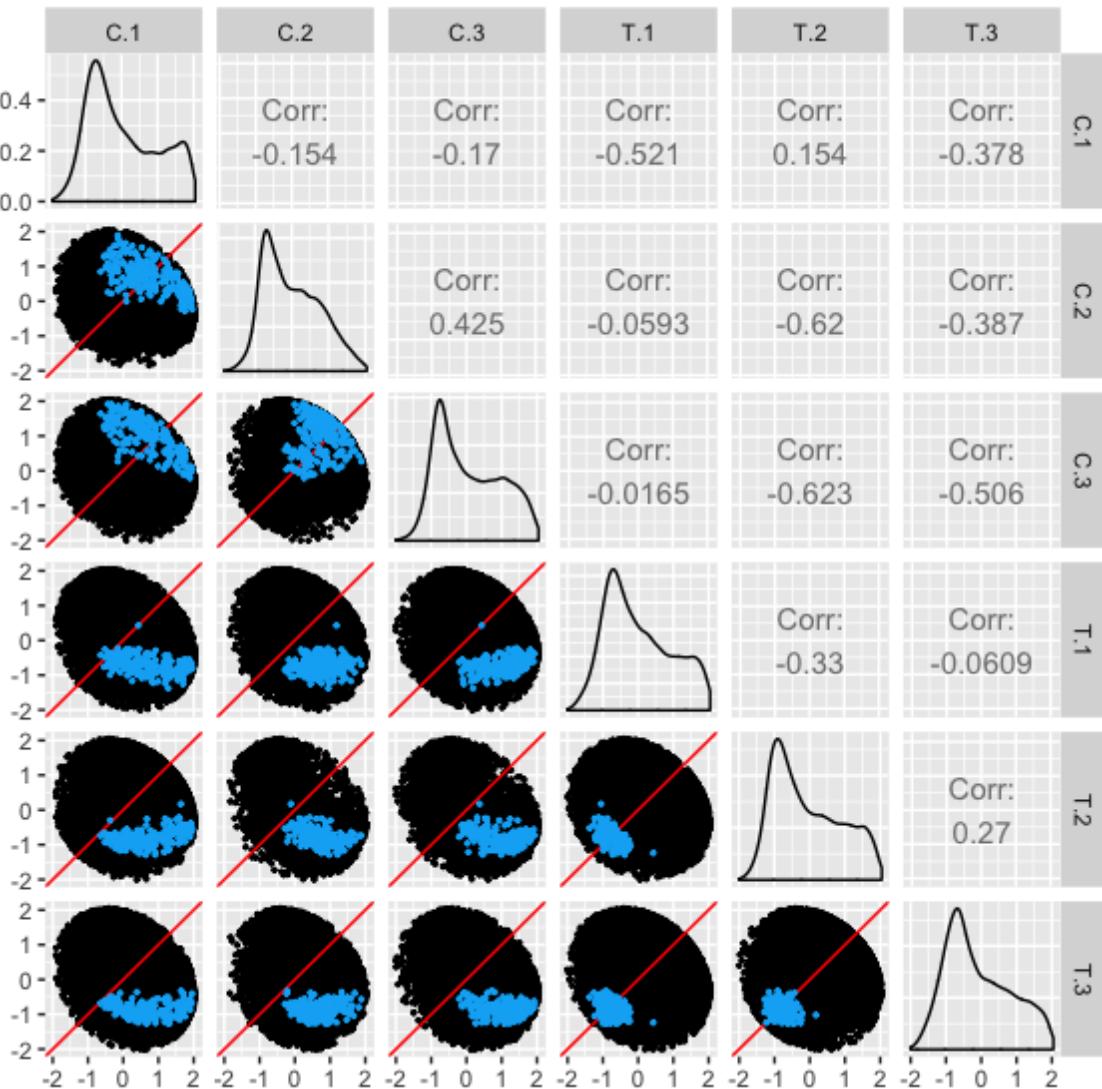


Figure 1.12: The 224 DEGs from the third cluster of the Galbraith dataset (shown in Figure 1.2) superimposed as turquoise dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the $x=y$ line in the treatment scatterplots, but adhering to the $x=y$ line in the replicate scatterplots.

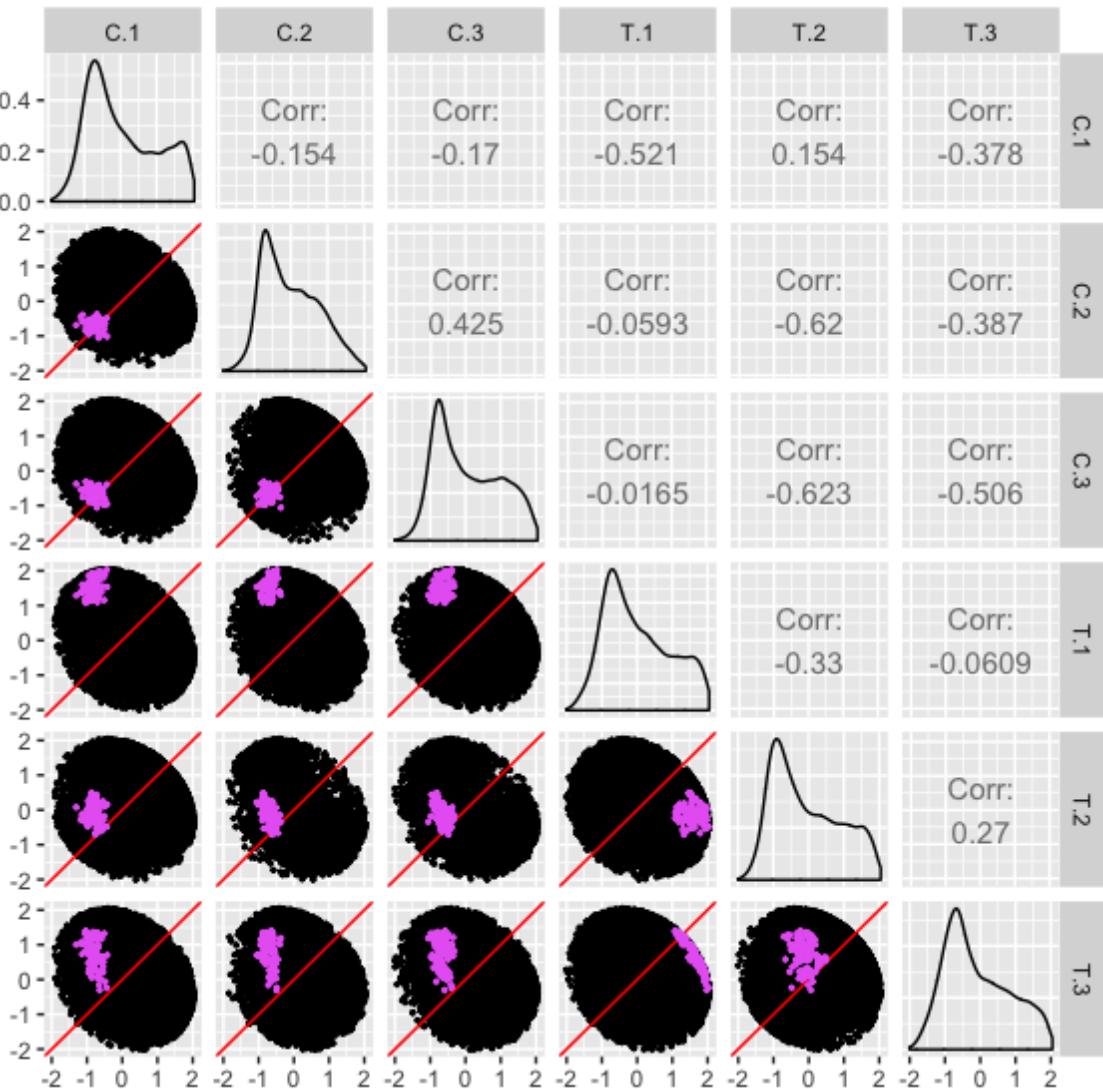


Figure 1.13: The 103 DEGs from the fourth cluster of the Galbraith dataset (shown in Figure 1.2) superimposed as pink dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the $x=y$ line in the treatment scatterplots, but adhering to the $x=y$ line in the replicate scatterplots. We also see that the second replicate from the virus-treated sample (“T.2”) may be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots results in the DEGs unexpectedly deviating from the $x=y$ line and its presence in the treatment scatterplots results in the DEGs unexpectedly adhering to the $x=y$ line.

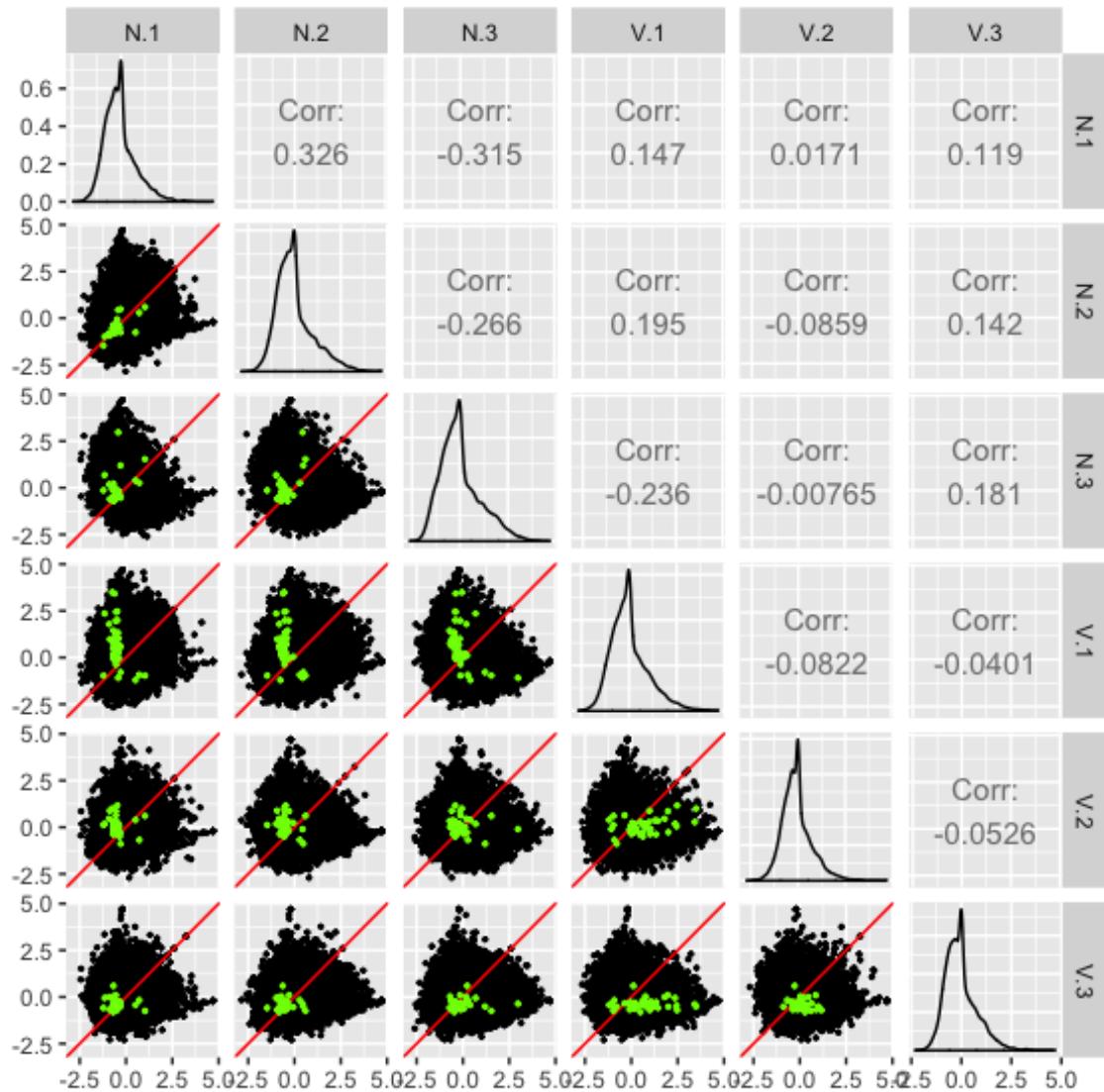


Figure 1.14: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 1, 2, and 3 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from the Galbraith data, the 43 DEGs from this subset of six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the $x=y$ line in the replicate plots and sometimes unexpectedly adhering to the $x=y$ line in the treatment plots.

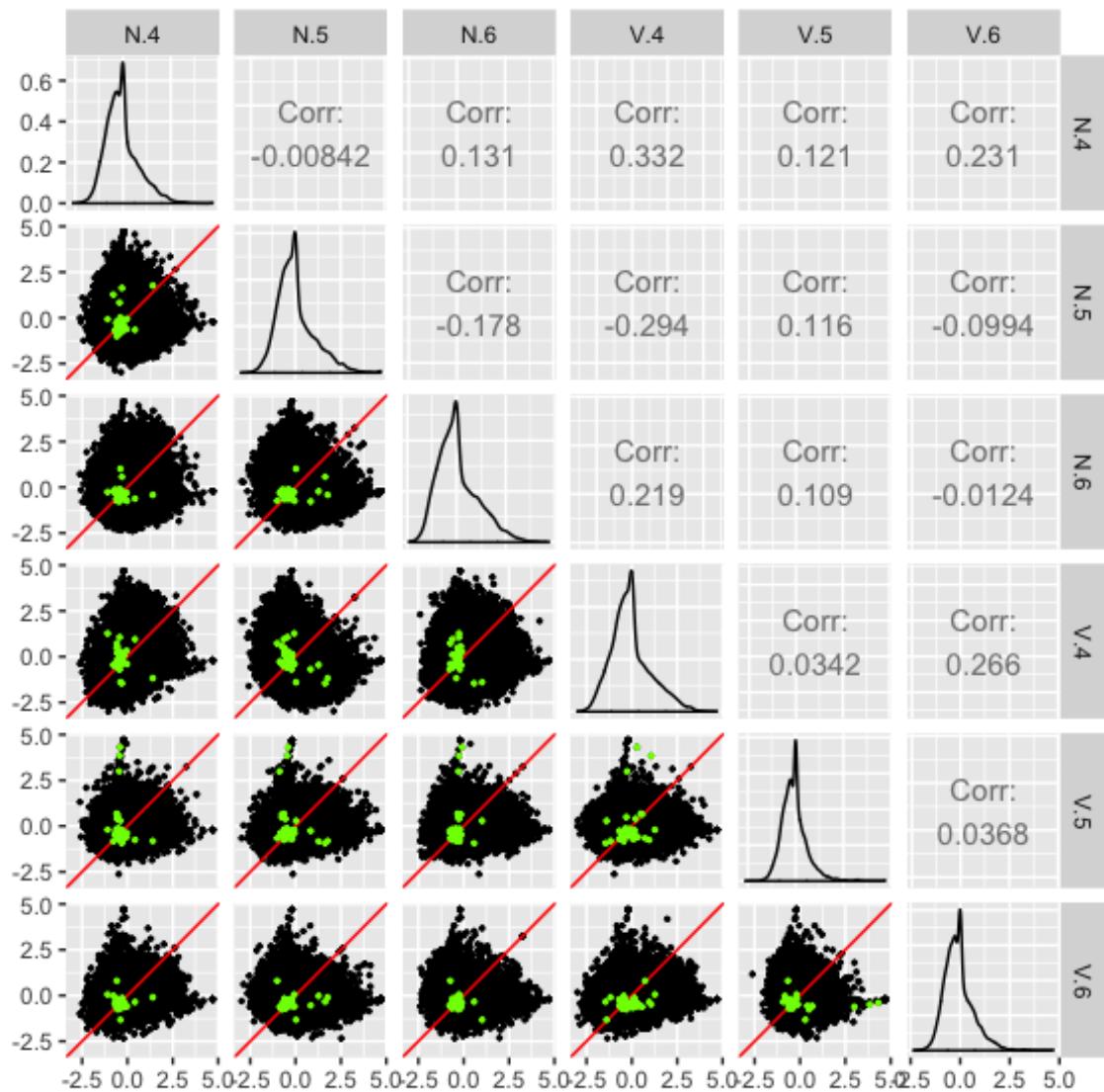


Figure 1.15: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 4, 5, and 6 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from the Galbraith data, the 43 DEGs from this subset of six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the $x=y$ line in the replicate plots and sometimes unexpectedly adhering to the $x=y$ line in the treatment plots.

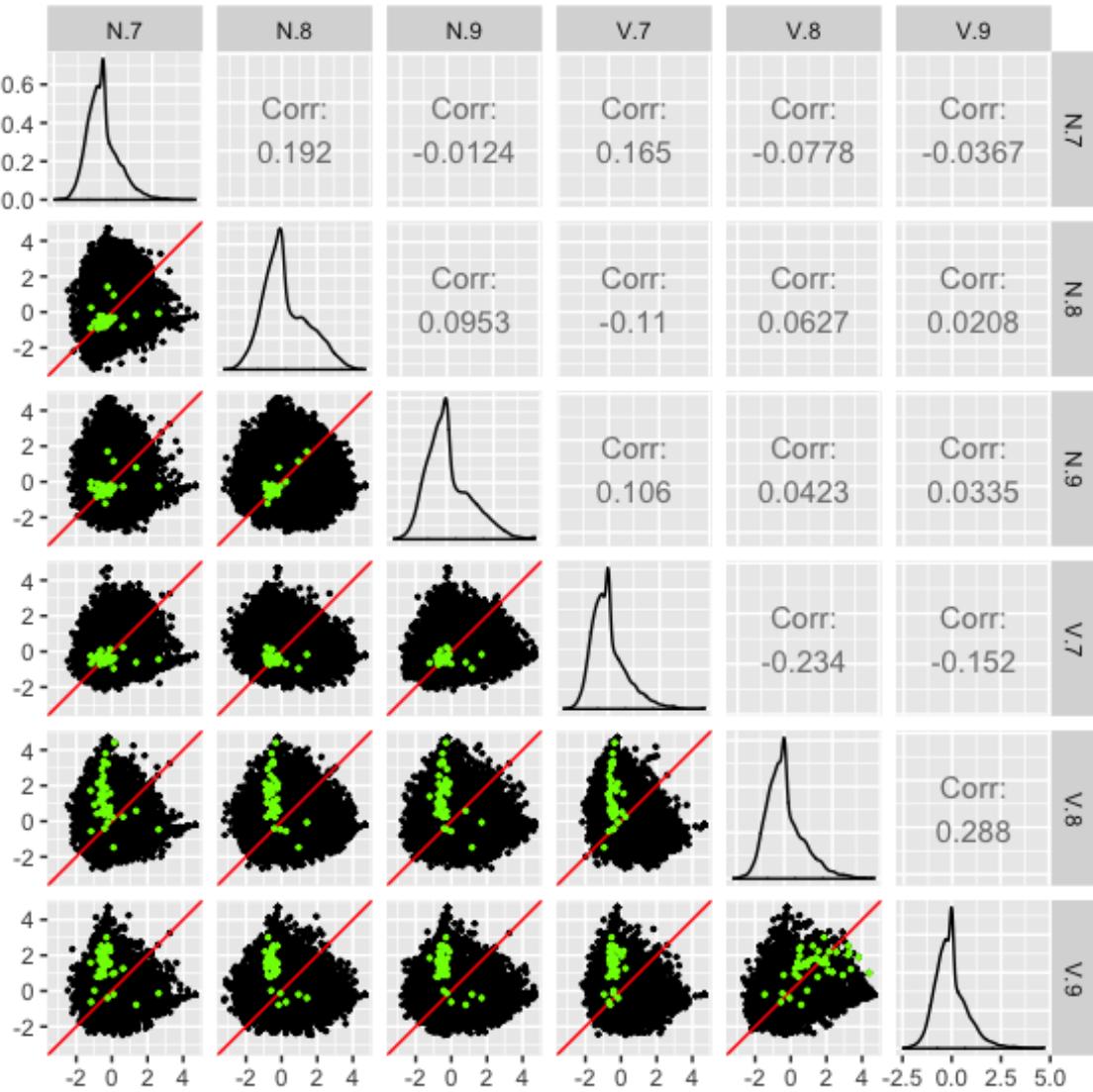


Figure 1.16: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 7, 8, and 9 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from the Galbraith data, the 43 DEGs from this subset of six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the $x=y$ line in the replicate plots and sometimes unexpectedly adhering to the $x=y$ line in the treatment plots.

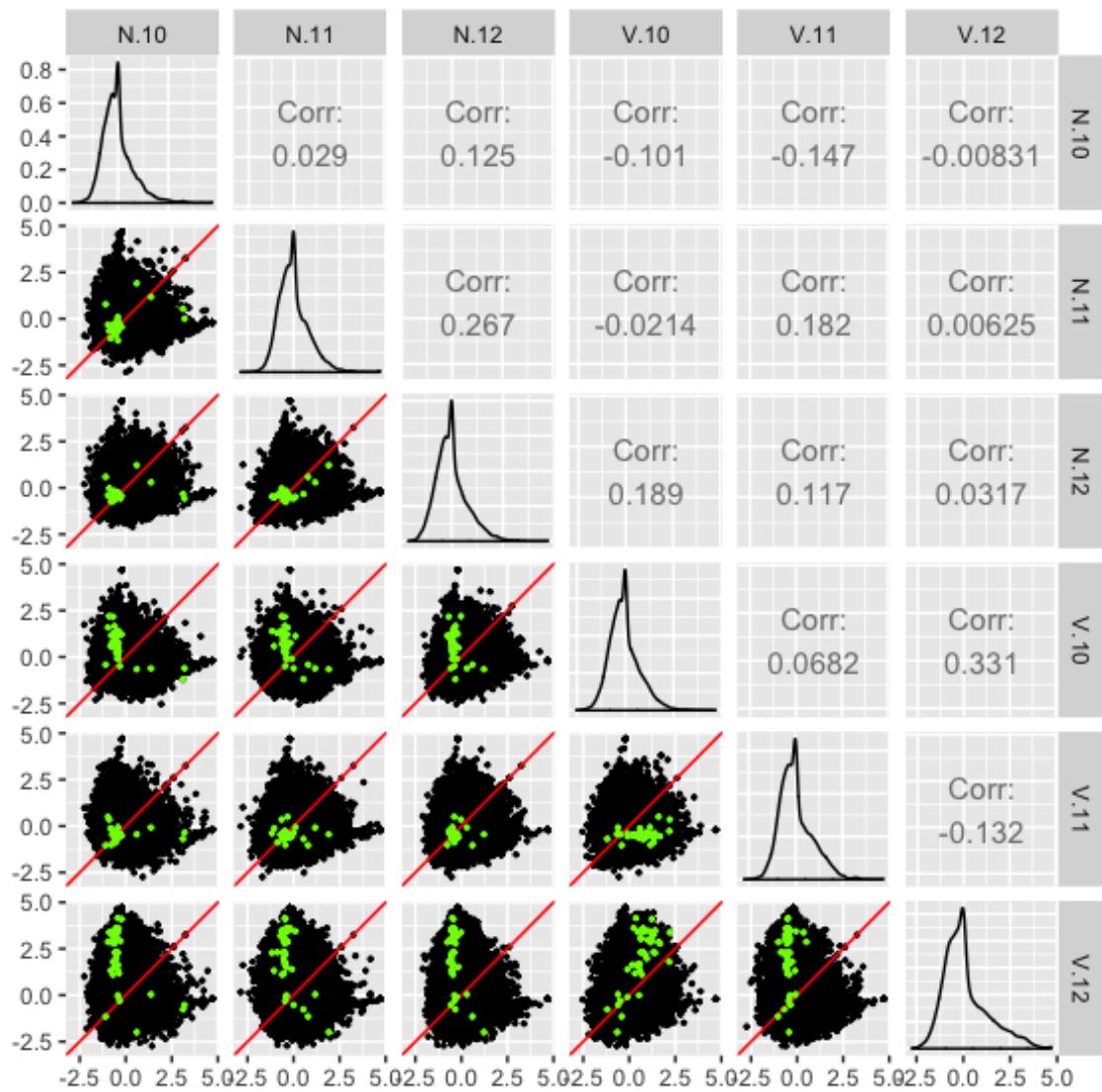


Figure 1.17: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 10, 11, and 12 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from the Galbraith data, the 43 DEGs from this subset of six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the $x=y$ line in the replicate plots and sometimes unexpectedly adhering to the $x=y$ line in the treatment plots.

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	11	2.20E-04	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, C-terminal-binding protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase C, protein kinase shaggy, protein prickle-like, serine/threonine-protein kinase NLK
Circadian rhythm	4	2.40E-02	casein kinase I-like, period circadian protein, protein kinase shaggy, thyrotroph embryonic factor
Hippo signaling pathway	7	5.60E-02	actin, muscle-like, casein kinase I-like, division abnormally delayed protein, hemicentin-2, protein dachshous, serine/threonine-protein kinase Warts
Phototransduction	5	7.30E-02	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, G protein-coupled receptor kinase 1, actin (muscle-like), calcium/calmodulin-dependent protein kinase II, protein kinase C
FoxO signaling pathway	6	1.50E-01	casein kinase I-like, histone acetyltransferase p300-like, insulin-like receptor-like (InR-2), phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, serine/threonine-protein kinase NLK
Notch signaling pathway	4	1.80E-01	C-terminal-binding protein, histone acetyltransferase p300-like, protein jagged-1, protein numb
Insulin resistance	5	2.10E-01	insulin-like receptor-like (InR-2), phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform
mRNA surveillance pathway	6	2.30E-01	cleavage and polyadenylation specificity factor subunit CG7185, heterogeneous nuclear ribonucleoprotein 27C, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform
Jak-STAT signaling pathway	3	2.50E-01	histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform
Phosphatidylinositol signaling system	5	2.70E-01	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, diacylglycerol kinase theta, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase C

Table 1.6: GO analysis results for the 601 DEGs that were upregulated in the NC treatment in the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honeybees are given high quality Chestnut pollen compared to being given low quality Rockrose pollen.

Pathway Term	# of Genes	Benjamini	Example Genes
Sphingolipid metabolism	4	6.00E-01	alkaline ceramidase, putative neutral sphingomyelinase, serine palmitoyltransferase 1, sphingosine-1-phosphate phosphatase 1-like
SNARE interactions in vesicular transport	4	7.00E-01	BET1 homolog, Golgi SNAP receptor complex member 2, syntaxin-7, vesicle transport protein USE1
Basal transcription factors	4	7.30E-01	cyclin-dependent kinase 7, general transcription factor IIF subunit 2, transcription initiation factor IIE subunit beta, transcription initiation factor TFIID subunit 10-like

Table 1.7: GO analysis results for the 340 DEGs that were upregulated in the NR treatment in the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honeybees are given low quality Rockrose pollen compared to being given high quality Chestnut pollen.

Pathway Term	# of Genes	Benjamini	Example Genes
Hippo signaling pathway	5	7.50E-02	actin (muscle-like), cadherin-related tumor suppressor, casein kinase I-like, hemicentin-2, stress-activated protein kinase JNK
Wnt signaling pathway	4	3.00E-01	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, casein kinase I-like, stress-activated protein kinase JNK
Circadian rhythm	2	5.50E-01	casein kinase I-like, thyrotroph embryonic factor

Table 1.8: GO analysis results for the 247 DEGs that were upregulated in the VC treatment in the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honeybees are given high quality Chestnut pollen compared to being given low quality Rockrose pollen.

CHAPTER 1. GENE EXPRESSION RESPONSES TO DIET QUALITY AND VIRAL
32 INFECTION IN *APIS MELLIFERA*

Pathway Term	# of Genes	Benjamini	Example Genes
Fanconi anemia pathway	4	1.60E-02	breast cancer type 2 susceptibility protein homolog, DNA polymerase eta, E3 ubiquitin-protein ligase FANCL, Fanconi anemia group M protein

Table 1.9: GO analysis results for the 129 DEGs that were upregulated in the VR treatment in the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honeybees are given low quality Rockrose pollen compared to being given high quality Chestnut pollen.

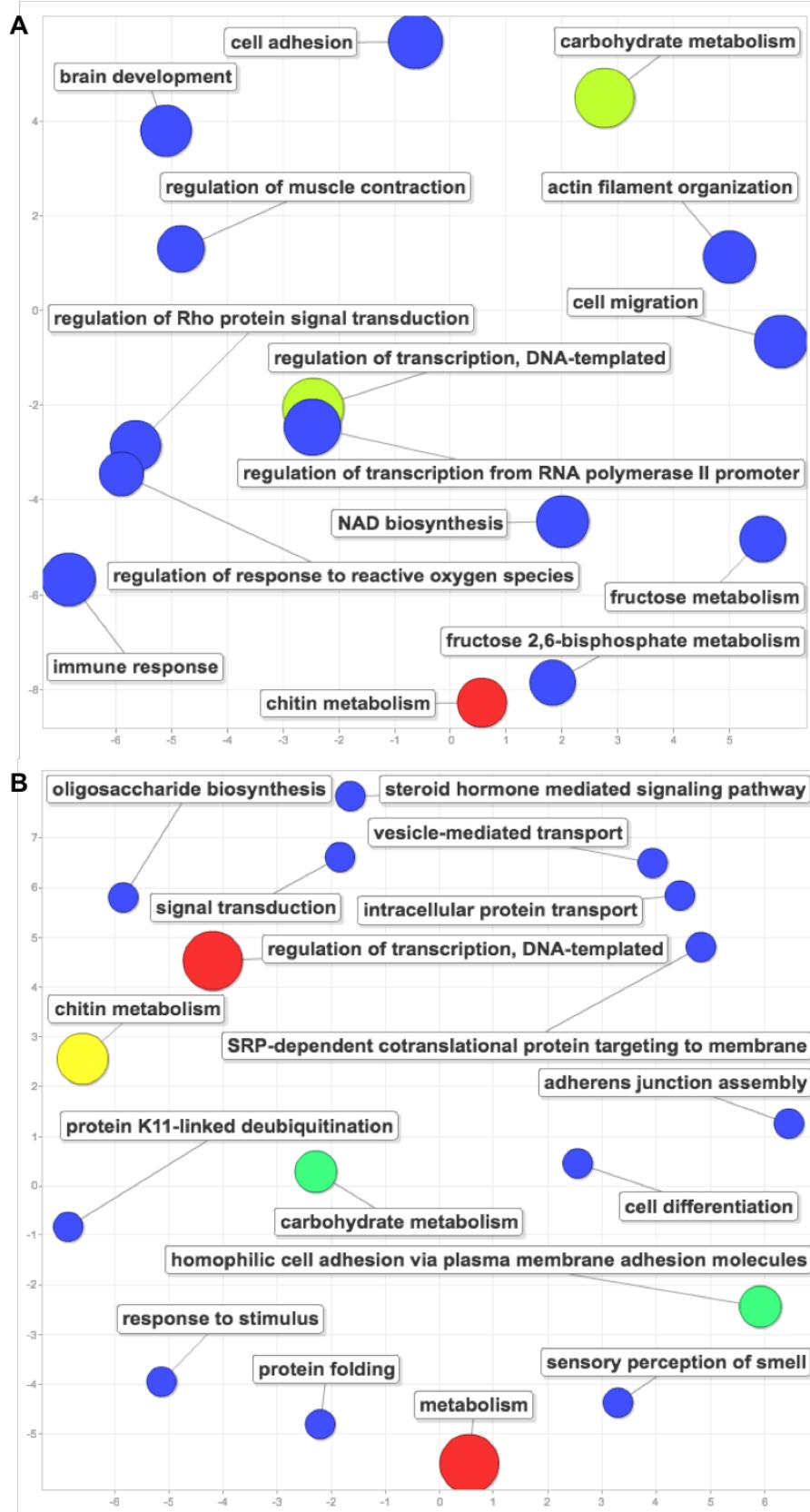


Figure 1.18: GO analysis results for the 122 DEGs related to our “resilience” hypothesis (A) and for the 125 DEGs related to our “resistance” hypothesis (B).

Contrast	DEGs	Interpretation	Results
V vs N	43	Genes that change expression due to virus effect regardless of diet status in bees	Table 1.5
NC vs NR	941	Genes that change expression due to diet effect in uninfected bees	Tables 1.6 and 1.7
VC vs VR	376	Genes that change expression due to diet effect in infected bees	Tables 1.8 and 1.9
VC upregulated in VC vs VR overlapped with NC upregulated in NC vs NR	122	“Resilience” genes that are turned on by good diet regardless of virus infection status in bees	Figure 1.18A
VC upregulated in VC vs VR but NC is not upregulated in NC vs NR	125	“Resistance” genes that are turned on by good diet only in infected bees	Figure 1.18B

Table 1.10: Contrasts in our study for assessing GO and pathways analysis.

339

Bibliography

- 340 C. Alaux, F. Ducloz, and D. Crauser Y. Le Conte. Diet effects on honeybee immunocompe-
341 tence. *Biol. Lett.*, 6:562–565, 2010.
- 342 C. Alaux, C. Dantec, H. Parrinello, and Y. Le Conte. Nutrigenomics in honey bees: digital
343 gene expression analysis of pollen’s nutritive effects on healthy and varroa-parasitized
344 bees. *BMC Genomics*, 12:496, 2011.
- 345 Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and
346 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
347 (*Methodological*), 57:289–300, 1995.
- 348 J. Bond, K. Plattner, and K. Hunt. *Fruit and Tree Nuts Outlook: Economic Insight U.S.*
349 *Pollination- Services Market*. USDA. Economic Research Service Situation and Outlook
350 FTS-357SA, 2014.
- 351 R. Brodschneider and K. Crailsheim. Nutrition and health in honey bees. *Apidologie*, 41:
352 278–294, 2010.
- 353 R. Brodschneider, G. Arnold, N. Hrassnigg, and K. Crailsheim. Does patriline composition
354 change over a honey bee queen’s lifetime? *Insects*, 3:857–869, 2012.
- 355 D. Caron and R. Sagili. Honey bee colony mortality in the Pacific Northwest: Winter
356 2009/2010. *Am Bee J*, 151:73–76, 2011.
- 357 J. Carrillo-Tripp, A.G. Dolezal, M.J. Goblirsch, W.A. Miller, A.L. Toth, and B.C. Bonning.
358 In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep*, 6:22265, 2016.
- 359 Y.P. Chen and R. Siede. Honey bee viruses. *Adv Virus Res*, 70:33–80, 2007.
- 360 Y.P. Chen, J.S. Pettis, M. Corona, W.P. Chen, C.J. Li, M. Spivak, P.K. Visscher,
361 G. DeGrandi-Hoffman, H. Boncristiani, Y. Zhao, D. van Engelsdorp, K. Delaplane,
362 L. Solter, F. Drummond, M. Kramer, W.I. Lipkin, G. Palacios, M.C. Hamilton, B. Smith,

- 363 S.K. Huang, H.Q. Zheng, J.L. Li, X. Zhang, X.F. Zhou, L.Y. Wu, J.Z. Zhou, M-L. Lee,
364 E.W. Teixeira, Z.G. Li, and J.D. Evans. Israeli acute paralysis virus: Epidemiology,
365 pathogenesis and implications for honey bee health. *Plos Pathog*, 10:e1004261, 2014.
- 366 Honey Bee Genome Sequencing Consortium. Finding the missing honey bee genes: lessons
367 learned from a genome upgrade. *BMC Genomics*, 15:86, 2014.
- 368 Y. Le Conte, J-L. Brunet, C. McDonnell, and C. Alaux. *Interactions between risk factors
369 in honey bees*. CRC Press, 2011.
- 370 R.S. Cornman, D.R. Tarpy, Y. Chen, L. Jeffreys, D. Lopez, and J.S. Pettis. Pathogen webs
371 in collapsing honey bee colonies. *Plos One*, 7:e43562, 2012.
- 372 D.L. Cox-Foster, S. Conlan, E.C. Holmes, G. Palacios, J.D. Evans, N.A. Moran, P-L. Quan,
373 T. Brieske, M. Hornig, D.M. Geiser, V. Martinson, D. vanEngelsdorp, A.L. Kalkstein,
374 A. Drysdale, J. Hui, J. Zhai, L. Cui, S.K. Hutchison, J.F. Simons, M. Egholm, J.S. Pettis,
375 and W.I. Lipkin. A metagenomic survey of microbes in honey bee colony collapse disorder.
376 *Science*, 318:283–287, 2007.
- 377 K. Crailsheim. The flow of jelly within a honeybee colony. *J Comp Physiol B*, 162:681–689,
378 1992.
- 379 K. Crailsheim, L.H.W Schneider, N. Hrassnigg, G. Bühlmann, U. Brosch, R. Gmeinbauer,
380 and B. Schöffmann. Pollen consumption and utilization in worker honeybees (*Apis
381 mellifera carnica*): dependence on individual age and function. *J Insect Physiol*, 38:
382 409–419, 1992.
- 383 R.H. Crozier and R.E. Page. On being the right size: Male contributions and multiple
384 mating in social Hymenoptera. *Behav. Ecol. Sociobiol.*, 18:105–115, 1985.
- 385 A. Decourtye, E. Mader, and N. Desneux. Landscape enhancement of floral resources for
386 honey bees in agro-ecosystems. *Apidologie*, 41:264–277, 2010.
- 387 G. DeGrandi-Hoffman and Y. Chen. Nutrition, immunity and viral infections in honey
388 bees. *Current Opinion in Insect Science*, 10:170–176, 2015.
- 389 G. DeGrandi-Hoffman, Y. Chen, E. Huang, and M.H Huang. The effect of diet on protein
390 concentration, hypopharyngeal gland development and virus load in worker honey bees
391 (*Apis mellifera L.*). *J Insect Physiol*, 56:1184–1191, 2010.
- 392 A.G. Dolezal and A.L. Toth. Feedbacks between nutrition and disease in honey bee health.
393 *Current Opinion in Insect Science*, 26:114–119, 2018.
- 394 A.G. Dolezal, J. Carrillo-Tripp, T. Judd, A. Miller, B. Bonning, and A. Toth. Interacting
395 stressors matter: Diet quality and virus infection in honey bee health. *In prep*, 2018.

- 396 C.G. Elsik, A. Tayal, C.M. Diesh, D.R. Unni, M.L. Emery, H.N. Nguyen, and D.E. Hagen.
397 Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine.
398 *Nucleic Acids Research*, 4:D793–800, 2016.
- 399 D. Van Engelsdorp and M.D. Meixner. A historical review of managed honey bee populations
400 in Europe and the United States and the factors that may affect them. *J Invertebr Pathol*,
401 103:S80–S95, 2010.
- 402 D. Van Engelsdorp, J. Jr. Hayes, R.M Underwood, and J. Pettis. A survey of honey bee
403 colony losses in the U.S., fall 2007 to spring 2008. *Plos One*, 3:e4071, 2008.
- 404 D.A. Galbraith, X. Yang, E.L. Niño, S. Yi, and C. Grozinger. Parallel epigenomic and
405 transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *Plos Pathogens*,
406 11:e1004713, 2015.
- 407 N. Gallai, J-M. Salles, J. Settele, and B.B. Vaissière. Economic valuation of the vulnerability
408 of world agriculture confronted with pollinator decline. *Ecol. Econ.*, 68:810–821, 2009.
- 409 D. Goulson, E. Nicholls, C. Botías, and E.L. Rotheray. Bee declines driven by combined
410 stress from parasites, pesticides, and lack of flowers. *Science*, 347:1255957, 2015.
- 411 M.H. Haydak. Honey bee nutrition. *Annu Rev Entomol*, 15:143–156, 1970.
- 412 C. Hou, H. Rivkin, Y. Slabezki, and N. Chejanovsky. Dynamics of the presence of israeli
413 acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses*, 6:
414 2012–2027, 2014.
- 415 D.W. Huang, B.T. Sherman, and R. Lempicki. Systematic and integrative analysis of large
416 gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4:44–57, 2009a.
- 417 D.W. Huang, B.T. Sherman, and R.A. Lempicki. Bioinformatics enrichment tools: paths
418 toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37:
419 1–13, 2009b.
- 420 A-M. Klein, B.E. Vaissière, J.H. Cane, I. Steffan-Dewenter, S.A. Cunningham, C. Kremen,
421 and T. Tscharntke. Importance of pollinators in changing landscapes for world crops.
422 *Proc Biol Sci*, 274:303–313, 2007.
- 423 K. Kulhanek, N. Steinhauer, K. Rennich, D.M. Caron, R.R. Sagili, J.S. Pettis, J.D. Ellis,
424 M.E. Wilson, J.T. Wilkes, D.R. Tarpy, R. Rose, K. Lee, J. Rangel, and D. vanEngelsdorp.
425 A national survey of managed honey bee 2014–2015 annual colony losses in the USA.
426 *Journal of Apicultural Research*, 56:328–340, 2017.
- 427 Johan Larsson. *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*, 2018.
428 URL <https://cran.r-project.org/package=eulerr>. R package version 4.0.0.

- 429 M. Laurent, P. Hendrikx, M. Ribiere-Chabert, and M-P. Chauzat. A pan-European
430 epidemiological study on honeybee colony losses 2012–2014. *Epilobee*, 2013:44, 2016.
- 431 M.I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion
432 for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.
- 433 E. Maori, N. Paldi, S. Shafir, H. Kaled, E. Tsur, E. Glick, and I. Sela. IAPV, a bee-affecting
434 virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion.
435 *Insect Mol Biol*, 18:55–60, 2009.
- 436 H.R. Mattila and T.D. Seeley. Genetic diversity in honey bee colonies enhances productivity
437 and fitness. *Science*, 317:362–364, 2007.
- 438 J.R. De Miranda, G. Cordoni, and G. Budge. The acute bee paralysis virus-Kashmir bee
439 virus-Israeli acute paralysis virus complex. *J Invertebr Pathol*, 103:S30–47, 2010.
- 440 D. Naug. Nutritional stress due to habitat loss may explain recent honeybee colony collapses.
441 *Biol Conserv*, 142:2369–2372, 2009.
- 442 P. Neumann and N.L. Carreck. Honey bee colony losses. *J Apicult Res*, 49:1–6, 2010.
- 443 R.E. Page and H.H. Laidlaw. Full sisters and supersisters: A terminological paradigm.
444 *Anim. Behav.*, 36:944–945, 1988.
- 445 G.D. Pasquale, M. Salignon, Y.L. Conte, L.P. Belzunces, A. Decourtey, A. Kretzschmar,
446 S. Suchail, J-L. Brunet, and C. Alaux. Influence of pollen nutrition on honey bee health:
447 Do pollen quality and diversity matter? *Plos One*, 8:e72016, 2013.
- 448 S.G. Potts, J.C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W.E. Kunin. .
449 *Global pollinator declines: trends, impacts and drivers*, 25:345–353, 2010.
- 450 M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. limma
451 powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic
452 Acids Research*, 43(7):e47, 2015.
- 453 M.D. Robinson, D.J. McCarthy, and G.K. Smyth. edger: a bioconductor package for
454 differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–
455 140, 2010.
- 456 P. Rosenkranz, P. Aumeier, and B. Ziegelmann. Biology and control of Varroa destructor.
457 *J Invertebr Pathol*, 103:S96–S119, 2010.
- 458 T.H. Roulston and S.L. Buchmann. A phylogenetic reconsideration of the pollen starch-
459 pollination correlation. *Evol Ecol Res*, 2:627–643, 2000.
- 460 J.O. Schmidt. Feeding preference of *Apis mellifera* L. (Hymenoptera: Apidae): Individual
461 versus mixed pollen species. *J. Kans. Entomol. Soc.*, 57:323–327, 1984.

- 462 J.O. Schmidt, S.C. Thoenes, and M.D. Levin. Survival of honey bees, *Apis mellifera*
463 (Hymenoptera: Apidae), fed various pollen sources. *J. Econ. Entomol.*, 80:176–183, 1987.
- 464 M.Q. Shen, L.W. Cui, N. Ostiguy, and D. Cox-Foster. Intricate transmission routes and
465 interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with
466 the honeybee host and the parasitic varroa mite. *J Gen Virol*, 86:2281–2289, 2005.
- 467 P.W. Sherman, T.D. Seeley, and H.K. Reeve. Parasites, pathogens, and polyandry in social
468 Hymenoptera. *Am. Nat*, 131:602–610, 1988.
- 469 M. Spivak, E. Mader, M. Vaughan, and N.H. Euliss. The Plight of the Bees. *Environ Sci
470 Technol*, 45:34–38, 2011.
- 471 R.G. Stanley and H.F. Linskens. *Pollen: Biology, biochemistry, management*. Springer
472 Verlag, 1974.
- 473 F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc. REVIGO summarizes and visualizes long
474 lists of Gene Ontology terms. *Plos ONE*, 6:e21800, 2011.
- 475 D.R. Tarpy. Genetic diversity within honeybee colonies prevents severe infections and
476 promotes colony growth. *Proc. R. Soc. Lond. B*, 270:99–103, 2003.
- 477 D. van Engelsdorp, J.D. Evans, C. Saegerman, C. Mullin, E. Haubrige, B.K. Nguyen,
478 M. Frazier, J. Frazier, D. Cox-Foster, Y. Chen, R. Underwood, D.R. Tarpy, and J.S.
479 Pettis. Colony collapse disorder: A descriptive study. *PLoS One*, 4:e6481, 2009.
- 480 K.P. Weinberg and G. Madel. The influence of the mite *Varroa Jacobsoni Oud.* on the
481 protein concentration and the haemolymph volume of the brood of worker bees and
482 drones of the honey bee *Apis Mellifera L.* *Apidologie*, 16:421–436, 1985.
- 483 T.D. Wu, J. Reeder, M. Lawrence, G. Becker, and M.J. Brauer. GMAP and GSNAp
484 for genomic sequence alignment: Enhancements to speed, accuracy, and functionality.
485 *Methods Mol Biol*, 1418:283–334, 2016.
- 486 X. Yang and D. Cox-Foster. Effects of parasitization by *Varroa destructor* on survivorship
487 and physiological traits of *Apis mellifera* in correlation with viral incidence and microbial
488 challenge. *Parasitology*, 134:405–412, 2007.
- 489 X.L. Yang and D.L. Cox-Foster. Impact of an ectoparasite on the immunity and pathology
490 of an invertebrate: Evidence for host immunosuppression and viral amplification. *P Natl
491 Acad Sci USA*, 102:7470–7475, 2005.