Supplemental Methods

Metabolite profiling

Mature seeds were dehulled manually and processed following **?**. Briefly, dehulled seeds were homogenized, and 100 mg of pulverized tissue was used to separate polar and non-polar compounds using a biphasic extraction method. A set of QC samples was created by combining 60 μL of the upper organic layer from each sample, as well as 60 μL of the lower aqueous phase. 600 μL of the upper organic layer and 1.7 mL of the lower aqueous phase were transferred to new glass vials, and were dried under nitrogen gas overnight. Aqueous fractions were re-suspended in 1.5 mL of 50% methanol 50% water (v/v), while organic fractions were re-suspended in 0.7 mL of 50% methanol 50% methyl tert-butyl ether (MTBE). Organic fractions were used to quantify fatty acid methyl esters, however these data were not included in our study.

LC-MS Phenyl-Hexyl Analysis

A total of 5 μL of the aqueous fraction was injected onto a Waters Acquity UPLC system in five blocks, with every fifth sample being a pooled QC sample. A Waters Acquity UPLC CSH Phenyl Hexyl column (1.7 μM, 1.0 x 100 mm) using a gradient from solvent A (2mM ammonium hydroxide, 0.1% formic acid) to solvent B (Acetonitrile, 0.1% formic acid) was used for separation. Injections were made in 100% A, held at 100% A for 1 min, increased to 98% B over 12 minutes, held at 98% B for 3 minutes, and then returned to starting conditions over 0.05 minutes and allowed to re-equilibrate for 3.95 minutes, with a 200 μL min^{-1} constant flow rate. The column was held at 65 °C, while samples were maintained at 6 °C. The column eluent was infused into a Waters Xevo G2 TOF-MS with an electrospray source in positive mode, scanning 50-2000 m/z at 0.2 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium iodide with 1 ppm mass accuracy. The capillary voltage was held at 2200 V, source temp at 150 °C, and nitrogen desolvation temperature at 350 °C with a flow rate of 800 L hr^{-1}.

GC-MS Non-targeted Analysis

A 40 μL aliquot of the aqueous phase re-suspension was dried under nitrogen, re-suspended in 50 μL of pyridine containing 25 mg/mL of methoxyamine hydrochloride, incubated at 60° for 1 h, sonicated for 10 min, and incubated for an additional 1 h at 60°C. Next, 50 μL of N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS, Thermo Scientific) was added and samples were incubated at 60 °C for 45 min, briefly centrifuged, cooled to room temperature, and 100 μL of the supernatant was transferred to a 150 μL glass insert in a GC-MS autosampler vial. Metabolites were detected using a Trace 1310 GC coupled to a Thermo ISQ mass

spectrometer (Thermo Scientific). Samples were injected in a 1:10 split ratio. Separation occurred using a 30 m TG-5MS column (Thermo Scientific, 0.25 mm i.d., 0.25 μm film thickness) with a 1.2 mL/min helium gas flow rate, and the program consisted of 80 °C for 30 sec, a ramp of 15 °C per min to 330°C, and an 8 min hold. Masses between 50-650 m/z were scanned at 5 scans sec^{-1} after electron impact ionization.

Metabolite data analysis

Preprocessing

Molecular features, defined by retention time and mass-charge (m/z), were generated for each sample using `XCMS` package in R (**??**). The `matchedFilter` algorithm was used for GC-MS data, and the `centWave` algorithm for LC-MS data. `RAMClustR` was used to group features, with normalization set to `'none'` (**?**). Annotation of GC-MS features were performed by searching the GOLM metabolome database using the `RAMSearch` program (**?**).

Annotation of LC-MS feature was performed using a combination of approaches. First, searches were conducted against an in-house spectra and retention time database using `RAMSearch`. Searches against the MassBank database using MSFinder was also used to perform a spectral search against the MassBank database (**??**). Finally, all results were imported into R and a collective annotation was derived with the prioritization of annotations derived from RAMSearch placed over those from MSFinder.

The pooled QC samples were used to correct for signal intensity drift within and between batches on a metabolite by metabolite basis using a random forest (RF) approach implemented in the `statTarget` package in R (**?**). Prior to signal correction, one sample was omitted due to low metabolite detection. The RF-based approach learns a correction factor based on regression of metabolite intensities within each QC sample on the injection order of the QC samples. Principal component analysis was used to assess grouping/clustering of QC and experimental samples. This approach showed tight clustering of QC samples across batches and reduced clustering of experimental samples across batches. All data was transformed using a cube-root transformation to encourage normality of relative intensities for each metabolite.

Calculation of Best Linear Unbiased Predictors for metabolites

To remove systematic effects, best linear unbiased predictors (BLUPs) were calculated for each metabolite and accession. The linear mixed model is given by

$$y = \mu + DTH + check + new : entry + block + batch + e \tag{1}$$

where *check* is a fixed effect for each of the seven check varieties; *new* is an indicator variable where 0 indicates a check variety and 1 indicates an unreplicated entry, and is nested within entry; DTH is a

fixed covariate that provides days to heading for each observation; *block* and *batch* are random effects to account for field blocks and injection batch for MS, respectively. The terms μ and e represent the overall mean and the vector of residuals, respectively. We assume entries are unrelated. The above model was fitted using the `sommer` package in R (**?**). Deregressed BLUPs for each entry i and metabolite j were calculated following **?** using

$$\hat{g}^*_{ij} = \frac{\hat{g_{ij}}}{1 - \frac{PEV_{ij}}{\sigma^2_{g_j}}} \tag{2}$$

where $\hat{g_{ij}}$ is the BLUP for entry i and metabolite j, PEV_{ij} is the prediction error variance, and $\sigma^2_{g_j}$ is the total genetic variance. This provided BLUPs for the 368 unreplicated entries.

To identify potential outlier samples, the model above was fit using each of the top 20 principal components (PCs) for GC-MS and top 20 PCs LC-MS as response variables. For each PC and MS approach, we inspected diagnostic plots (i.e., Q-Q plots and standardized residual vs fitted values) to identify outlier samples. Particular emphasis was placed on leading PCs, as these account for more variance in metabolites. One sample showed consistent deviation from normality in multiple PCs and was excluded from downstream analyses. Thus, metabolomic data for 367 entries were used in subsequent analyses.

Genotyping, imputation and assessing population structure

Single-nucleotide polymorphism (SNP) data were compiled from 11 genotyping experiments for 539 lines, which consist of accessions in the diversity panel and those used for seed quality trials (discussed below; Table 1). Missing marker genotypes were imputed using the `glmnet` approach described by **?**. Prior to imputation, markers were excluded based on the following criteria: allele frequency < 0.02, proportion of missing data across individuals > 0.6, and heterozygosity > 0.1. While individuals were excluded if more than 70% of markers were missing or if 10% of the markers were heterozygous. Three hundred thirty-five lines with metabolomics data were extracted from these data. After imputation markers were removed with a minor allele frequency < 0.05 in the 335 lines, leaving 62,049 markers.

To identify subpopulations, a k-means clustering approach was performed for ten different sets of clusters/centers using the `kmeans` function in R (**?**), with 10,000 iterations and 100 randomly selected cluster centers. Prior to clustering, the marker matrix was scaled and centered so that each column (i.e., marker) had a mean of zero and unit variance. For each model, the Bayesian Information Criteria (BIC) was calculated as $BIC = D + mk\log(n)$, where D is the total within cluster sum of squares, m is the number of markers, n is the number of individuals, and k is the number of clusters.

GWAS on metabolites

We performed GWAS on deregressed BLUPs for 1,668 metabolites using a mixed linear model approach. The model and marker data is identical to what was used for GWAS on factor scores. We used the approach described by **?** to determine the number of effective tests ($M_{eff.}$) and used Sidak's correction to account for multiple testing (**?**). This resulted in a p-value threshold of $1.54x10^{-8}$. This analysis returned a total of 2,466 marker-metabolite associations for 129 compounds with 1,191 markers associated with one or more factors.

These results were compared to those obtained through factor score-based GWAS. We assigned the 1,191 markers recovered from metabolite-based GWAS into two groups: those that were identified with both GWAS on factor scores and GWAS on metabolites, and markers that were only recovered by GWAS on metabolites. Figure S30 compares the number of compounds that were associated with each marker for each group. These results show that associations that were detected with both factor score-based GWAS and metabolite-based GWAS are primarily associated with more than one compound, while marker-metabolite associations that were detected only with metabolite-based GWAS primarily affect one compound.

Table 1. Genotyping-by-sequencing experiments in Triticeae Toolbox used in this study.

Study Name in T3	Link
CornellMetabolomics_2016	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=CornellMetabolomics_2016
POGI_2015_SDSU	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_SDSU
POGI_2015_UMN	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_UMN
POGI_2015_other	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_Other
POGI_2015_ABERDEEN	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_Aberdeen
POGI_2015_SOAP	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_SOAP
POGI_2015_UOPN	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_UOPN
POGI_2017	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2017
POGI_2015_ENCORE	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_ENCORE
POGI_2015_UIL	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_UIL
POGI_2015_UWM	https://triticeaetoolbox.org/oat/genotyping/display_genotype.php?trial_code=POGI_2015_UWM

Table 2. Trials in Triticeae Toolbox used for training and testing genomic prediction models.

Traits	Study Name in T3	Link
Lipid and protein NIRS	Founders_2017_Lamberton	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_2017_Lamberton
	Founders_2017_Morris	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_2017_Morris
	Founders_2017_Crookston	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_2017_Crookston
	Founders_2018_Lamberton	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_2018_Lamberton
	Founders_Late_2018_Lamberton	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_Late_2018_Lamberton
	Founders_2018_Crookston	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=Founders_2018_Crookston
Fatty Acids	CornellMetabolomics_2014_McGowan	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=CornellMetabolomics_2014_McGowan
	CornellMetabolomics_2014_Caldwell	https://triticeaetoolbox.org/oat/display_phenotype.php?trial_code=CornellMetabolomics_2014_Caldwell

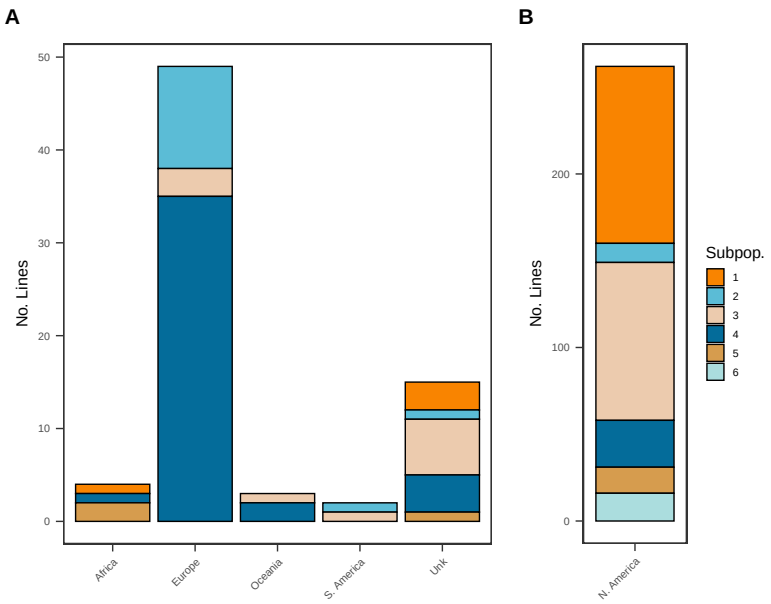

Figure 1. Summary of subpopulation clusters based on major geographic regions. Demographic histories for each line was obtained from T3/oat (`https://triticeaetoolbox.org/oat`). Breeding programs were used to assign lines to major geographical regions. In cases where lines were obtained from Germplasm Resources Information Network(GRIN; `www.ars-grin.gov`), the passport history was used to assign lines to a major geographical region. Unk indicates lines with an unknown origin.

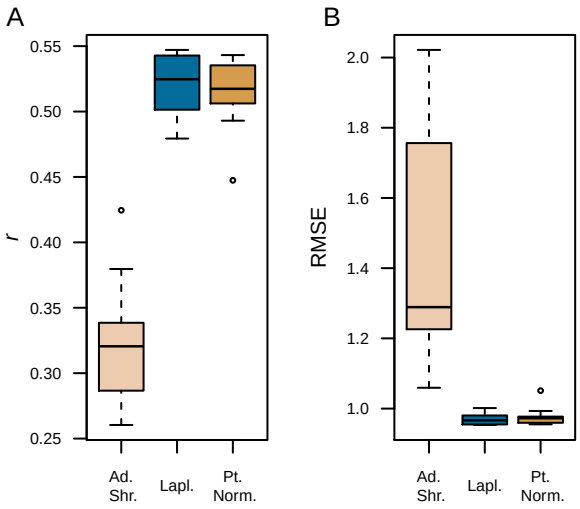

Figure 2. Three-fold orthogonal cross validation results for three EBMF approaches. Pearson's correlation (r) between predicted and observed values is shown in panel A, while root mean square error (RMSE) is shown in panel B. The metrics were calculated based on data in the testing set. CV was performed ten times, and results were averaged across folds within each resampling run.

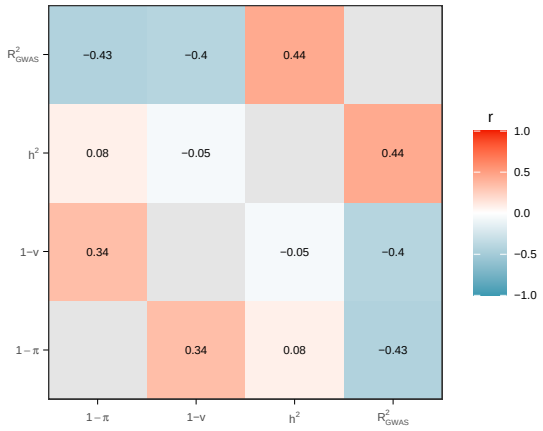

Figure 3. Spearman's correlation between factor density $(1-v)$, polygenicity $(1-\pi)$, narrow sense heritability (h^2), and the proportion of phenotypic variance explained by GWAS hits $(\mathbf{R^2_{GWAS}})$. The proportion of variance explained by significant GWAS hits was determined by comparing a model that included the first 10 principal components of the genomic relationship matrix and any marker with $p < 2.57 \times 10^{-7}$ with a model that included only the first 10 principal components. All terms were considered as fixed.

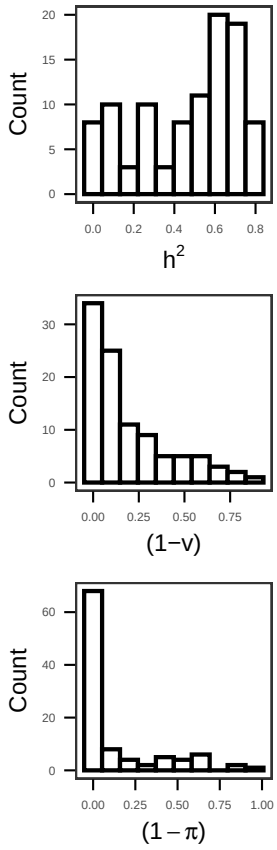

Figure 4. Distribution of narrow sense heritability (h^2), density ($1 - v$), and polygenicity ($1 - \pi$) estimates. Posterior means were used as estimates for polygenicity ($1 - \pi$) and narrow sense heritability (h^2).

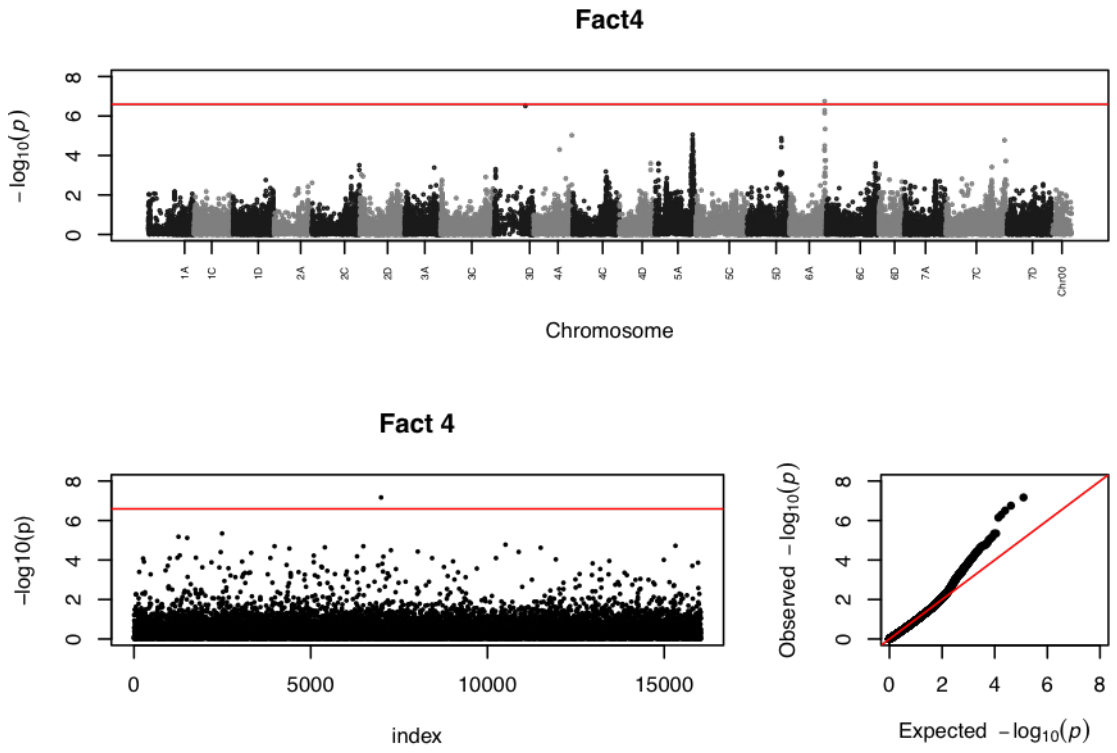

Figure 5. GWAS using scores for factor 4. The top panel shows $-log_{10}(p - values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

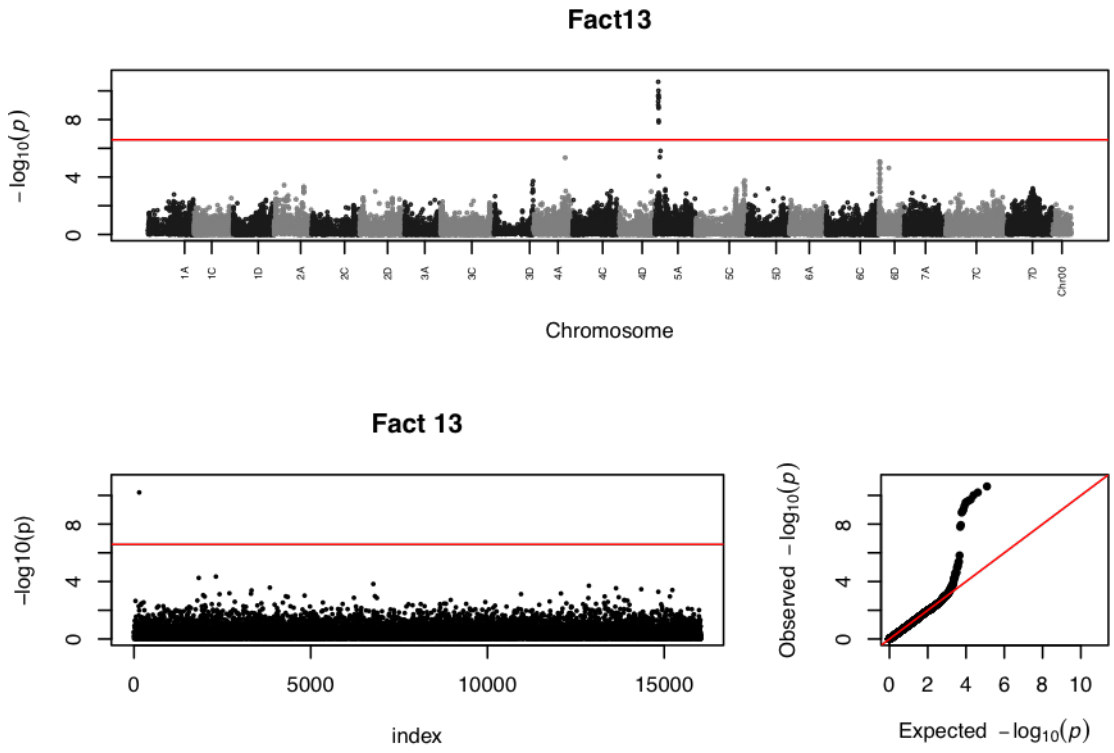

Figure 6. GWAS using scores for factor 13. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

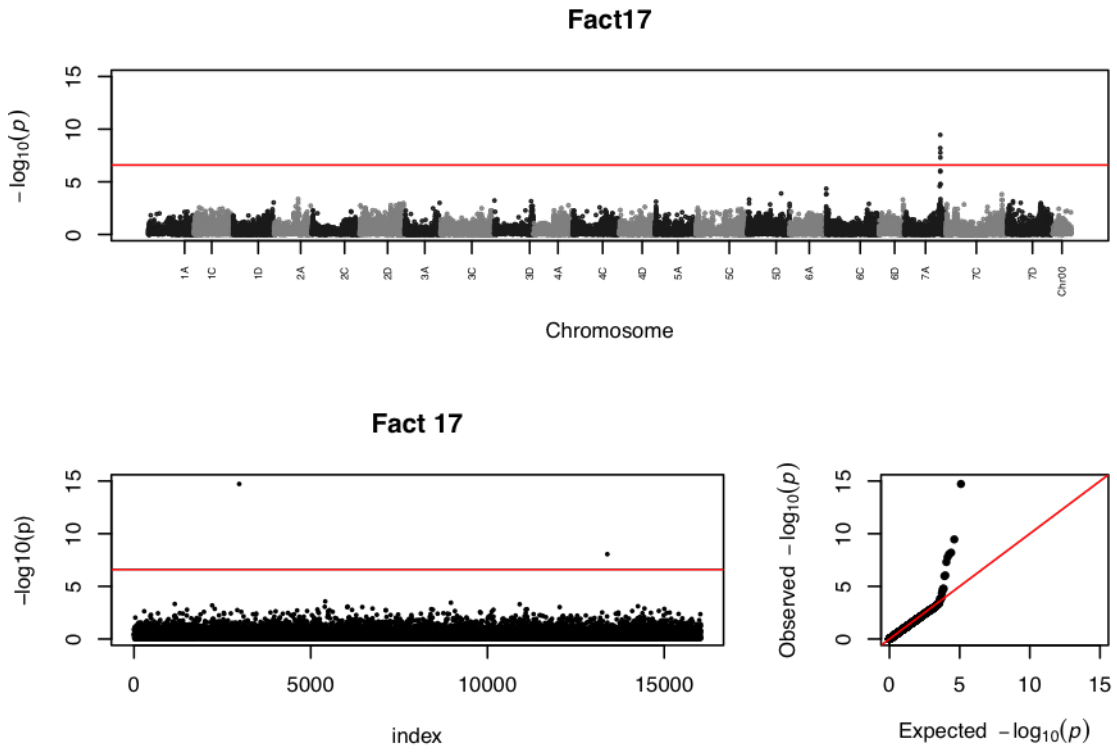

Figure 7. GWAS using scores for factor 17. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

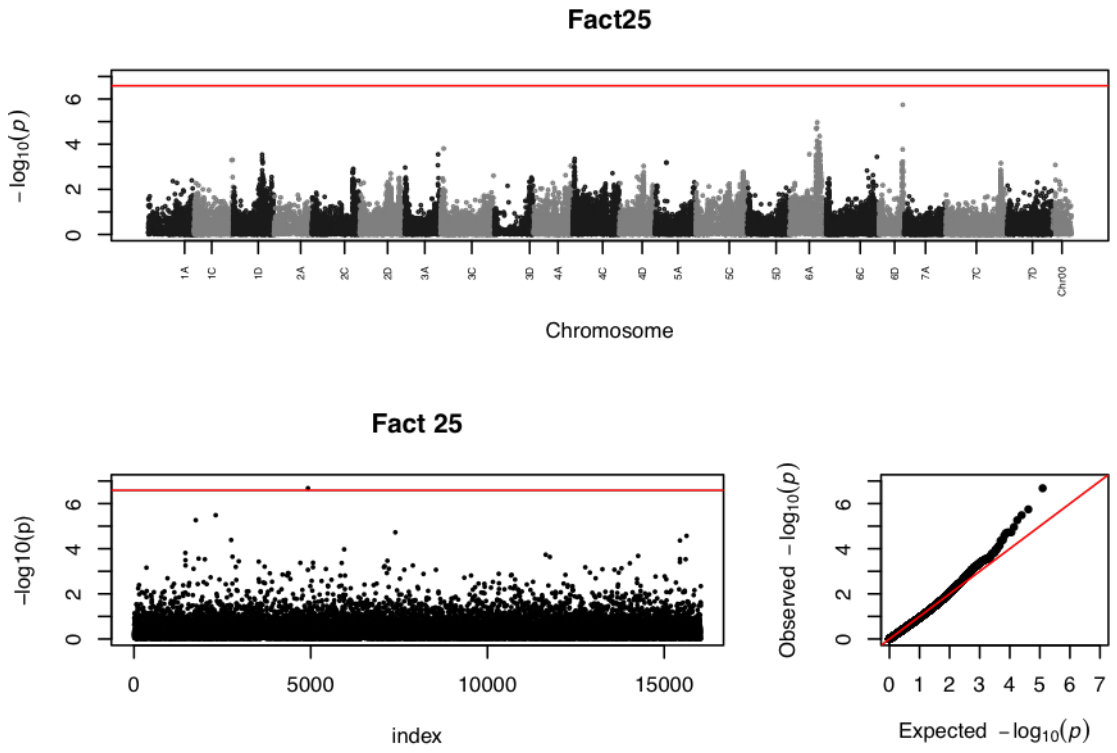

Figure 8. GWAS using scores for factor 25. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

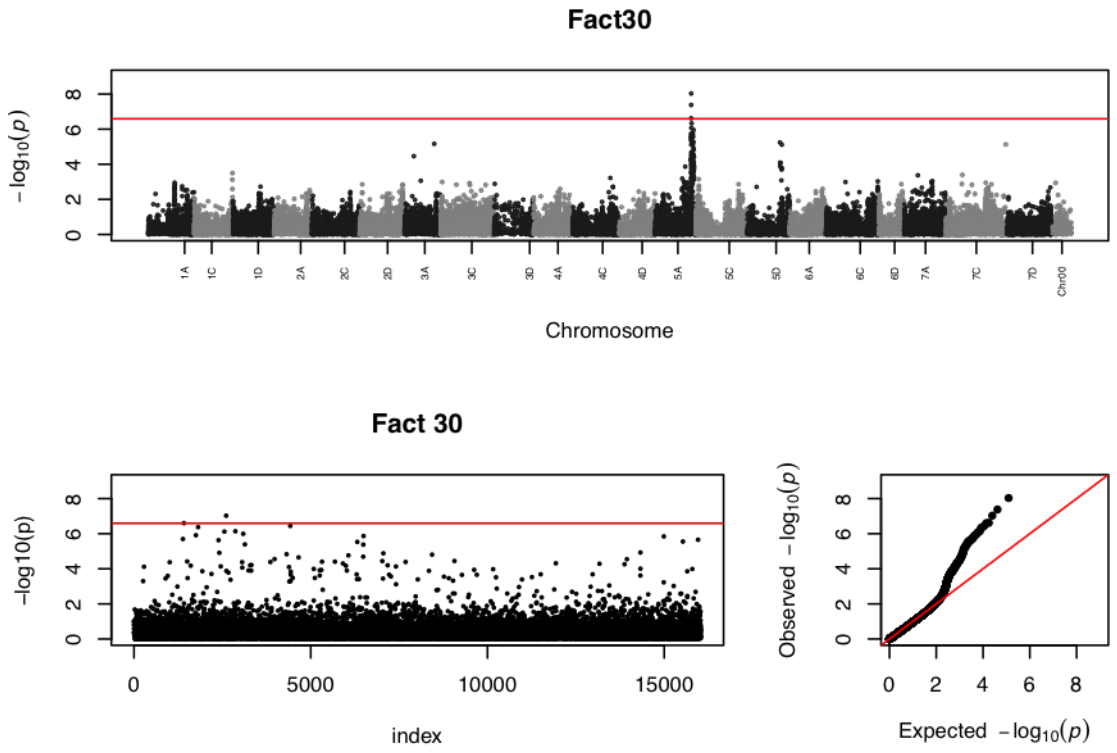

Figure 9. GWAS using scores for factor 30. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

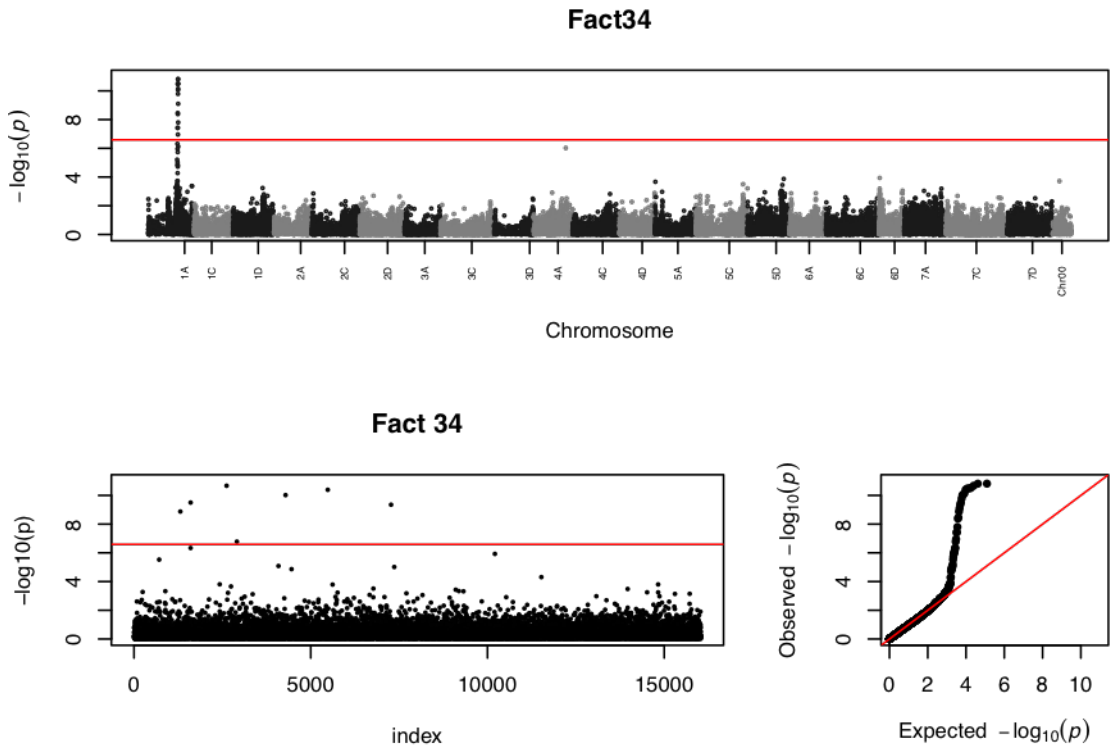

Figure 10. GWAS using scores for factor 34. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

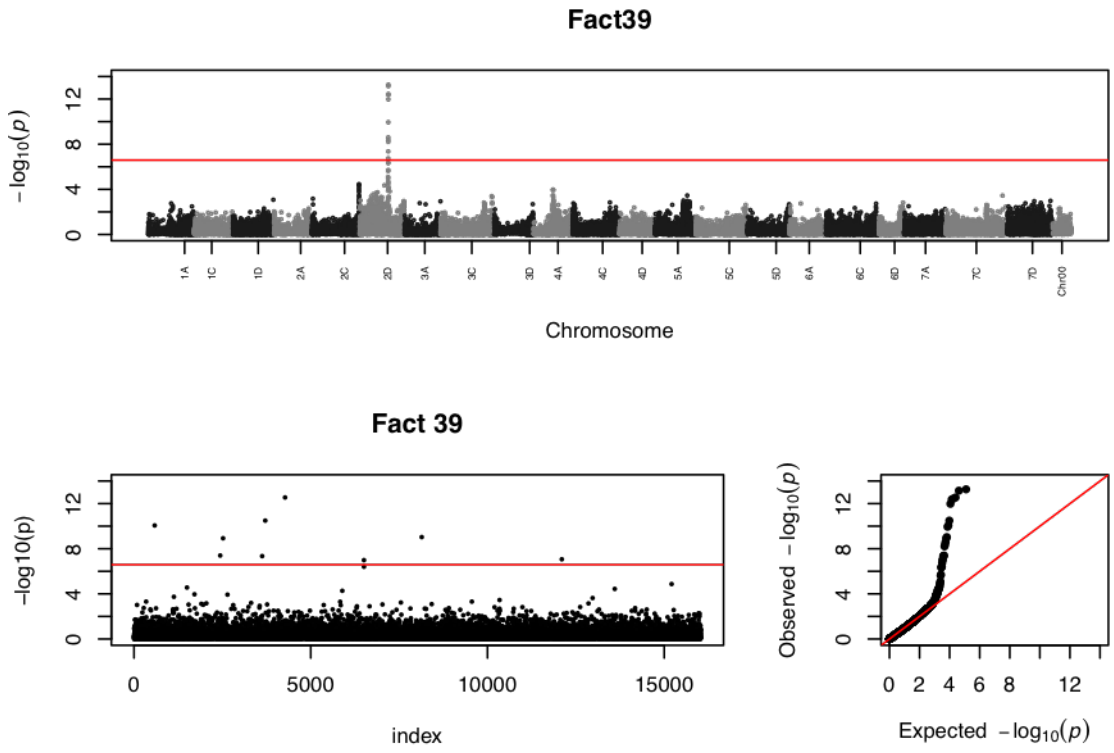

Figure 11. GWAS using scores for factor 39. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

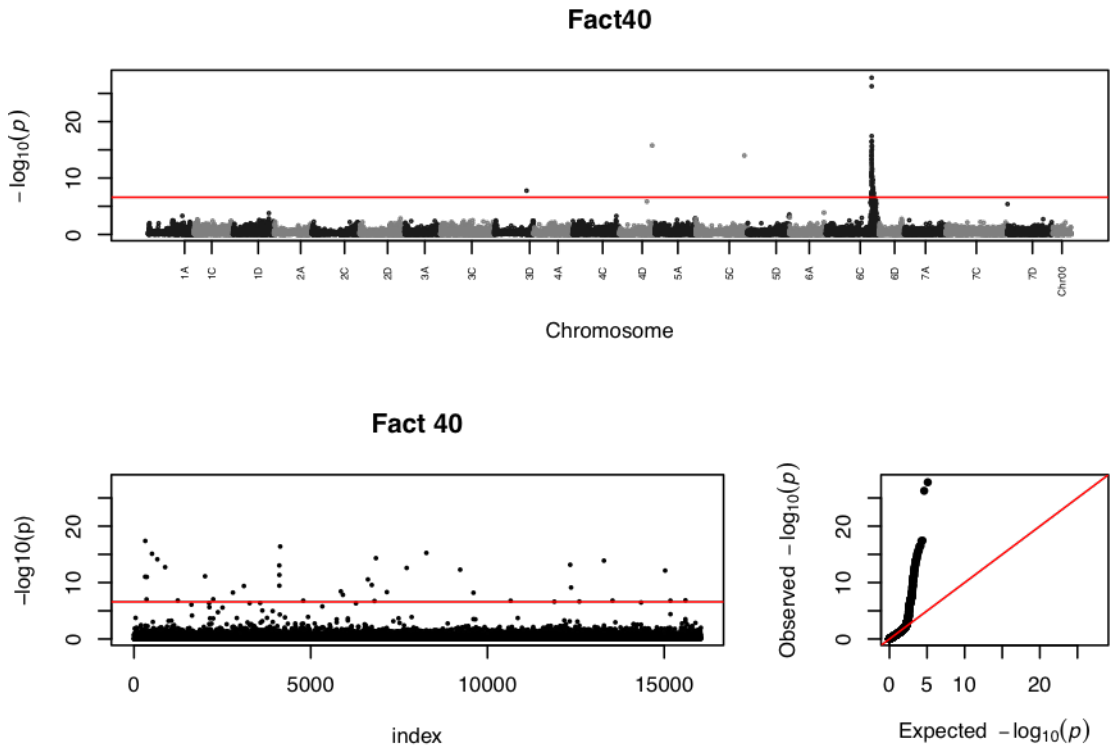

Figure 12. GWAS using scores for factor 40. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

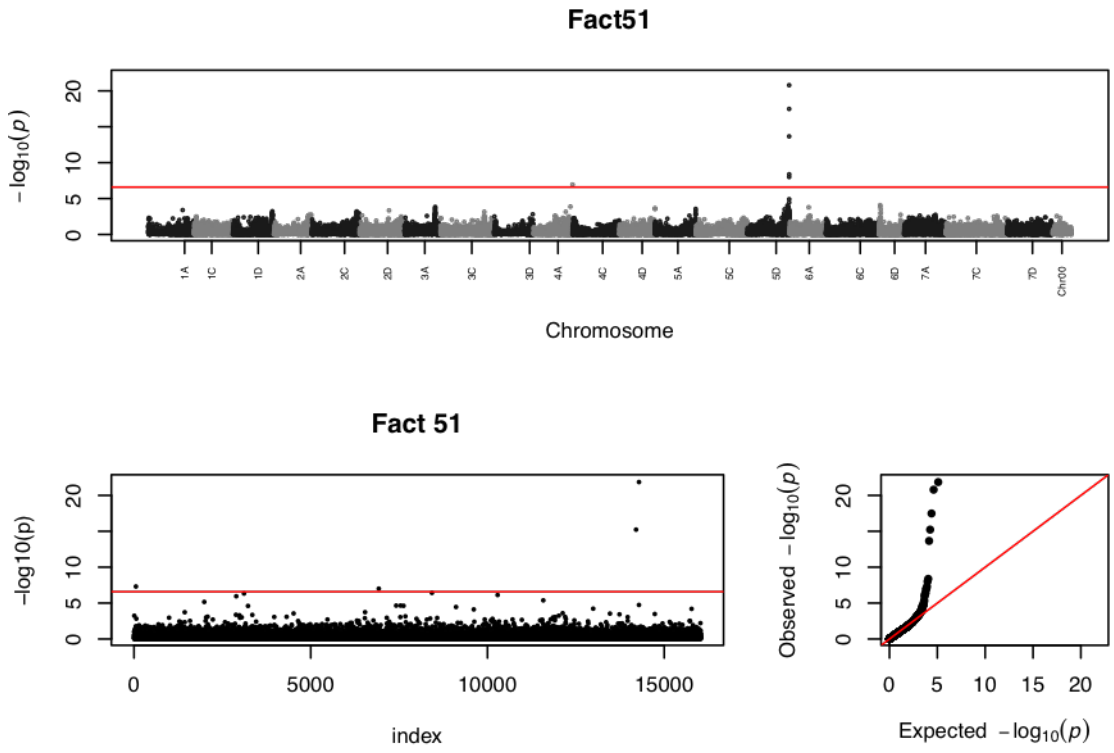

Figure 13. GWAS using scores for factor 51. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

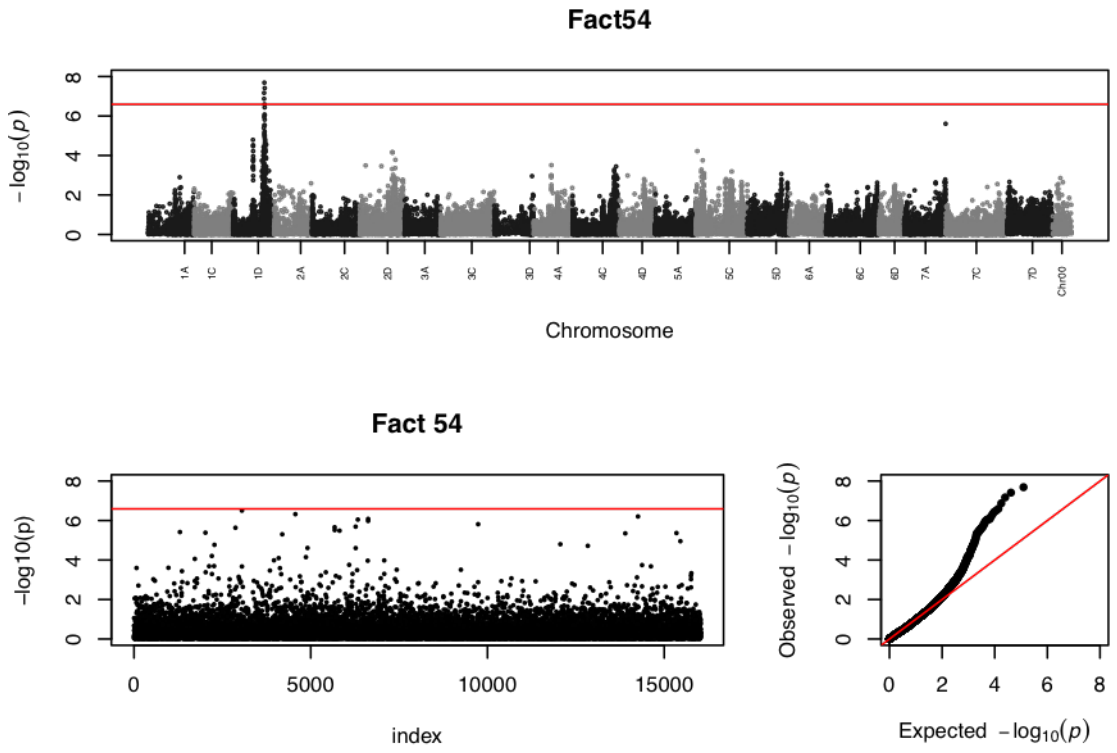

Figure 14. GWAS using scores for factor 54. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

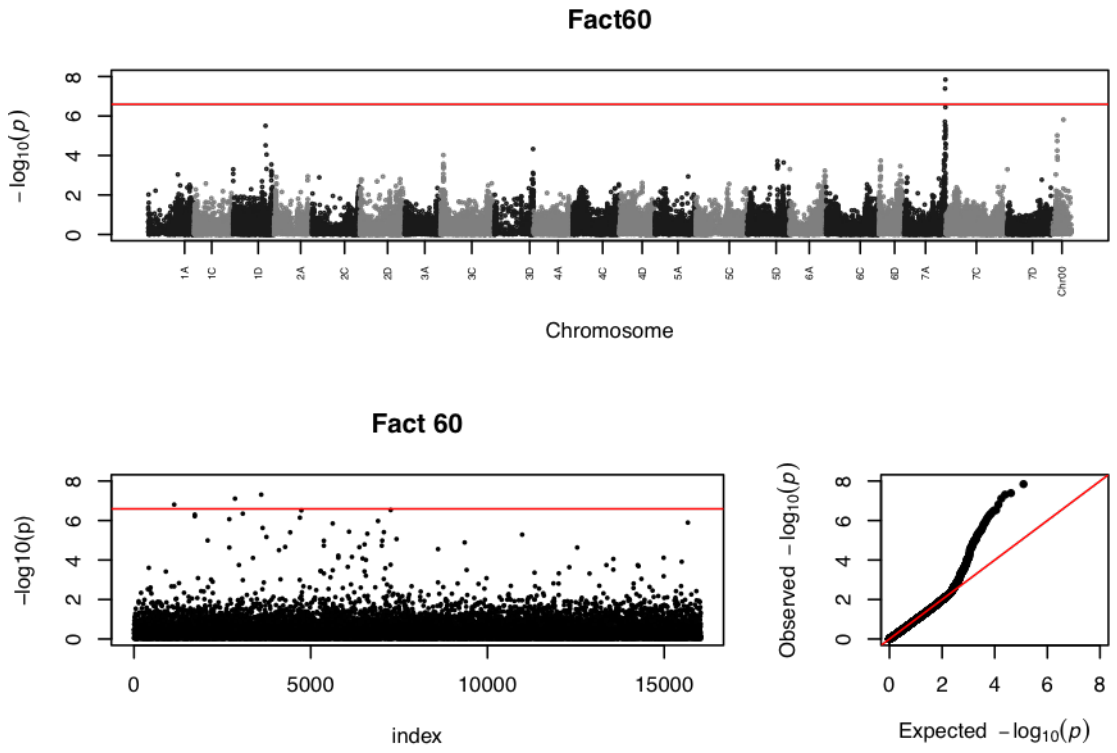

Figure 15. GWAS using scores for factor 60. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

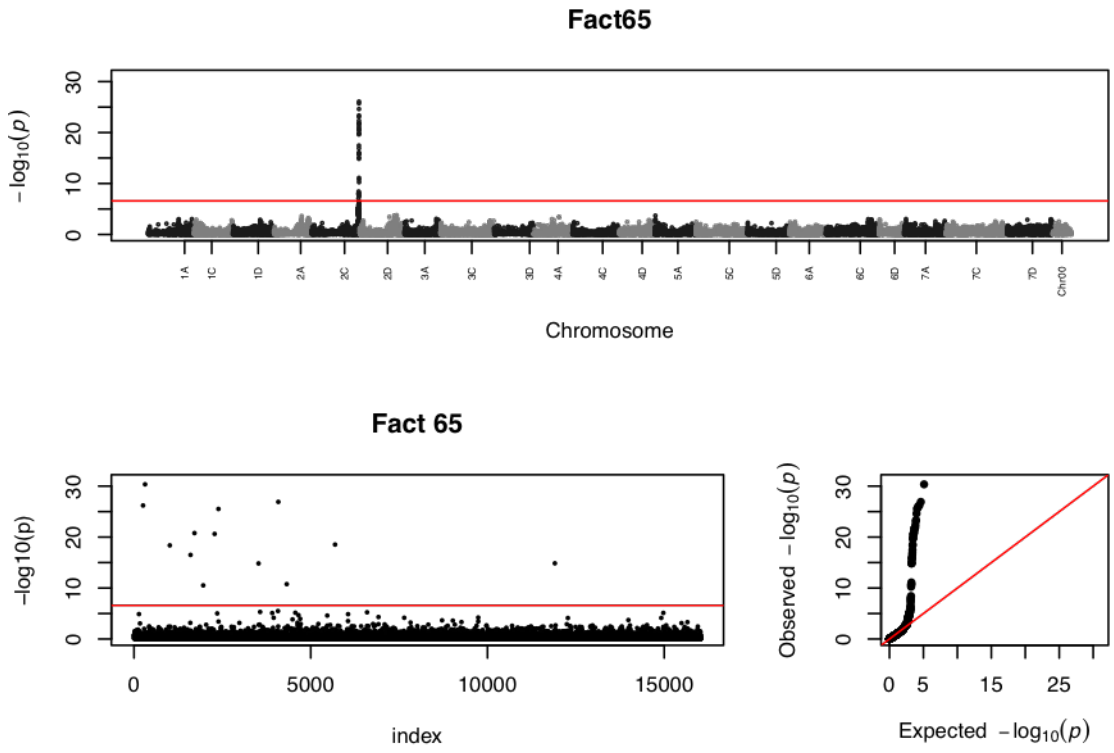

Figure 16. GWAS using scores for factor 65. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

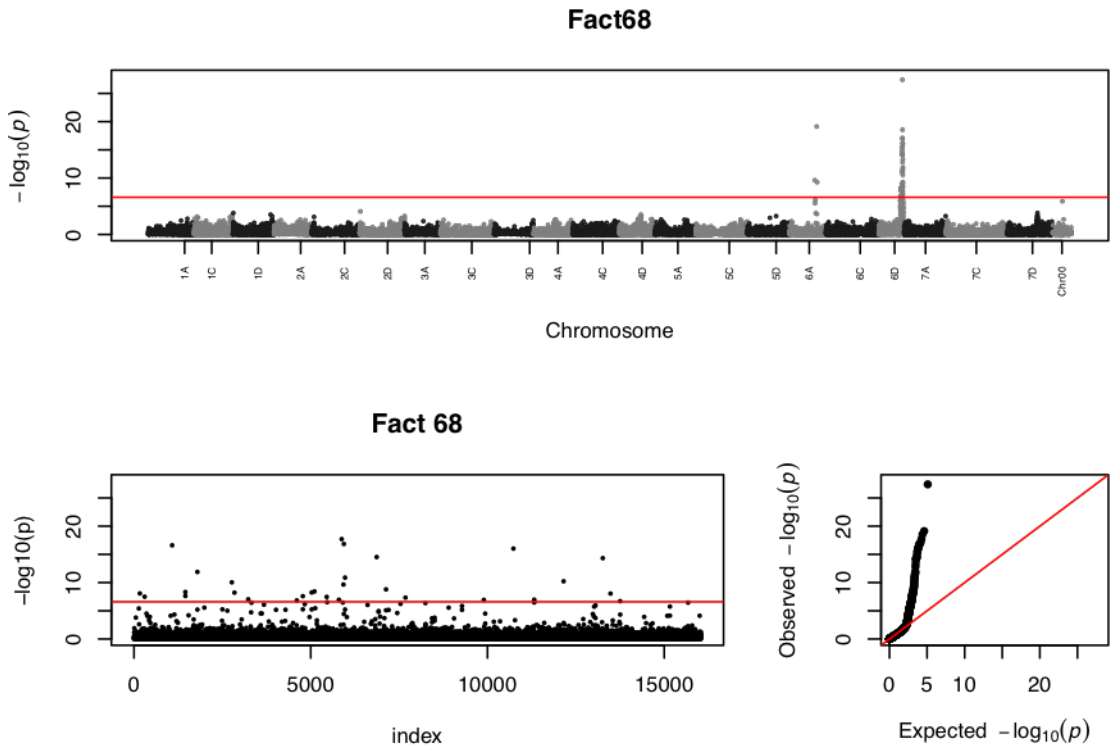

Figure 17. GWAS using scores for factor 68. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

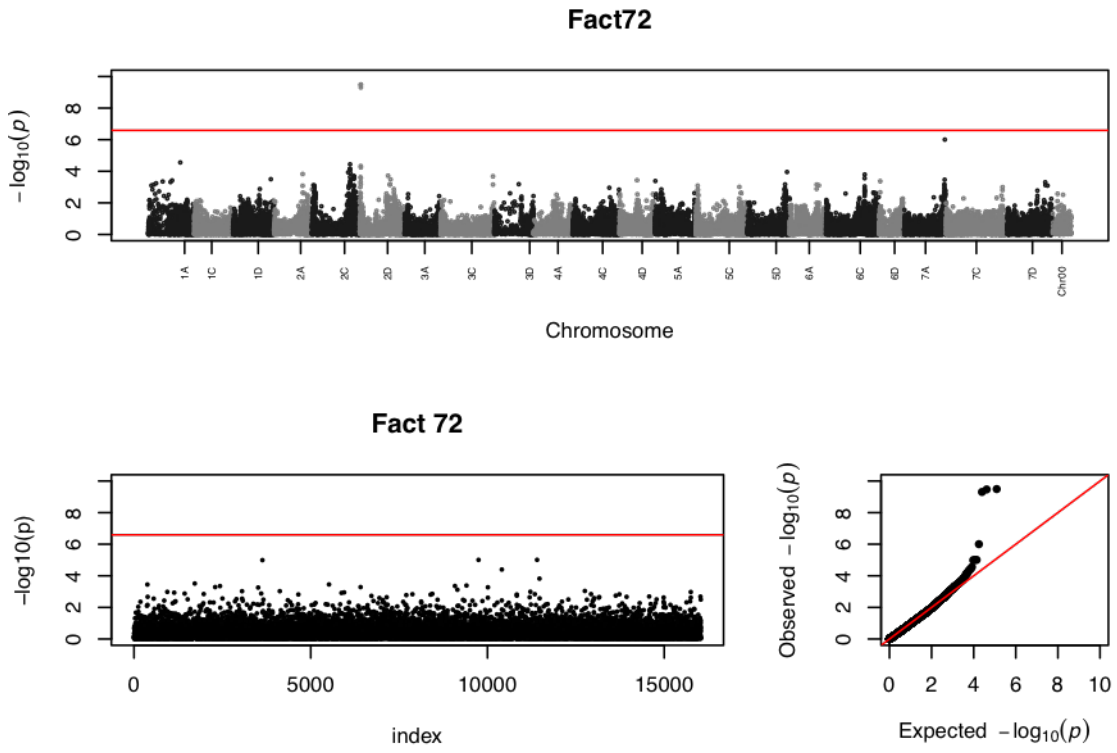

Figure 18. GWAS using scores for factor 72. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

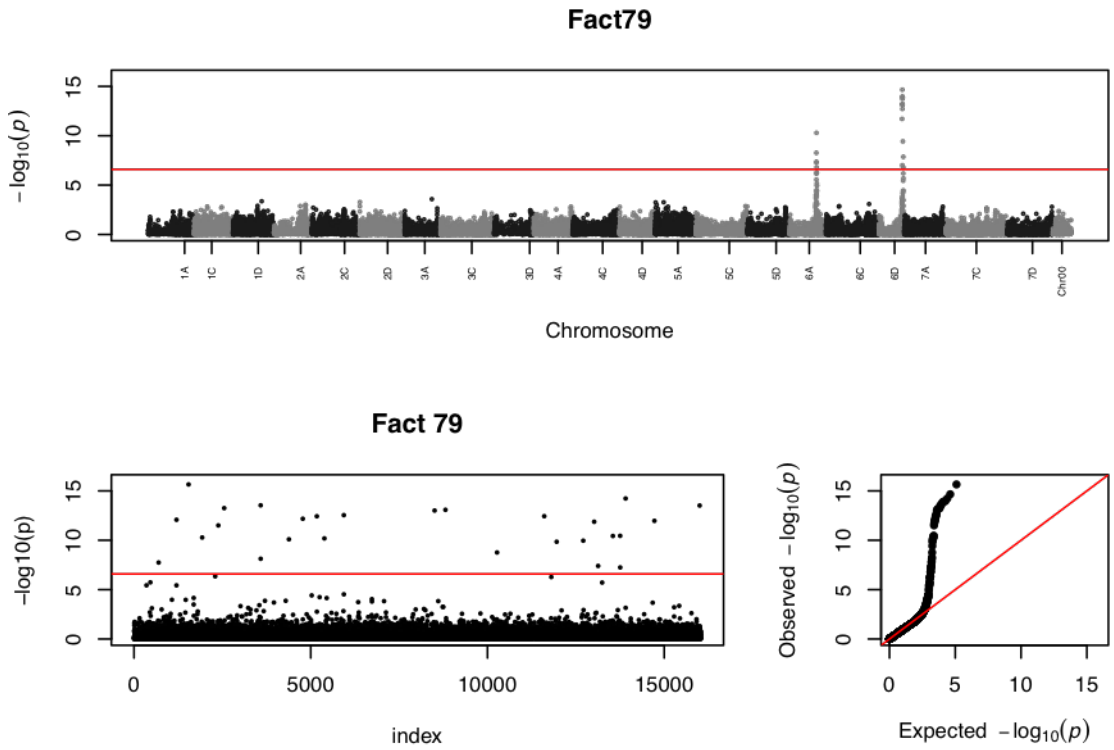

Figure 19. GWAS using scores for factor 79. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

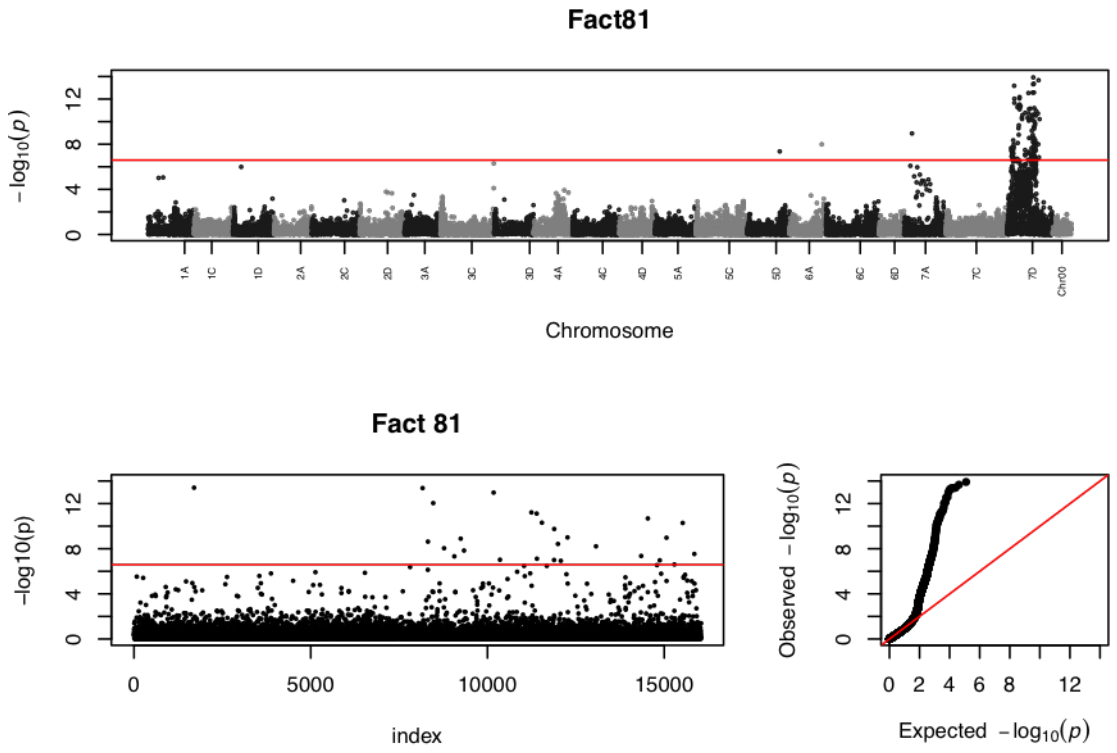

Figure 20. GWAS using scores for factor 81. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

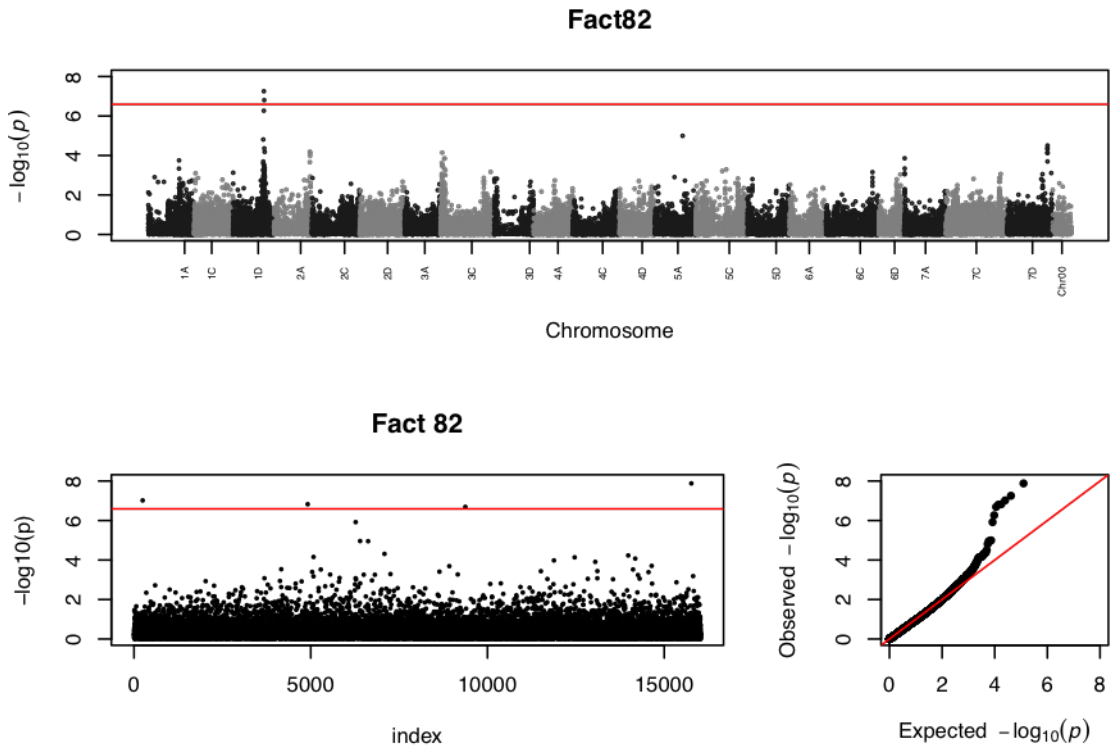

Figure 21. GWAS using scores for factor 82. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

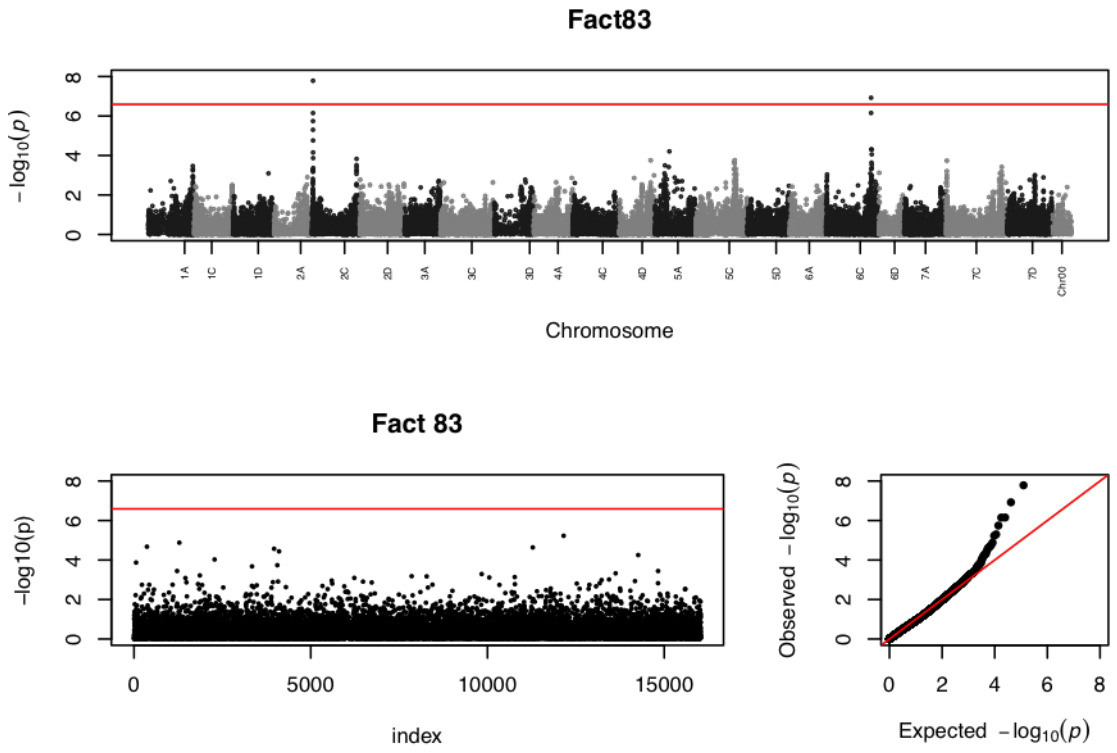

Figure 22. GWAS using scores for factor 83. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

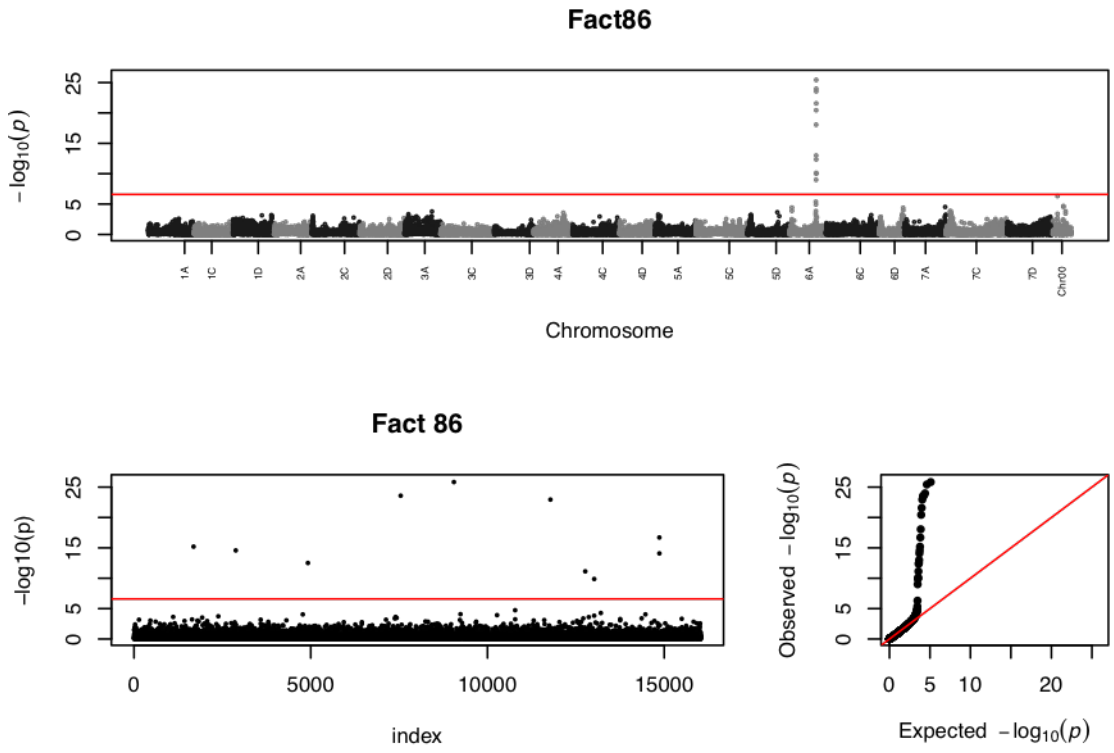

Figure 23. GWAS using scores for factor 86. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

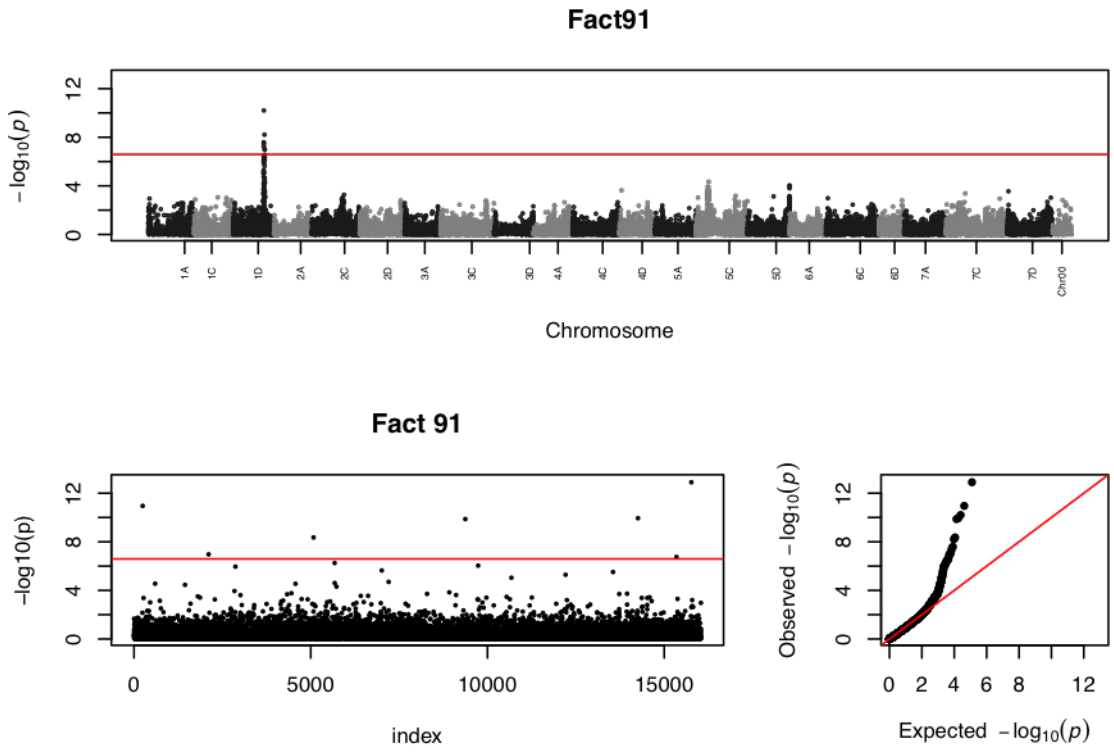

Figure 24. GWAS using scores for factor 91. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

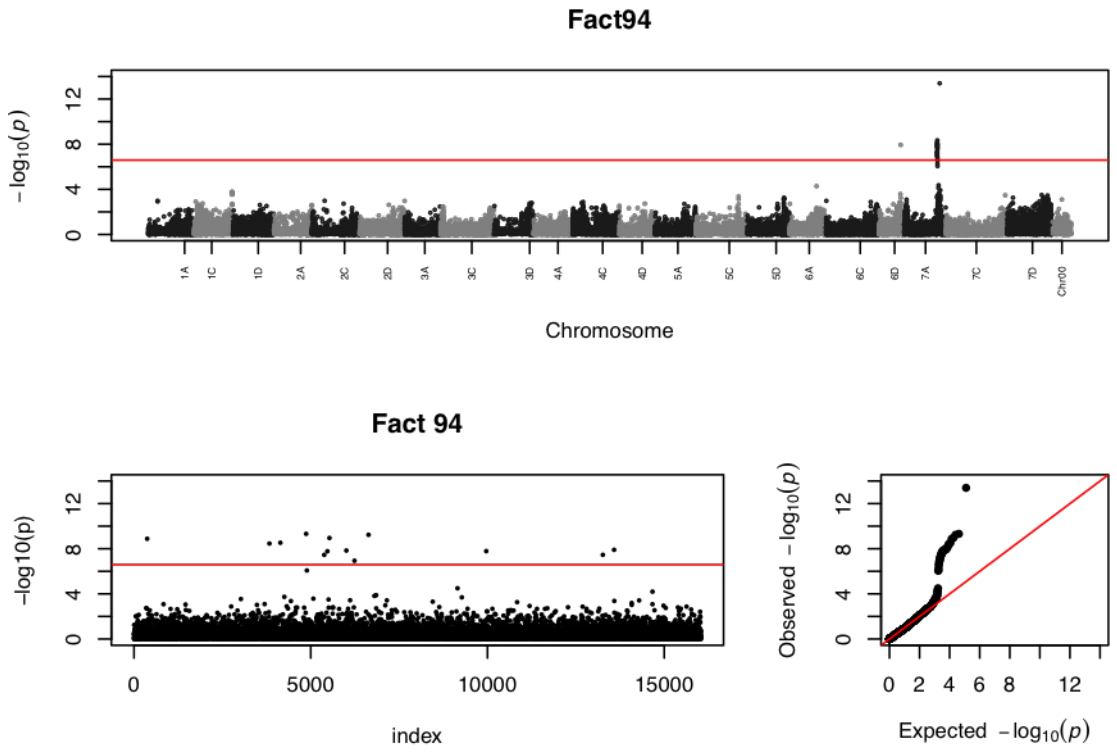

Figure 25. GWAS using scores for factor 94. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

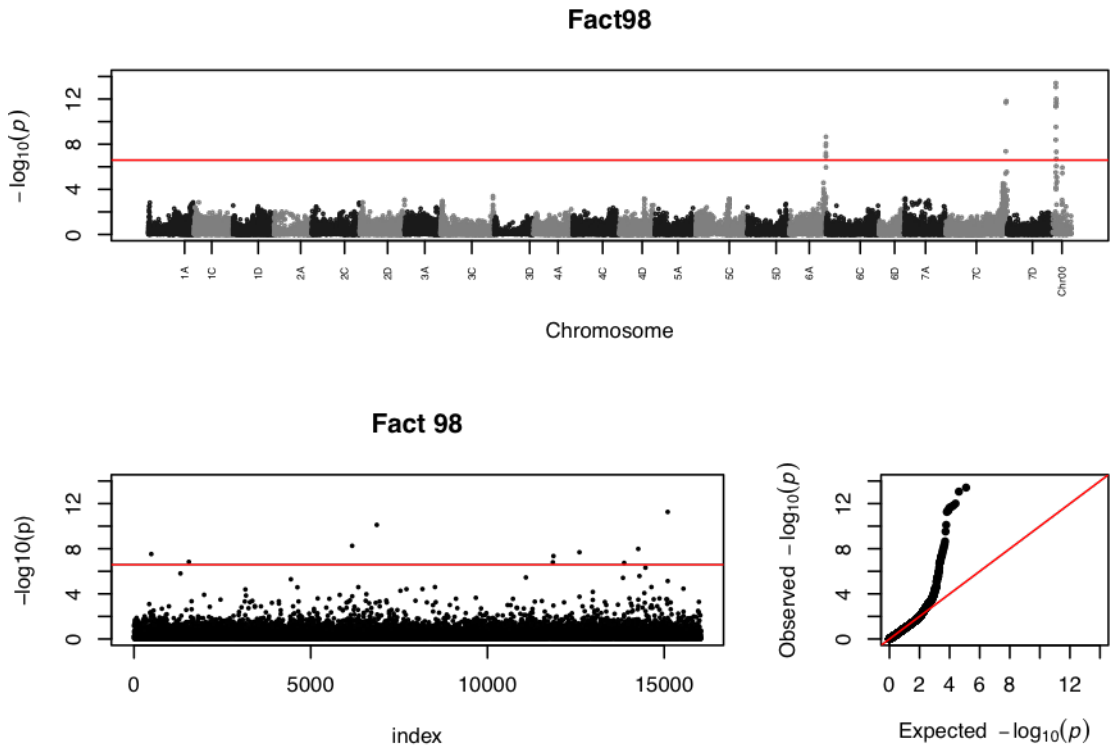

Figure 26. GWAS using scores for factor 98. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

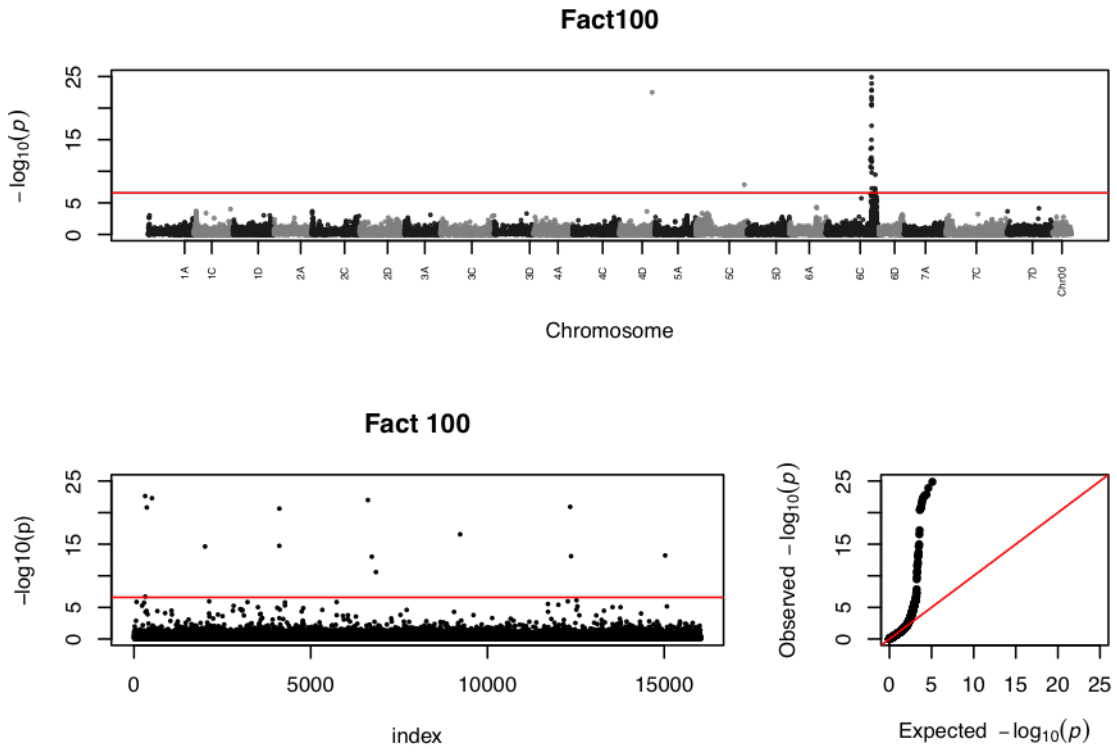

Figure 27. GWAS using scores for factor 100. The top panel shows $-log_{10}(p-values)$ for markers that could be anchored to the genetic map, while the lower left panel shows associations for unanchored markers. The red horizontal line indicates the p-value threshold used to identify significant associations ($p < 2.57 \times 10^{-7}$). Physical positions are based on the Avena sativa – OT3098 v1, PepsiCo genome assembly (https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico).

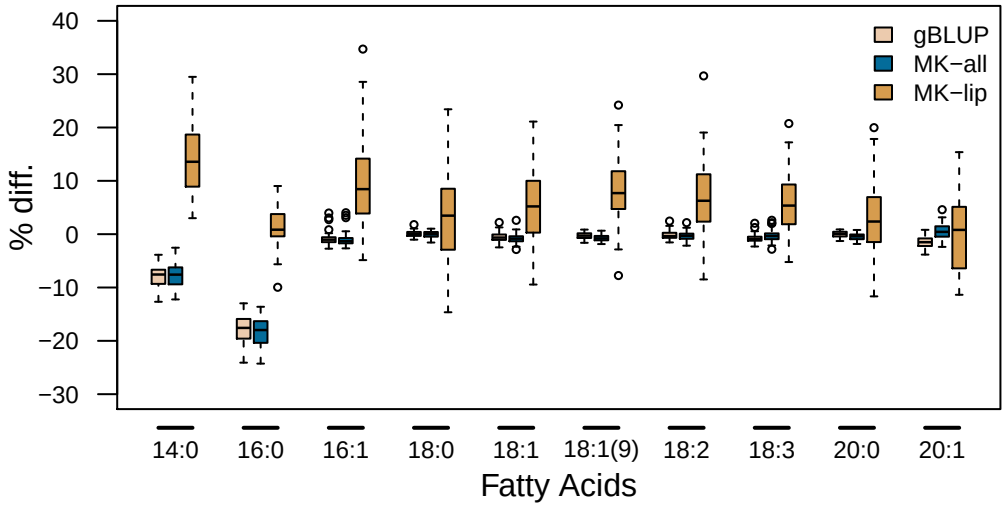

Figure 28. Percent change in prediction accuracy for ten fatty acid phenotypes using BayesB as a reference. Prediction accuracy (r) was assessed using five-fold cross validation with 50 resampling runs. The suffixes '-all' and '-lip' indicate models where the biologically-informed kernel was constructed from markers associated with any latent factor or lipid-enriched factors, respectively. Twelve lines used in this study were also included in the untargeted metabolomics study.

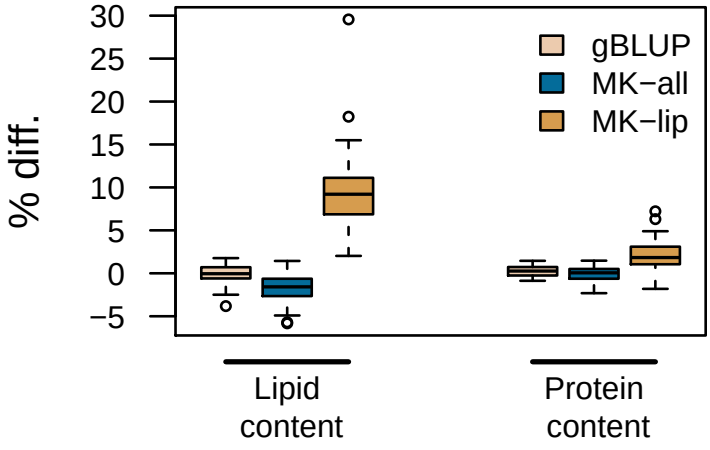

Figure 29. Percent change in prediction accuracy for NIRS seed compositional traits using BayesB as a reference. Prediction accuracy (r) was assessed using five-fold cross validation with 50 resampling runs. The suffixes '-all' and '-lip' indicate models where the biologically-informed kernel was constructed from markers associated with any latent factor or lipid-enriched factors, respectively. Twelve lines used in this study were also included in the untargeted metabolomics study.

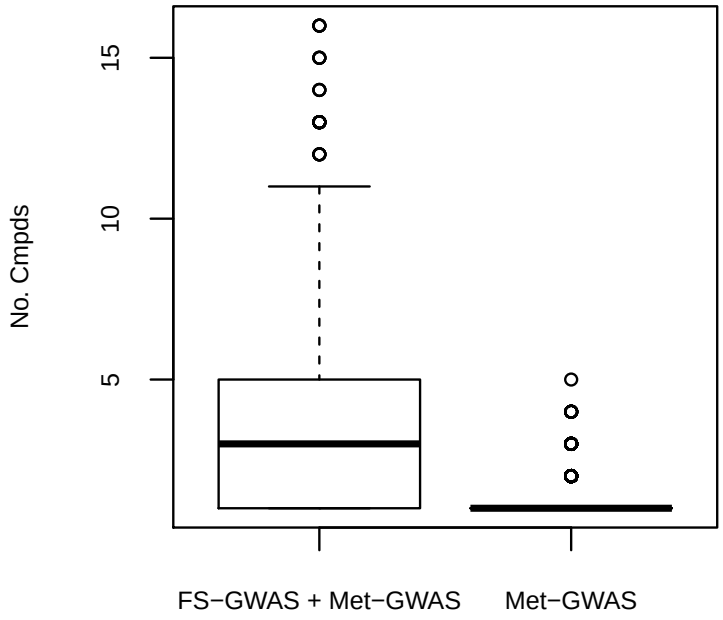

Figure 30. Comparing the number of compounds associated with significant GWAS hits detect using both factor score (FS-GWAS) and metabolite-based GWAS (Met-GWAS), and significant GWAS hits that were uniquely identified using metabolite-based GWAS. The 1,191 markers recovered from metabolite-based GWAS were partitioned into into two groups: those that were identified with both GWAS on factor scores and GWAS on metabolites (FS-GWAS + Met-GWAS), and markers that were only recovered by GWAS on metabolites.