Supplementary material for:

Genomic Analysis of Genotype by Social Environment Interaction for *Drosophila* Aggressive Behavior

Palle Duun Rohde[*, †, ‡, 1], Bryn Gaertner[§, **], Kirsty Ward[§, **], Peter Sørensen[*], and Trudy F. C. Mackay[§,**]

[*]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark; [†]*i*PSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, 8000 Aarhus, Denmark; [‡]*i*SEQ, Center for Integrative Sequencing, Aarhus University, 8000 Aarhus, Denmark; [§]Department of Biological Sciences and Program in Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA; [**]W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, North Carolina 27695, USA

1

Supplementary Tables

Supplementary Figures

Tables attached as Excel files

Table S1   Line mean, standard error (SE), adjusted line mean and number of replicates (n) for DGRP lines reared in a social or socially isolated environment. A measure of genotype-by-social-environment interaction (GSEI) was computed as the difference between adjusted line mean for the social and socially isolated environment within each DGRP line.

Table S2   Single marker regression results ($p < 1 \times 10^{-5}$).

Table S3   Predictive ability (PA) for GO terms that significantly (adjusted p-value < 0.05) increased the PA compared to the GBLUP model using all markers. SE: standard error. p-value: the p-value from a t-test comparing the PA of the GFBLUP and GBLUP (NULL) models. Adjusted p-value: p-value adjusted for multiple testing. BP: biological processes. MF: molecular function. CC: cellular component.

Table S4   Genetic decomposition of predictive GO terms (adjusted p-value < 0.05) to the genes constituting each GO term. V_G: Genetic variance. Scaled V_G: The genetic variance explained per gene, adjusted for the number of SNPs, was scaled such the total genetic variance within the GO term summed to one. Only genes explaining $>10\%$ Scaled V_G are listed below.
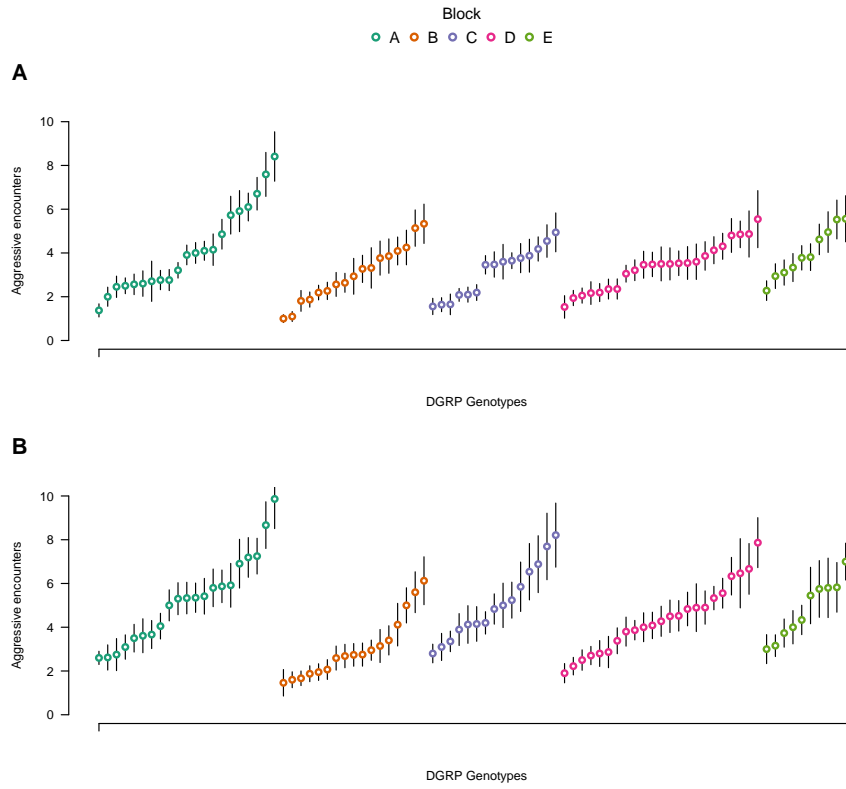
Supplementary Figures

Figure S1: Average number of aggressive encounters (unadjusted) for each DGRP genotype in (A) the socialized and (B) socially isolated environment, sorted according to increasing trait value within the block they were assayed.
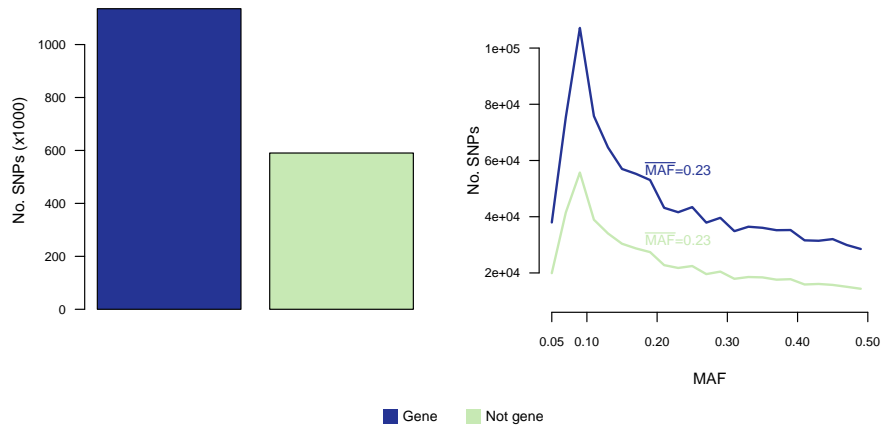
Figure S2: The left panel shows the number of SNPs (x1000) that were annotated to gene regions (i.e. transcribed part), and SNPs having no gene annotation. The right panel shows the minor allele frequency (MAF) distribution of the two groups.
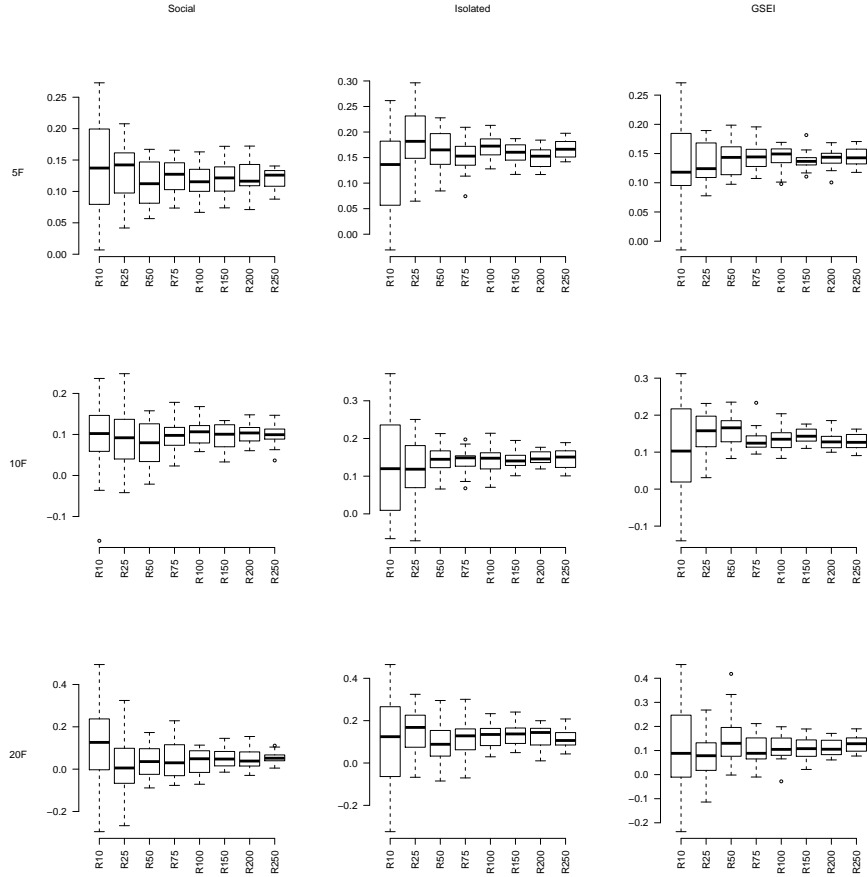
Figure S3: Box plots of the predictive ability for the NULL model (GBLUP) using different cross validation schemes; 5-fold (first row), 10-fold (second row) and 20-fold (third row). For each trait (socially reared flies (first column), socially isolated flies (second column), and GSEI (third column)), and cross validation scheme, the number of training and validation sets varied, $[10, 25, 50, 75, 100, 150, 200, 250]$, to find the number required to obtain a stable value of the predictive ability.
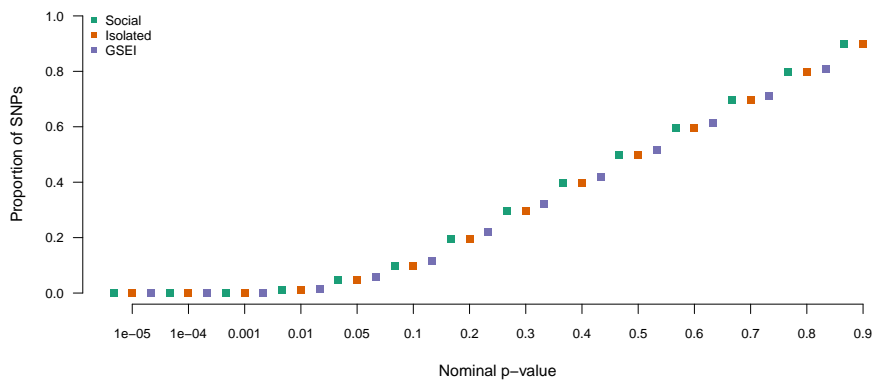
Figure S4: Average proportion of SNPs within each *p*-value cutoff for the 100 training sets for the social (green), isolated (orange) and GSEI (blue). Standard deviations are not shown because they were very small, and cannot been easily seen on the plot.
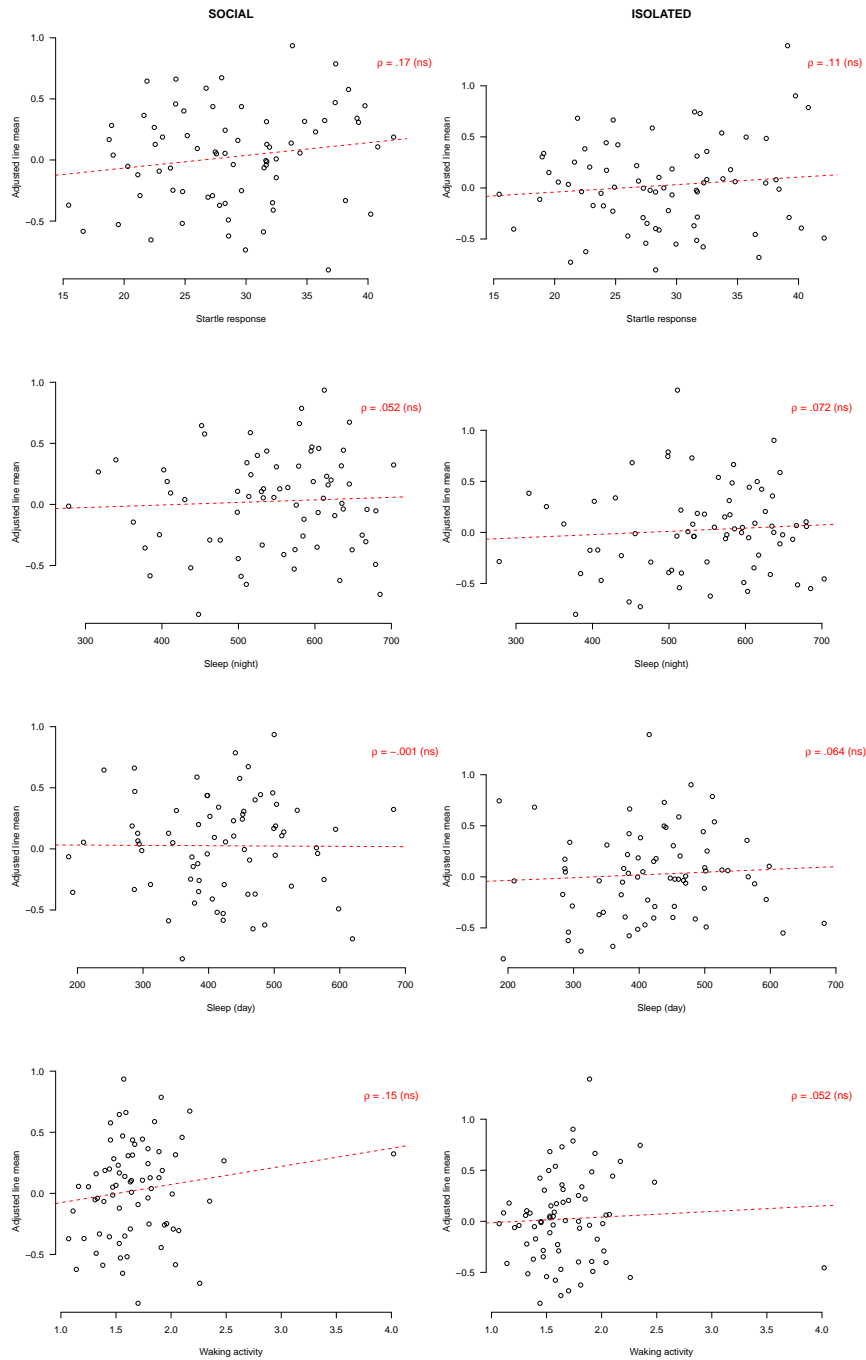
Figure S5: Phenotypic correlations between adjusted line means of aggressive behavior of DGRP lines from the social environment (left panel) or from the socially isolated environment (right panel) and different measures of locomotor activity; startle response (Mackay *et al.* 2012), sleep during night, sleep during day and waking activity (Harbison *et al.* 2013). The Pearson correlation, ρ, is noted on each panel (all correlations were non-significant, ns).
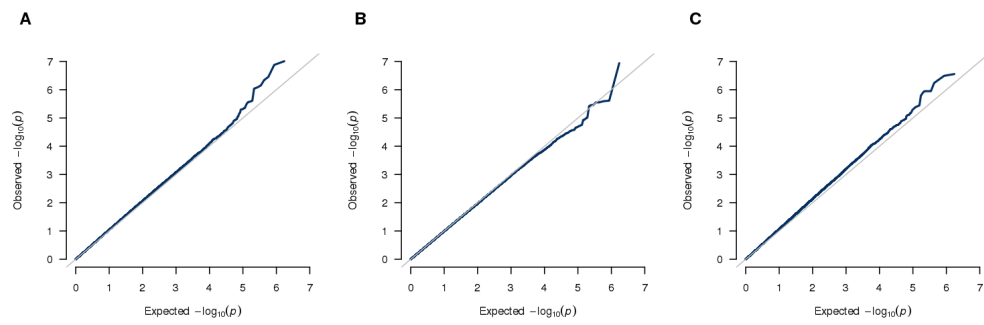
8

Figure S6: Quantile-quantile plots of genome-wide association study using single marker associations for aggressive behavior of flies reared in a socialized environment (**A**), socially isolated environment (**B**), and for GSEI (**C**).
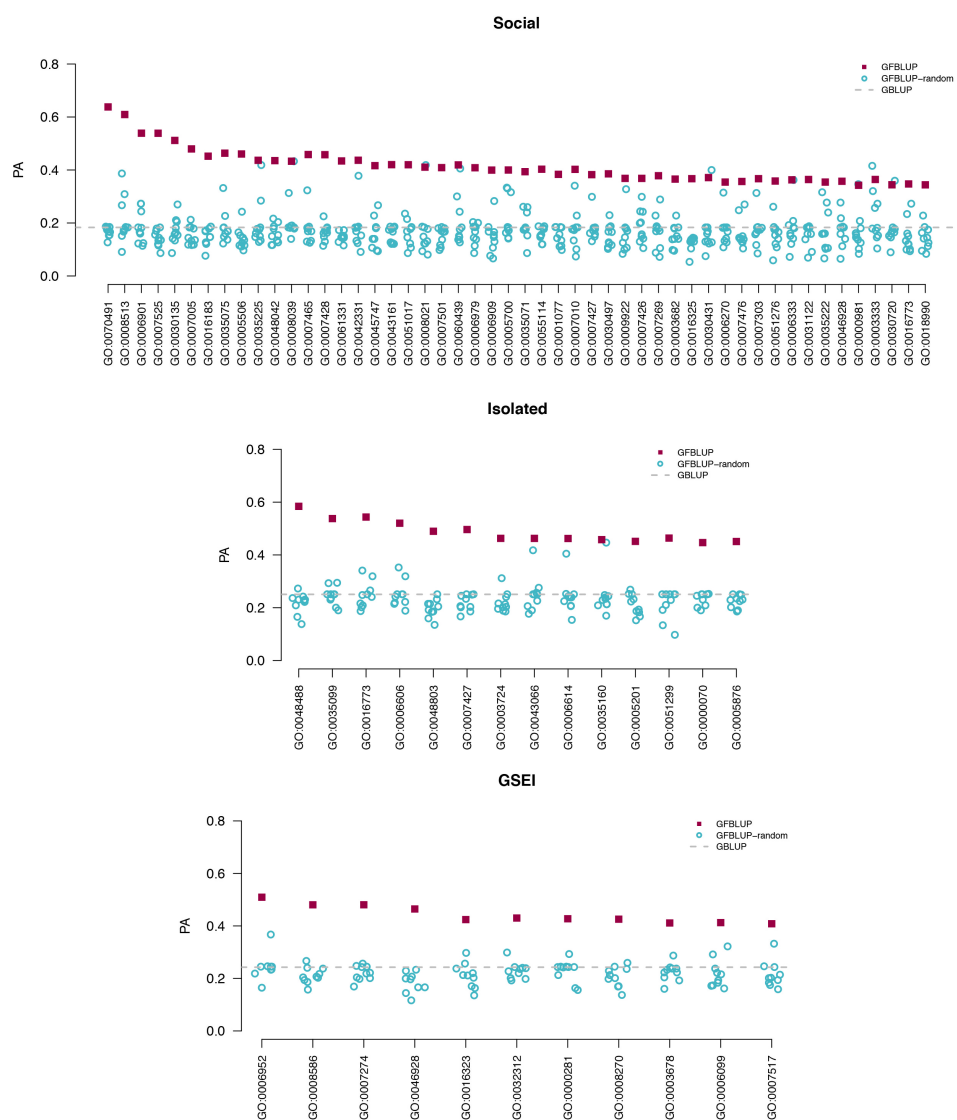
Figure S7: Mean predictive ability of the predictive GO terms (squares) and their corresponding randomly generated GO terms (circles). Horisontal dashed lines indicate the mean predictive ability of the GBLUP model.
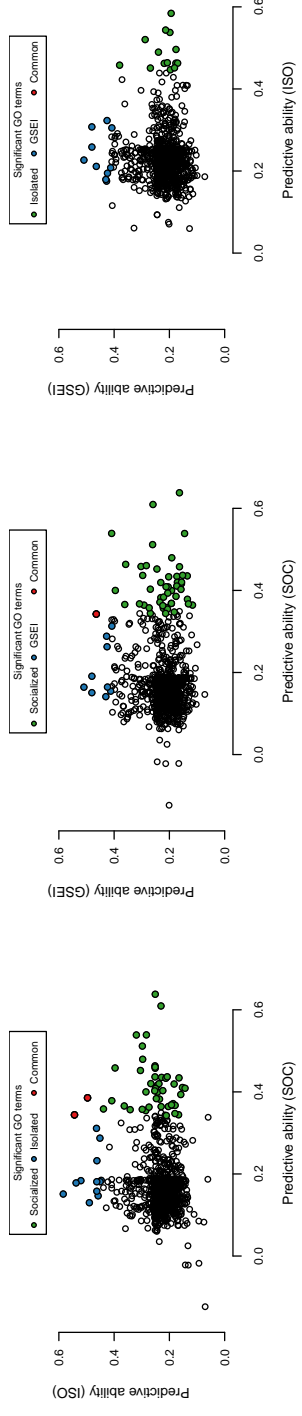
Figure S8: Comparison of the GO term predictive abilities for flies reared in a socialized (SOC), socially isolated (ISO), and for GSEI. Each panel shows the different comparisons. Filled symbols indicate if a GO term significantly increased the predictive ability (see legends).
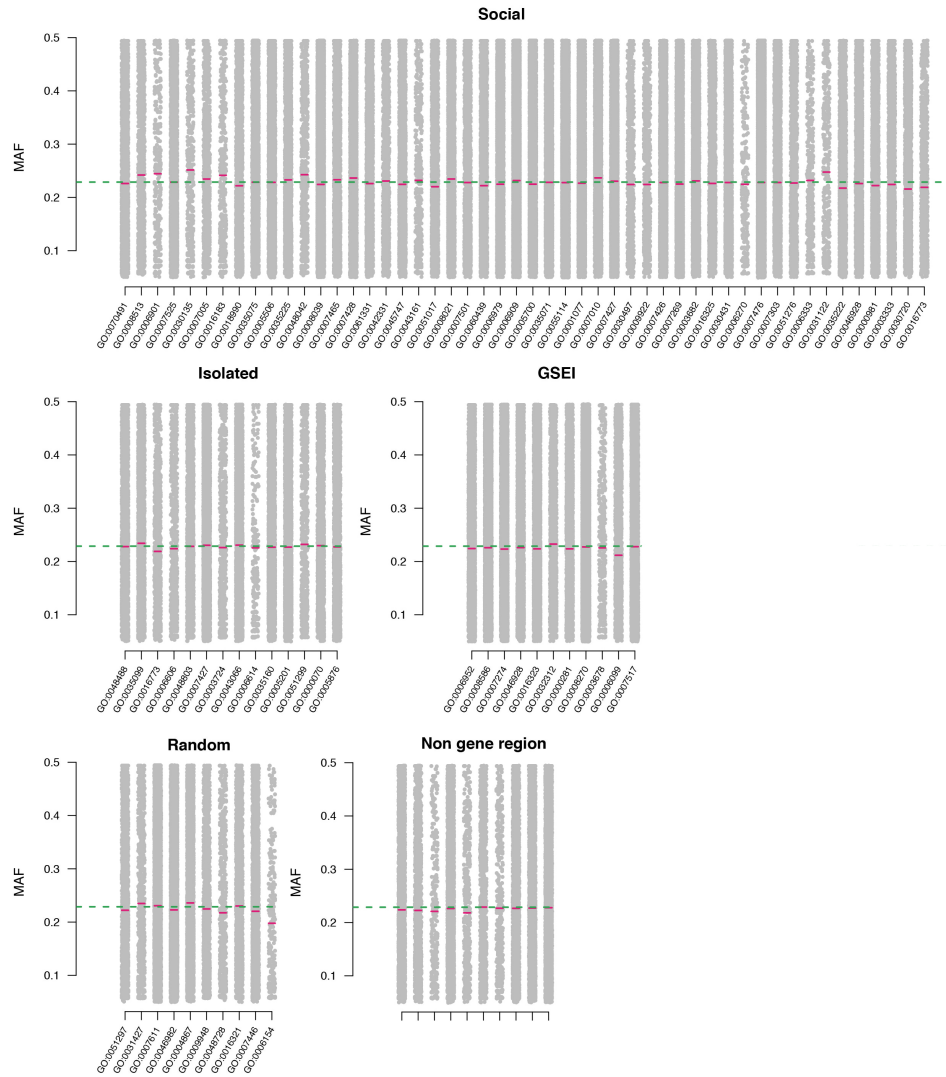
Figure S9: Allele frequency distribution within predictive GO terms for the social and isolated environment, as well as for GSEI. These are compared to ten randomly sampled GO terms, and to ten groups of randomly selected SNPs outside the gene regions (the ten groups contain the same number of SNPs as the first ten GO terms for the social environment). Horizontal pink lines indicate GO specific mean MAFs, whereas the horizontal green lines are the overall mean MAF.
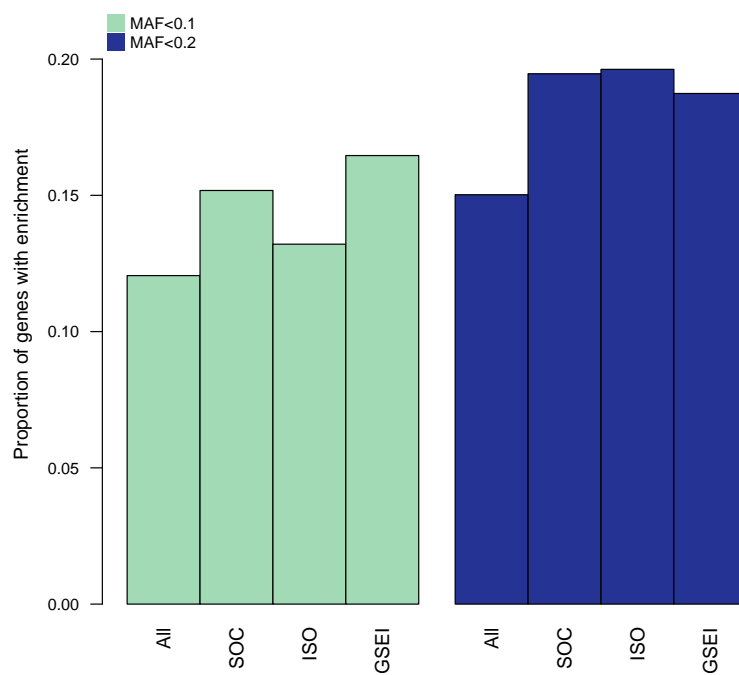
Figure S10: Proportion of all genes ('All') and genes within the predictive GO terms for socialized ('SOC') and socially isolated ('ISO') flies, and for GSEI, that were significantly enriched for SNPs with a minor allele frequency (MAF) below 0.1 (green) and 0.2 (blue).
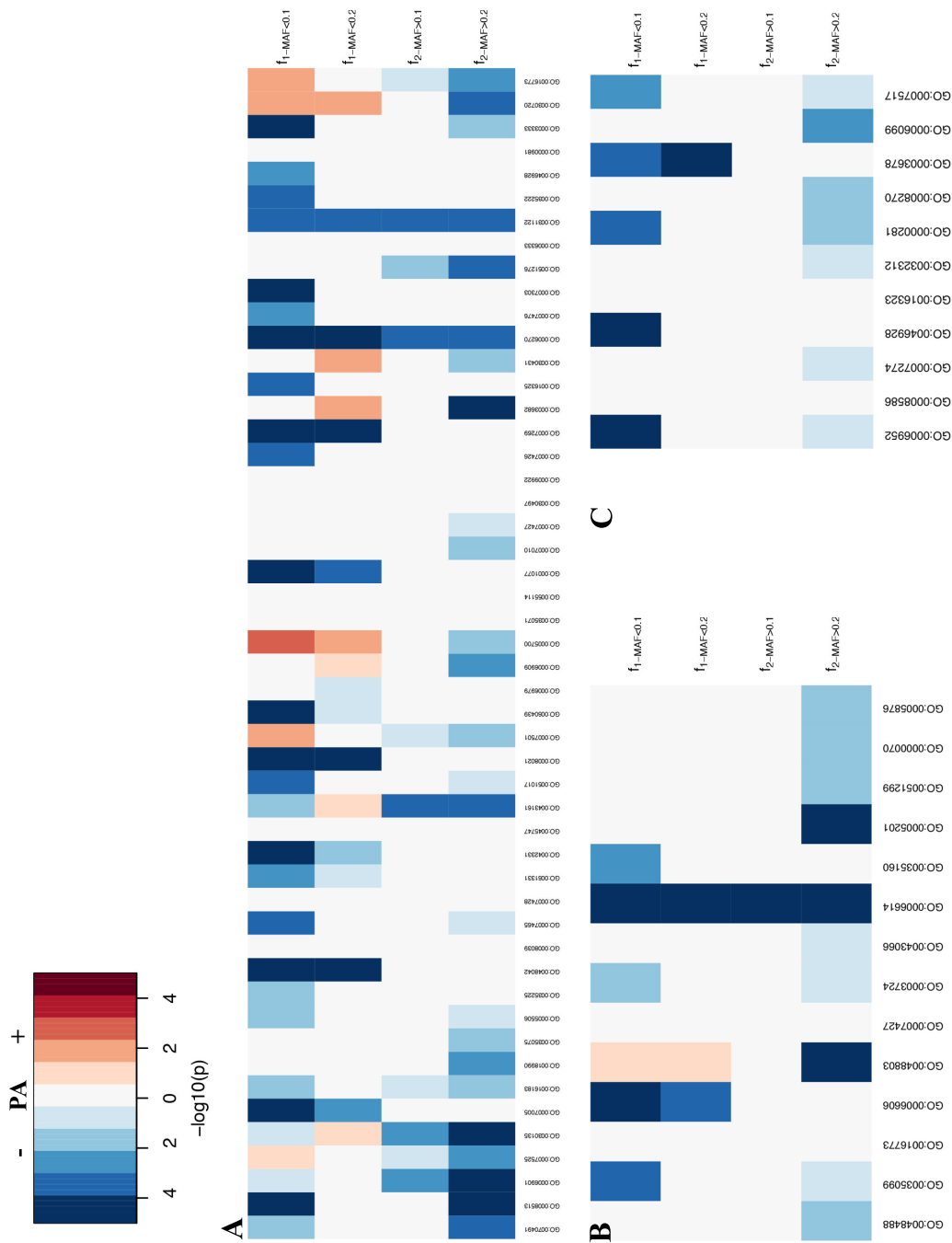
Figure S11: Heatmaps (A: social environment, B: socially isolated environment, C: GSEI) showing the direction (blue scale: predictive ability reduced, red scale: predictive ability incrased), and significance (as $-log_{10}(p)$)) of GFBLUP models when the feature only contains variants from the feature group with a minor allele frequency (MAF) below 0.1 ($f_{1-MAF<0.1}$ and $f_{1-MAF<0.2}$), or is based on variants from the feature group with a minor allele frequency (MAF) above 0.1 ($f_{2-MAF>0.1}$ and $f_{2-MAF>0.2}$). The p-value is from a t-test between the GFBLUP PA for the GO term and the PA from the GFBLUP model with the reduced feature group.