**Supplementary Table**

**Supplementary Table 1:** Anxiety-related traits measured on 258 Diversity Outbred mice used in case study of Reference Trait Analysis.

|  |  |
| --- | --- |
| Group | Phenotype |
| Light-dark box | Distance traveled |
| Light-dark box | Light-dark transitions |
| Light-dark box | Percent time in light (first four minutes) |
| Light-dark box | Percent time in light (total) |
| Light-dark box | Percent time in light, slope |
| Open-field | Distance traveled (first four minutes) |
| Open-field | Distance traveled (total) |
| Open-field | Distance change (first – last) |
| Open-field | Percent time in corner |
| Open-field | Percent time in corner, slope |
| Open-field | Percent time in periphery |
| Open-field | Percent time in periphery, slope |
| Open-field | Percent time in center (square-root transformed) |
| Open-field | Percent time in center, slope |
| Open-field | Percent time mobile |
| Open-field | Fecal boli count |

**Supplementary Figure Legends**

**Supplementary Figure 1:** Schematic comparing overall strategies of Reference Trait Analysis and TWAS. For Reference Trait Analysis, canonical correlation analysis is used to relate traits of interest to reference traits (blue, 1) and coefficients derived from this model are applied to reference traits in the cohort without measurements of traits of interest (blue, 2). Finally, these projected reference traits are compared to gene expression to identify trait-gene expression correlations (blue, 3). In the TWAS approach, genotypes are used to build models that predict gene expression through eQTL (red, 1). These models are applied to genotypes in the cohort without gene expression measurements (red, 2) and imputed gene expression is compared with traits of interest to identify trait-gene expression correlations (red, 3). Note that training and testing cohort labels are switched for the two methods but that the end result of each is to compare traits of interest with gene expression (grey dashed line, middle).

**Supplementary Figure 2:** Comparison of TWAS and Reference Trait Analysis using a single random division of the mouse anxiety dataset. For both panels we take the true trait of interest to be the first canonical variable of open-field traits (open-field CC1). For TWAS we used genotypes to impute gene expression. Left panel shows correlation of individual transcripts to the trait of interest, where the *x*-axis plots correlations based on true transcript abundance and the *y*-axis plots correlations based on imputed transcript abundance. Right panel shows the analogous result but using Reference Trait Analysis, where gene expression is fixed and predictors of open-field behavior are represented by projected traits.

**Supplementary Datasets**

**Supplementary Dataset 1:** Normalizedhippocampal gene expression matrix. RNA-Seq data were processed as described (Methods). To obtain normalized gene expression matrix, raw counts in each sample were normalized to the upper quartile value and transformed to normal scores.

**Supplementary Dataset 2:** Traits derived from open-field arena exploration assay and used in case study of Reference Trait Analysis. Supplementary Table 1 provides basic information on phenotypes, while complete details of animal rearing, husbandry and phenotyping are presented in Logan et al. (2013).

**Supplementary Dataset 3:** Traits derived from light-dark box behavior assay and used in case study of Reference Trait Analysis. Supplementary Table 1 provides basic information on phenotypes, while complete details of animal rearing, husbandry and phenotyping are presented in Logan et al. (2013).