**TRANSCRIPTIONAL REGULATION OF INTRAGENOMIC CONFLICT IN BEES**

**David A. Galbraith, Rong Ma, and Christina M Grozinger**

Summary

1. **Does PSGE in the brain also support kin selection theory of intragenomic conflict?**
   1. Paternally inherited alleles show higher expression in reproductive individuals than sterile individuals
2. **Do the same genes show PSGE in the brain and reproductive tissues?**
   1. **Didn’t find this in the results.**
3. **What is mechanistic basis of the tissue differences?** 
   1. Reproductive genes are just not expressed in brain
      1. There was an overlap of 2,413 genes that could be found in both tissues, 1,532 unique to the brain, and 2,551 unique to the fat bodies and ovaries
      2. There were 12 genes that were differentially expressed between reproductive and sterile individuals in both brains and FBO.
         1. This number seems low at first, but it should be noted that this represents 50% of the DEGs in the brain.
   2. Reproductive genes are expressed but do not show POE in brain
      1. epigenetic marks are lost in brain tissue
      2. tissue-specific epigenetic marks
4. **Are PSGE’s the same as the DEGs that differ between sterile and reproductive bees?** 
   1. None of the genes that were significantly differentially expressed exhibit parent specific expression as well.
5. **Are the PSGEs linked with the DEGs through networks?** 
   1. Determine if PSGEs and DEGs are in the same modules in the brain. If so, see if PSGE are hubs and DEGs are spokes
      1. There is one module that is significantly upregulated for both, so potentially yes
      2. however the general trend is that they are in different modules in the brain
   2. Determine if PSGEs and DEGs are in the same modules in the ovaries
      1. Yes, this seems to be the case. There are many modules that show significant enrichment of both DEGs and Ovaries

**RESULTS**

*Differential Gene Expression.*  Differential gene expression analysis revealed 24 genes whose expression were significantly different between brain samples taken from reproductive and sterile individuals (GLM, FDR < 0.05). Thirteen of the genes were upregulated in reproductive individuals (and downregulated in sterile individuals), including genes encoding an ADAMTS-like protein, a nucleoporin, a protein lin-9 homolog, and several uncharacterized proteins. Eleven of the genes were upregulated in sterile individuals (and downregulated in reproductive individuals), including a cytochrome P450 and a protein kinase. A full list of genes and differential gene expression results can be found in Supplementary Tables (REF).

*Allele Specific Expression Analysis*. There were 120 genes with transcripts that showed allele-specific expression (Table 1). Paternally inherited genes showed significantly higher expression in reproductive individuals compared to sterile individuals (Fisher’s exact test, P=0.0001). There was no such effect in either maternally inherited genes or genes associated with Africanized or European genetic backgrounds. There were no overlaps between genes with transcripts that showed allele-specific expression and genes which were differentially expressed in reproductive and sterile individuals.

\*\*Add the comparisons of the brain vs ovaries here? Or do this as an analysis after the WGCA to ask about tissue specific effects?\*\*

*Weighted gene co-expression network analysis (WGCNA)*. Because there were no overlaps between differentially expressed genes and genes with allele specific expression, WGCNA was used to construct networks of genes based on their expression patterns (termed ‘modules’). Because these modules are constructed independently from sample trait information, the modules can be used to assess how groups of genes with correlated expression respond to parental cross, experimental block, and reproductive status treatments.

WGCNA also identified 22 genetic modules in brain tissue. Of these modules, 8 were correlated with sample traits. Most of these (7/8) were correlated with experimental block, while only one was correlated with parental cross.

In addition, the modules in the co-expression network can be assessed for enrichment of genes that show significant differential expression (DEG) and significant parent-specific gene expression (PSGE). In brain tissue, modules identified by WGCNA were compared to DEGs and PSGEs identified in brain tissue in the present study. Eleven modules showed significant enrichment in DEGs or PSGEs, two of which were significantly enriched for both PSGEs and DEGs. Module 8 was significantly enriched for both DEGs and PSGEs from sterile workers while module 1 was enriched for both DEGs and PSGEs from reproductive workers. In addition to these two modules, there were 3 modules that were significantly enriched for sterile PSGEs and 4 modules enriched for active PSGEs.

WGCNA identified 22 genetic modules in reproductive tissue, which included fat bodies and ovaries. In reproductive tissue, 12 modules were correlated with at least one sample trait. Of these modules, most were correlated with reproductive status of the individual (68%), half were correlated with experimental block, and only two were correlated with parental cross background.

In the comparison of FBO-specific gene network to DEGs and PSGEs identified in (Galbraith et al 2016), 11 modules showed significant overlaps with at least one list of significant genes, and 5 modules showed enrichment for both DEGs and PSGEs. For example, module 18 was significantly enriched for PSGEs associated with both active and sterile individuals, DEGs. Except for Module 18, most modules in FBO tissue were enriched with sterile PSGEs or active PSGEs, but not both.

*Methods*

*Biological Samples:* The samples used here were generated in a previous study (Galbraith et al. 2016). Three Africanized *Apis mellifera* swarms were obtained from the Mojave Desert in southern California, and their genotypes were confirmed. These were used to create reciprocal crosses with three European (*Apis mellifera ligustica*) colonies that were managed by Glenn Apiaries. These colonies were separated into three blocks (labeled A, B, and C), with one Africanized and one European colony per block. Three queens from each colony per block were crossed with three drones from the reciprocal source colony by instrumental insemination, resulting in a total of 18 colonies (9 headed by an Africanized queen and 9 headed by a European queen). Colonies were generated and maintained at Glenn Apiaries. For this study, we only used samples from blocks A and B.

One queenless and broodless experimental colony was created for each source colony block using 500 callow workers from each of the 6 reciprocal crosses within the block, for a total of 3 experimental colonies. The callow workers were paint marked according to their colony of origin to be able to identify the different colonies. Each experimental colony was provisioned with two frames of honey and one frame of empty drone come to encourage egg laying. The colonies were left to mature for at least 15 days to allow for worker ovary activation and subsequent egg laying. At the end of the experiment, all individuals were collected onto dry ice and shipped to Penn State University for further processing.

*Brain Dissection and RNA extraction*: In our previous study (Galbraith et al. 2016), we analyzed the fat body and ovary tissue from selected individuals. The head and thoraces of these individuals had been stored at -80ºC. Here, the heads from these individuals were used for further analysis of the brain transcriptome of reproductive and sterile workers. In some cases, the heads of the individuals were missing (identified in supplementary table X), so these individuals were replaced with individuals that had been previously identified as reproductive and sterile, based on ovary score, from the same crosses as the missing samples.

The heads of these individuals were lyophilized for 45 minutes and the brains were then dissected on dry ice, and the individual brains were homogenized using a Fastprep tissue homogenizer. RNA was extracted using a Qiagen AllPrep DNA/RNA mini kit following the manufacturer’s standard protocol. DNA from the brain tissue was also extracted and saved for a potential future study. The RNA was then shipped to Duke University for library preparation and transcriptome sequencing on an Illumina HiSeq 4000 with 151bp paired end reads across. All samples were run on each of the two lanes to minimize the impact of potential lane effects. Transcriptomic data generated from this study has been deposited in NCBI’s Gene Expression Omnibus and can be accessed through the accession number XXXXXXX.

*RNAseq Quality Control and Mapping:* To assess the quality of the RNA sequencing data, the reads from the brain transcriptomes were pass through FastQC (**REF**) to identify issues associated with read quality, overrepresented sequences, and sequence duplication rate. The reads were then preprocessed to remove adaptor sequences and low quality reads from the data using Trimmomatic (REF).

In our previous study, pseudogenomes were created by identifying the genomic variants associated with each parent genome (maternal and paternal) from the four experimental colonies and inserting those variants into the most recent honey bee reference genome build (Amel\_4.5) (**REF**), see Galbraith et al. 2016 for detailed methods on the generation of these pseudogenomes. To identify genes exhibiting allele specific expression, the preprocessed reads were aligned to their individual parent pseudogenomes using the same pipeline as in our previous manuscript (Galbraith et al. 2016).

To detect changes in brain gene expression between the reproductive and sterile individuals, the preprocessed transcriptomic reads were aligned to the most recent honey bee reference genome build (Amel\_4.5) (**REF**) using the same pipeline as in our previous manuscript.

*Detection of Allele Specific Expression:* To detect allele specific expression in reproductive and sterile workers from the different genotypes, we used the same pipeline that we previously used to detect allele specific expression in ovaries and fat bodies of the same reproductive and sterile workers. This conservative pipeline used the intersection of two methods to detect allele specific expression, including a Storer-Kim test (REF) and a generalized linear interactive mixed model (GLIMMIX). For the GLIMMIX, parent of origin (maternal vs. paternal), race of origin (Africanized honey bee vs. European honey bee), and their interaction (parent × race) were included as fixed effects in the model, while SNP and replicate were considered as random effects. A false-discovery threshold of *FDR* < 0.05 was used to correct for multiple testing for both tests. Only transcripts that exhibited a significant allelic expression bias in both methods were used for further analyses.

*Detection of Transcriptomic Changes in Reproductive Individuals:* Read counts from the preprocessed transcriptomic reads were imported into R statistical software (www.r-project.org) for differential expression analysis. Genes filtered to remove those genes with fewer than five reads per sample and normalized using a trimmed mean of M-values (TMM) method (**REF**). Differential brain gene expression was assessed by LIMMA (**REF**) using a linear model with parent of origin, race of origin, and their interaction used as fixed effects. To determine the significance of the differential gene expression, a false-discovery threshold of FDR < 0.05 was used to correct for multiple testing.

*Weighted gene co-expression network analysis (WGCNA).* To understand whether groups of genes were consistently expressed together to regulate intragenomic conflict in brain and reproductive tissues, we constructed tissue-specific gene networks. WGCNA is based on an unsupervised hierarchical clustering algorithm that constructs networks of genes based solely on the similarity of their expression patterns and organizes them into groups of co-expressed genes, called modules. Assignment of genes to modules is independent of sample information (e.g. sterile vs reproductive, European vs Africanized).

To generate tissue-specific gene networks, variance stabilized gene expression data from brains (24 samples) and ovaries (25 samples) were grouped into modules based solely on similarity of expression patterns in each data set separately. Genes with counts of at least 40 in 12 or more samples were kept in the analysis. Because genes within each module showed very highly correlated patterns, the first principal component of the genes within a module was used to represent the entire module (module ‘eigengene’). Then, these module representatives were correlated with sample traits using a generalized linear model, with maternal vs paternal cross (European vs Africanized), genetic block, and reproductive status of the individual as fixed effects.

Each of the gene networks were constructed using the same settings to ensure technical consistency. Minimum module size was set to 30, and deep split was set to 2. The ‘bicor’ correlation method was used, as recommended in the WGCNA package documentation. Modules were built with a standardized connectivity score of -2.5, and module definition was based on “hybrid” branch cutting at a cut height of 0.25. A signed gene co-expression network was constructed with a soft threshold of 10. Modules were merged based on a cut height of 0.1. Module eigengenes were correlated with sample traits using a generalized linear model with forager-type, pheromone exposure treatment, and their interaction as fixed effects.

*Overlaps between gene modules and genes of interest*. Using hypergeometric tests, we evaluated the degree of overlap between brain and FBO-specific gene networks to sets of genes showing differentially regulated or parent specific expression in brain or FBO samples.

**Table 1. Transcripts with allele-specific expression in the brains of sterile and reproductive workers.** Paternally inherited genes showed significantly higher expression in reproductive individuals compared to sterile individuals (Fisher’s exact test, P=0.0001). There was no such effect in either maternally inherited genes or genes associated with Africanized or European genetic backgrounds.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | paternal | maternal | European | Africanized |
| Reproductive | 62 | 13 | 5 | 1 |
| Sterile | **26** | 7 | 5 | 1 |

Table 2. Brain modules identified in weighted gene correlation network analysis are significantly enriched for PSGE and DEG and correlated; however, brain modules were not correlated with ovary activation.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Brain Module Name | Module Size | Active PSGE Module Overlaps | Sterile PSGE Module Overlap | DEG Brain Module Overlap | Active PSGE P-value | Sterile PSGE P-value | DEG Brain P-value | Ovary Activation P-value | Ovary Activation Correlation |
| ME5 | 877 | 5 | 8 | 0 | 0.278441636 | 0.038853194 | 0.757449519 | 0.786677527 | -0.024263209 |
| ME9 | 291 | 1 | 4 | 0 | 0.431692848 | 0.018095613 | 0.369234965 | 0.491588276 | 0.062642007 |
| ME22 | 38 | 1 | 0 | 0 | 0.015608243 | 0.180684478 | 0.05793042 | 0.209570745 | 0.111695185 |
| ME4 | 1249 | 13 | 8 | 2 | 0.003463152 | 0.203226152 | 0.311142463 | 0.800683033 | 0.019325607 |
| ME8 | 346 | 3 | 6 | 2 | 0.096507692 | 0.002188521 | 0.016284036 | 0.602501109 | -0.032893459 |
| ME1 | 2460 | 21 | 23 | 5 | 0.003934687 | 0.001290757 | 0.176505127 | 0.495391391 | -0.05283921 |
| ME7 | 552 | 7 | 3 | 1 | 0.006429334 | 0.32616119 | 0.213785237 | 0.565764264 | -0.042694876 |
| ME17 | 111 | 0 | 0 | 0 | 0.42969608 | 0.442070934 | 0.160320632 | 0.554285779 | -0.053971305 |
| ME18 | 92 | 0 | 1 | 0 | 0.371970131 | 0.083512451 | 0.134748579 | 0.055080424 | -0.164468531 |
| ME11 | 246 | 6 | 0 | 0 | 0.000234273 | 0.727195154 | 0.322252065 | 0.648749773 | 0.039770701 |
| ME19 | 71 | 0 | 0 | 0 | 0.301444198 | 0.311165238 | 0.105616674 | 0.802128711 | -0.016039419 |
| ME10 | 251 | 0 | 3 | 0 | 0.72078791 | 0.042375013 | 0.327633481 | 0.853431675 | -0.014453872 |
| ME21 | 56 | 0 | 0 | 0 | 0.246337521 | 0.254617827 | 0.084235153 | 0.456992368 | -0.063871618 |
| ME12 | 197 | 1 | 0 | 1 | 0.260732479 | 0.646038539 | 0.037753618 | 0.600594441 | -0.048131344 |
| ME14 | 147 | 0 | 0 | 1 | 0.525078127 | 0.538693236 | 0.022001551 | 0.61404942 | 0.040126877 |
| ME13 | 157 | 0 | 0 | 0 | 0.548653934 | 0.562464291 | 0.219267303 | 0.609507655 | 0.041584822 |
| ME2 | 1850 | 7 | 7 | 2 | 0.727184041 | 0.765764445 | 0.567522359 | 0.519498874 | 0.058660593 |
| ME3 | 1508 | 4 | 6 | 0 | 0.887166487 | 0.684457306 | 0.917096838 | 0.774245444 | 0.021838982 |
| ME15 | 129 | 1 | 0 | 0 | 0.137354082 | 0.492648504 | 0.183877538 | 0.934751468 | 0.007360854 |
| ME20 | 64 | 0 | 0 | 0 | 0.276241207 | 0.285324158 | 0.095698858 | 0.917197639 | -0.009364689 |
| ME6 | 617 | 1 | 1 | 1 | 0.822358951 | 0.838378751 | 0.251784038 | 0.643023248 | 0.038948854 |
| ME16 | 117 | 0 | 0 | 0 | 0.446812822 | 0.459460106 | 0.168244414 | 0.822622646 | 0.018823875 |

Table 3. Ovary modules identified in weighted gene correlation network analysis are significantly enriched for PSGE and DEG and correlated with ovary activation.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ovary Module Name | Module Size | Active PSGE Module Overlaps | Sterile PSGE Module Overlap | DEG Brain Module Overlap | Active PSGE P-value | Sterile PSGE P-value | DEG Brain P-value | Ovary Activation P-value | Ovary Activation Correlation |
| ME11 | 441 | 10 | 10 | 0 | 0.278441636 | 0.038853194 | 3.37E-08 | 0.165965164 | 0.119134918 |
| ME14 | 219 | 9 | 7 | 0 | 0.431692848 | 0.018095613 | 0.003222938 | 0.391286652 | -0.074528175 |
| ME2 | 2105 | 31 | 15 | 0 | 0.015608243 | 0.180684478 | 0.263674291 | 0.000112301 | 0.248071546 |
| ME4 | 1361 | 14 | 12 | 0 | 0.003463152 | 0.203226152 | 0.000146519 | 0.003247077 | 0.224923796 |
| ME13 | 390 | 2 | 2 | 0 | 0.096507692 | 0.002188521 | 0.742648465 | 0.989416855 | -0.001152406 |
| ME18 | 107 | 0 | 0 | 0 | 0.003934687 | 0.001290757 | 3.42E-14 | 0.002222495 | 0.223055039 |
| ME17 | 155 | 2 | 2 | 0 | 0.006429334 | 0.32616119 | 0.000391754 | 0.137731919 | -0.125825702 |
| ME21 | 83 | 3 | 1 | 0 | 0.42969608 | 0.442070934 | 0.998388813 | 0.4961957 | 0.038398146 |
| ME1 | 2235 | 22 | 23 | 0 | 0.371970131 | 0.083512451 | 0.118751601 | 0.055989685 | -0.161256646 |
| ME8 | 892 | 15 | 7 | 0 | 0.000234273 | 0.727195154 | 0.751152984 | 0.712418563 | -0.030481207 |
| ME5 | 1350 | 7 | 6 | 0 | 0.301444198 | 0.311165238 | 0.965736199 | 0.882006641 | 0.012712153 |
| ME9 | 635 | 1 | 1 | 0 | 0.72078791 | 0.042375013 | 0.999976863 | 0.001264697 | 0.246410528 |
| ME12 | 435 | 1 | 0 | 0 | 0.246337521 | 0.254617827 | 0.470942299 | 0.076911576 | 0.13971482 |
| ME10 | 545 | 3 | 4 | 0 | 0.260732479 | 0.646038539 | 0.052581167 | 0.288755476 | 0.086466371 |
| ME15 | 168 | 0 | 1 | 0 | 0.525078127 | 0.538693236 | 1.98E-08 | 0.634511394 | -0.041938261 |
| ME3 | 1639 | 16 | 22 | 0 | 0.548653934 | 0.562464291 | 0.307381071 | 0.052077547 | -0.13191166 |
| ME19 | 92 | 0 | 0 | 0 | 0.727184041 | 0.765764445 | 0.793265487 | 0.000338508 | -0.26673911 |
| ME22 | 36 | 0 | 0 | 0 | 0.887166487 | 0.684457306 | 0.972957698 | 0.126303993 | -0.084222671 |
| ME16 | 159 | 6 | 2 | 0 | 0.137354082 | 0.492648504 | 0.932823359 | 0.438283539 | -0.059235918 |
| ME7 | 954 | 30 | 22 | 0 | 0.276241207 | 0.285324158 | 0.020605069 | 5.51E-05 | -0.287554289 |
| ME6 | 1242 | 25 | 26 | 0 | 0.822358951 | 0.838378751 | 0.998347851 | 0.000505686 | -0.216883032 |
| ME20 | 84 | 4 | 1 | 0 | 0.446812822 | 0.459460106 | 0.843087661 | 0.009067785 | -0.190238646 |
| ME0 | 10 | 0 | 0 | 0 | 0.999933224 | 0.996644501 | 1 | 0.71877461 | -0.029813741 |

Figure 1. SK Plots for reproductive (left) and sterile (right) invididuals.

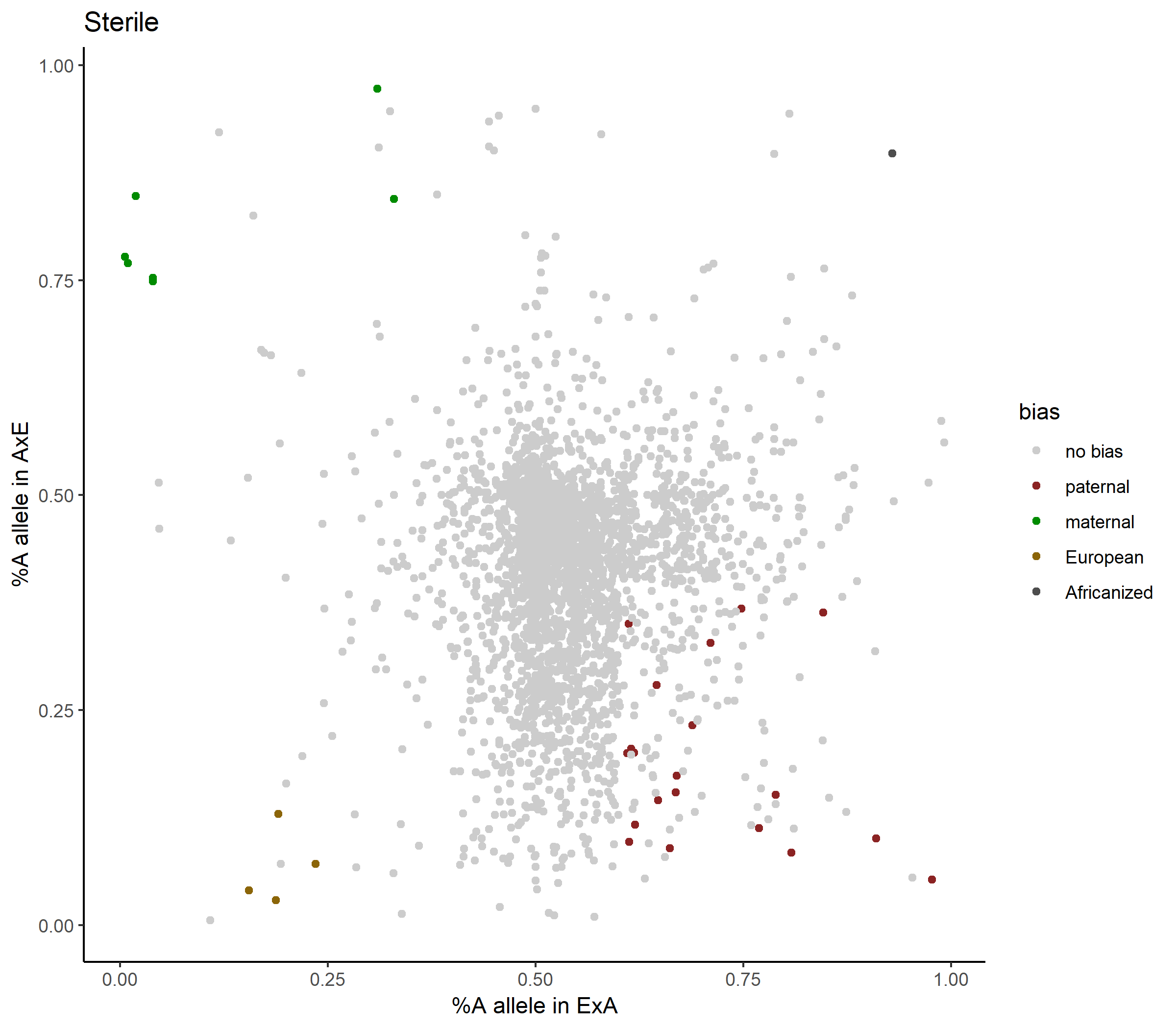
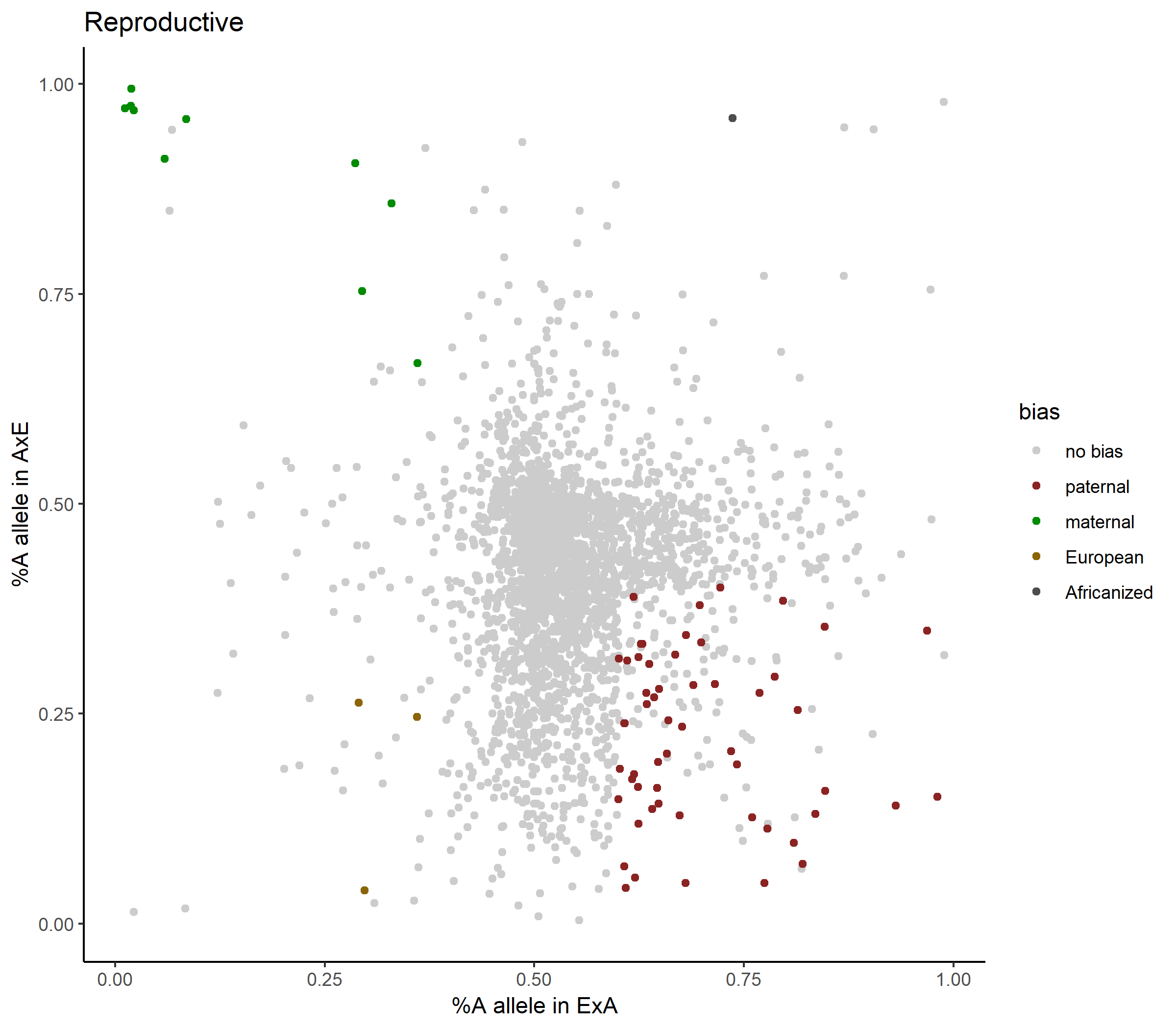
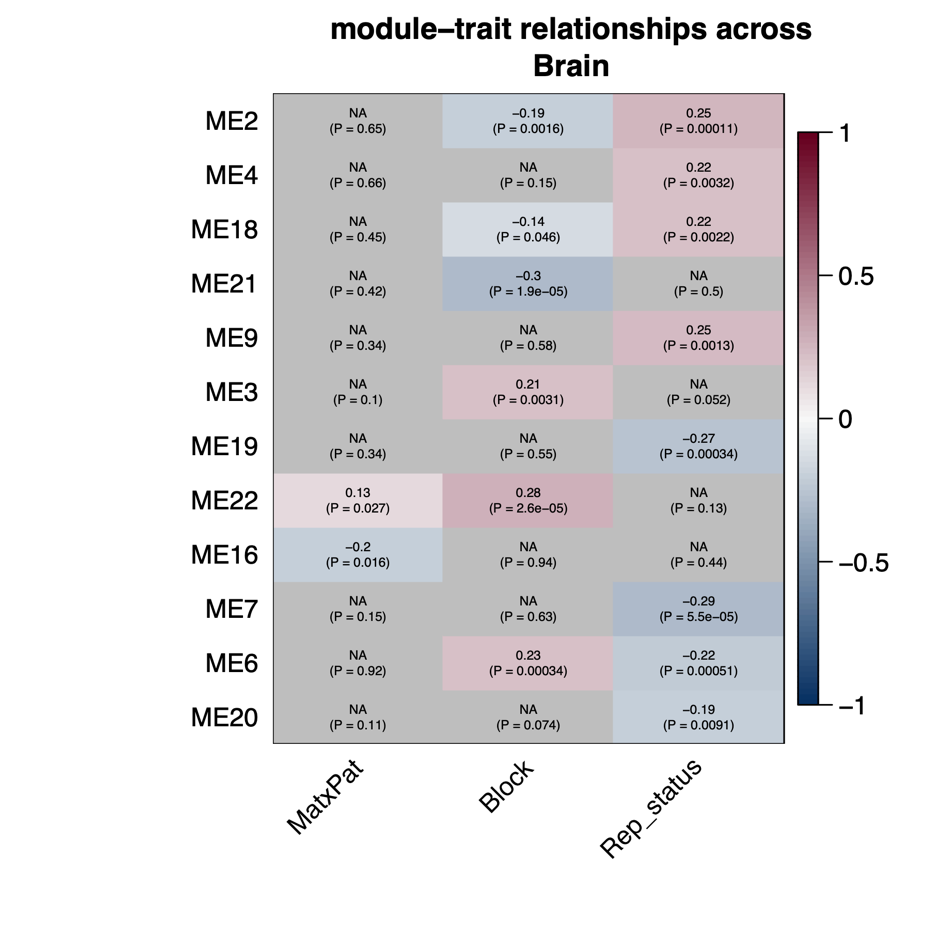
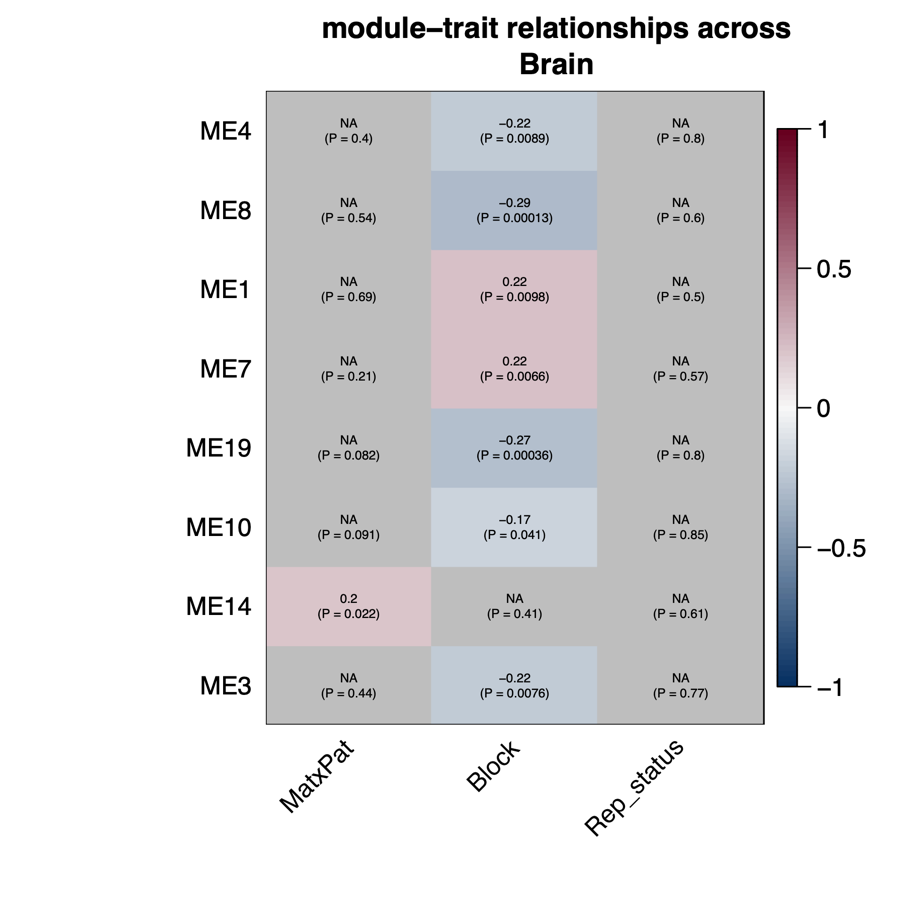
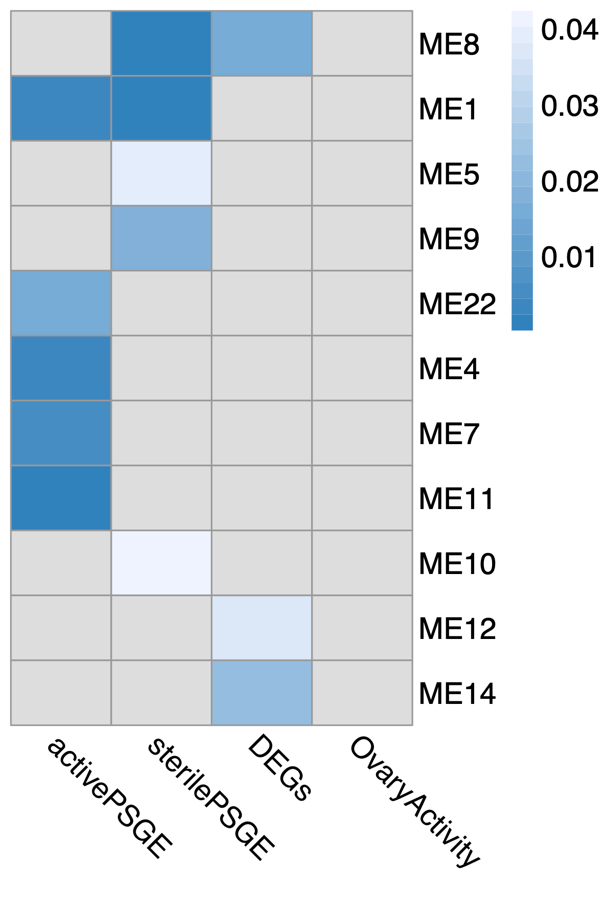
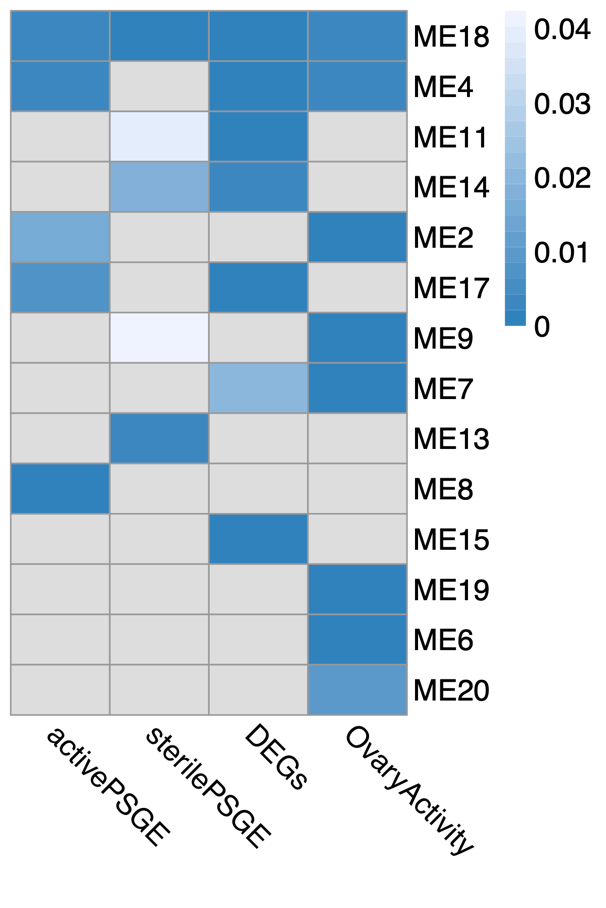


Figure 2. Gene networks in the brain and reproductive tissue. In brain tissue (left), eight gene modules identified in WGCNA are correlated with paternal background and genetic block (GLM, P<0.05). In reproductive tissue (right), 12 genetic modules were correlated with at least one fixed effect in the generalized linear model. Each row represents a separate generalized linear model for the relationship between average expression of genes in each module (using a module representative “eigengene”) and three traits, parental background, block, and reproductive status. Significant correlations are colorized according their correlation coefficient, with positive values in red and negative values in blue. P-values are shown in parentheses, and correlation coefficients are displayed for significant effects (GLM, P<0.05).



Figure 3. Gene modules were significantly enriched for both genes that showed differential gene expression and allele-specific expression (Fisher’s exact tests, P<0.05). (A) Brain tissue modules were compared to DEGs and PSGEs from the brain. (B) Reproductive tissue modules were compared to DEGs and PSGEs from reproductive tissue. The last column indicates whether the module was a significant correlation with ovary activation. Colors are proportional to the P-value associated with each significant overlap, and non-significant overlaps are shown in gray (P>0.05).