Hematological and gene co-expression network analyses of high-risk beef cattle defines immunological mechanisms and biological complexes involved in bovine respiratory disease and weight gain

[1]Matthew A. Scott, [2]Amelia R. Woolums, [3]Cyprianna E. Swiderski, [1]Abigail Finley, [4]Andy D. Perkins, [5]Bindu Nanduri, [6]Brandi B. Karisch

[1]Veterinary Education, Research, and Outreach Center, Texas A&M University and West Texas A&M University, Canyon, TX, USA, 79015; matthewscott@tamu.edu, afinley@cvm.tamu.edu

[2]Department of Pathobiology and Population Medicine, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA, 39762; amelia.woolums@msstate.edu

[3]School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, Arizona, USA, 85721; swiderski@arizona.edu

[4]Department of Computer Science and Engineering, Mississippi State University, Mississippi State, MS, USA, 39762; perkins@cse.msstate.edu

[5]Department of Comparative Biomedical Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA, 39762; bnanduri@cvm.msstate.edu

[6]Department of Animal and Dairy Sciences, Mississippi State University, Mississippi State, MS, USA, 39762; brandi.karisch@msstate.edu

Correspondence to Matthew A. Scott

matthewscott@tamu.edu

Abstract

Bovine respiratory disease (BRD), the leading disease complex in beef cattle production systems, remains highly elusive regarding diagnostics and disease prediction. Previous research has employed cellular and molecular techniques to describe hematological and gene expression variation that coincides with BRD development. Here, we utilized weighted gene co-expression network analysis (WGCNA) to leverage total gene expression patterns from cattle at arrival and generate hematological and clinical trait associations to describe mechanisms that may predict BRD development.

Gene expression counts of previously published RNA-Seq data from 23 cattle (2017; n=11 Healthy, n=12 BRD) were used to construct gene co-expression modules and correlation patterns with complete blood count (CBC) and clinical datasets. Modules were further evaluated for cross-populational preservation of expression with RNA-Seq data from 24 cattle in an independent population (2019; n=12 Healthy, n=12 BRD). Genes within well-preserved modules were subject to functional enrichment analysis for significant Gene Ontology terms and pathways. Genes which possessed high module membership and association with BRD development, regardless of module preservation ("hub genes"), were utilized for protein-protein physical interaction network and clustering analyses.

Five well-preserved modules of co-expressed genes were identified. One module ("steelblue"), involved in alpha-beta T-cell complexes and Th2-type immunity, possessed significant correlation with increased erythrocytes, platelets, and BRD development. One module ("purple"), involved in mitochondrial metabolism and rRNA maturation, possessed significant correlation with increased eosinophils, fecal egg count per gram, and weight gain over time. Fifty-two interacting hub genes, stratified into 11 clusters, may possess transient function involved in BRD development not previously described in literature. This study identifies co-expressed genes and coordinated mechanisms associated with BRD, which necessitates further investigation in BRD-prediction research.

Author Summary

Bovine respiratory disease (BRD), the leading disease in beef cattle, is a highly dynamic disease complex. Through simultaneous sequencing of thousands of genes active in the blood of cattle at arrival, we pursued the co-expression patterns of these genes to evaluate associations with BRD development and severity overtime. This approach allows for a better understanding of gene expression active in cattle at arrival, and the discovery of new molecules and biological complexes that may predict BRD before the onset of clinical signs. Our work provides evidence that genes related to T-cells, a type of immune cell, are strongly co-expressed when cattle arrive to beef production system, and correlate with increased red blood cell (RBC) factors and BRD development. Further analysis shows that genes involved in cellular energy production and the respiratory electron transport are strongly co-expressed when cattle arrive to beef production system, and correlate with increased eosinophils, a type of immune cell, and weight gain overtime. Additionally, using genes which strongly correlate with BRD development and severity overtime, we identify a novel protein interaction complex that may drive future research for discovering new ways to manage and treat BRD in beef cattle.

Introduction

Despite decades of research involved in discovering novel management tools, developing interventional systems, and advancing antimicrobial therapeutics, bovine respiratory disease (BRD) remains the leading cause of morbidity and mortality in beef cattle operations across North America [1,2,3]. Due to its widespread prevalence, BRD is considered one of the most economically devastating components of beef cattle production systems [2,3,4]. BRD is a polymicrobial, multifactorial disease complex, incorporating infectious agents, host immunity, and environmental elements as predisposing factors [5,6,7]. Previous research over the past several decades has greatly detailed these factors and risks associated with BRD, yet there is minimal evidence that overall rates of disease have improved [5,8,9,10]. Furthermore, diagnostic evaluation of BRD often relies on visual signs attributed to the disease complex, which are commonly non-specific to airway and lung disease, and lack clinical sensitivity [11,12]. Therefore, data driven approaches which capture the biological intricacies associated with clinical BRD development and provide candidate molecular targets capable of stratifying or

83 predicting risk of disease and/or production loss would offer a more precise method of managing
84 BRD.

85 Clinical BRD progression and severity often presents as an acute inflammatory disease
86 [13]. However, molecular and cellular changes precede physiological changes in terms of disease
87 development. As such, identifying consistent molecular and/or cellular components that relate to
88 BRD development would allow for the development of rapid diagnostics capable of being
89 performed with cattle at the time of facility arrival. Such a tool could facilitate precision
90 medicine practices in stocker and feedlot operations and improve both speed and success of
91 targeted therapy. Accordingly, hematological samples are ideal, as they represent a relatively
92 noninvasive, cost effective, and readily obtainable source that reflects dynamic biological
93 processes throughout the body [14,15].

94 Previous research has investigated cellular and molecular components that may indicate
95 or predict clinical BRD. Richeson and colleagues, utilizing complete blood count (CBC)
96 variables and castration status at facility arrival, identified significant associations with BRD in
97 calves with comparatively decreased numbers of eosinophils and increased numbers of
98 erythrocytes [16]. When evaluating the relationships between cytokine gene expression and CBC
99 data in cattle with concurrent BRD, Lindholm-Perry and colleagues discovered that cattle with
100 BRD possessed a comparative increase in numbers of neutrophils, decrease in numbers of
101 basophils, and increased expression of *CCL16*, *CXCR1*, and *CCR1* [17]. Recent RNA sequencing
102 studies, performed by both our group and others, have identified mechanisms and candidate
103 biomarkers in whole blood associated with BRD development [18,19,20]. However, these
104 studies primarily sought to identify differentially expressed genes (DEGs) between cattle that
105 were or were not treated for BRD based on clinical signs. Focus on identifying DEGs meant that
106 much of the data generated by these studies was neglected. Therefore, we aimed to leverage
107 global gene expression patterns across high-risk cattle, and incorporate available cellular-level
108 hematological data from the same cattle, to infer mechanisms associated with BRD development
109 with a more holistic approach.

110 As gene expression operates in tandem with biological regulatory networks and
111 complexes, investigation of gene co-expression levels may reveal transcriptional coordination,
112 distinguish protein production relationships, and measure cellular composition and function

relevant to specific disease states such as BRD [21,22]. This analysis approach falls into the field of systems biology, where, in contrast to reductionist biology, molecular components are pieced and scaled together to better understand disease and generate novel hypotheses [23,24]. In this respect, we sought to build networks of co-expressed genes, utilizing the full structure of previously published gene expression data [20], and discover relationships between gene expression and cellular hematological components, which may elucidate and/or further confirm genes and mechanisms related to BRD development or resistance.

Materials and Methods

Animal enrollment

All animal use and procedures were approved by the Mississippi State University Animal Care and Use Committee (IACUC protocol #17-120) and carried out in accordance with relevant IACUC and agency guidelines and regulations. This study was carried out in accordance with Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines (https://arriveguidelines.org). This study was conducted in accompaniment with previous work focused on differential gene expression analysis and candidate biomarker validation [20]; the RNA-Seq data of these animals were previously deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database under accession number GSE161396. Of the 24 cattle from the 2017 population having RNA-Seq data, one individual (ID: 162-2017_S24; GSM4906455) was not incorporated into the network analysis due to missing CBC data. The following clinical data were recorded for each animal: at-arrival fecal egg counts per gram via modified-Wisconsin procedure (FEC-d0), body weight in pounds (WT) at arrival, Day 12, Day 26, and Day 82, average daily weight gain at each time point (ADG), growth rate (slope of weight over days recorded; GR), at-arrival castration status (Sex), at-arrival rectal temperature (Temp-d0), development of clinical BRD within 28 days post-arrival (BRD), number of clinical BRD treatments (Treat_Freq), and timing to first BRD treatment (Risk_Days). Clinical data for these cattle are found in Supplemental Table S1.

Hematology analysis

Approximately 6 mL of whole blood was collected at arrival into K_3-EDTA glass blood tubes (BD Vacutainer; Franklin Lakes, NJ, USA) via jugular venipuncture. Blood samples were stored at 4°C and analyzed the same day of collection with the flow cytometry-based Advia 2120i hematology analyzer (Siemens Healthcare Diagnostics Inc., Tarrytown, NY, USA), testing for the following parameters: white blood cells (WBC; K/μL), erythrocytes (RBC; M/μL), hemoglobin (HGB; g/dL), hematocrit (HCT; %), mean corpuscular volume (MCV; fL), mean corpuscular hemoglobin (MCH; pg), mean corpuscular hemoglobin concentration (MCHC; g/dL), red blood cell distribution width (RDW; %), and platelets (PLT; K/μL). Blood smear staining was performed with a Hematek 3000 Slide Stainer (Siemens Healthcare Diagnostics Inc., Tarrytown, NY, USA) via Wright-Giemsa stain reagents. Stained blood smears were evaluated for leukocyte distribution via a manual 300-count white blood cell differential by trained clinical pathology technical staff at Mississippi State University College of Veterinary Medicine. Neutrophil, eosinophil, basophil, monocyte, and lymphocyte percentages were recorded, with accompanying neutrophil-to-lymphocyte ratios (NL Ratio). Hematology data for these cattle are found in Supplemental Table S2.

RNA-Seq data processing and normalization

The gene-level raw count matrix generated from our previous research was utilized for this study [20]. Briefly, RNA was isolated via Tempus Spin RNA Isolation Kits (Thermo Fisher Scientific; Waltham, MA, USA), following manufacturer's protocol. TruSeq RNA Library Kit v2 (Illumina; San Diego, CA, USA) was utilized for mRNA sequencing library preparation, following manufacturer's protocol. Single-lane, high-throughput RNA sequencing was performed with NovaSeq 6000 S4 reagent kit and flow cell (Illumina). Sequence read files were quality assessed and trimmed with FastQC v0.11.9 [25] and Trimmomatic v0.39 [26], respectively. Reference-guided (*Bos taurus*; ARS-UCD1.2) read mapping, indexing, and gene-level assembly were performed with HISAT2 v2.2.1 [27,28] and StringTie v2.1.2 [29,30], respectively. The python program prepDE.py [31] was utilized for gene-level count matrix construction.

Raw gene counts were imported to R v4.0.4 and processed with the filterByExpr toolkit [32], removing genes with a minimum total count of less than 200 and counts-per-million (CPM) below 1.0 across a minimum of 12 libraries. Libraries were normalized with the trimmed mean of M-values method (TMM) [33,34] and converted into log2-counts per million values (log2CPM). A total of 12,795 genes were identified after count processing and were utilized for weighted network analysis.

Weighted gene co-expression network analysis (WGCNA)

Weighted network analysis was performed with the R package WGCNA v1.70.3 [35]. Clinical and hematology trait data were compiled and aligned to each respective sample library. To remove any outlier sample, canonical Euclidean distance-based network adjacency matrices were estimated and used to identify outliers based on standardized connectivity. Estimated adjacency matrices had network connectivity standardized with the provided equation [36]:

$$Z.k_{\mu} = scale(k)_{\mu} = \frac{k_{\mu} - mean(k)}{\sqrt{var(k)}}.$$

Samples with a standardized connectivity < -5.0 were considered outliers and to be removed from further analysis; no samples were considered outliers in this study (Supplemental Figure S1). An adjacency matrix was constructed from the calculated signed Pearson coefficients between all genes across all samples. We utilized signed networks as they better capture gene expression trends (up- and down-regulation) and classify co-expressed gene modules which improve the ability to identify functional enrichment, when compared to unsigned networks [24,35,36,37]. Soft thresholding was used to calculate the power parameter (β) required to exponentially raise the adjacency matrix, to reach a scale-free topology fitting index (R^2) of >80%; β = 8 was selected for this study. The relationship between each unit β and R^2 is seen in Supplemental Figure S2. Co-expression modules were constructed with the automatic, one-step blockwiseModules function within the WGCNA R package, using the following parameters: power = 8, corType = "pearson," TOMType = "signed," networkType = "signed," maxBlockSize = 12795, minModuleSize = 30, mergeCutHeight = 0.25, and pamRespectsDendro = FALSE; all other parameters were set to default. Constructed co-expression modules were assigned a color by the WGCNA R package, with any gene not assembling into a specific

module placed in the "grey" module. Module-trait associations were identified with Pearson correlation between module eigengene (ME; first principal component of the co-expression matrix [38] and clinical and hematology data). Modules were considered weakly or strongly correlated with each trait having a p-value ≤ 0.10 and $|R| \geq 0.3$ or p-value ≤ 0.05 and $|R| > 0.4$, respectively. Color scaling was performed with the Bioconductor package viridis v0.6.1 [39] to allow ease of visual interpretation for individuals with color blindness.

Cross-population module preservation analysis

Based on our previous work, it can be inferred that host gene expression captured at facility arrival is variable across BRD severity cohorts [20,40,41]. Therefore, we assessed whether the at-arrival co-expression patterns and modules found in this study were well preserved across an RNA-Seq data set from an independent population of cattle. We investigated cross-populational module preservation across the whole blood transcriptomes of cattle previously assessed for differential gene expression (GSE161396; 2019 population (n=24)) with the modulePreservation function found within the WGCNA R package. The gene-level raw count matrix from previous analysis [20] was utilized and processed, filtered, and normalized in identical procedures as the 2017 RNA-Seq data set (see RNA-Seq data processing and normalization section); a total of 12,803 genes were identified in the 2019 data set after count processing and normalization. Permutation testing (n=200 permutations) was conducted to assess the significance of module preservation across the 2017 and 2019 RNA-Seq data sets, utilizing the two composite statistical measurements Zsummary and medianRank scores [36,42]. Briefly, the identified modules within the test network are randomly permuted n times, where, for each permuted index, the mean and standard deviation is calculated for defining the corresponding Z statistic [42,43]. Through the combination of additional preservation statistics (average of Zdensity and Zconnectivity), the calculated Zsummary statistic determines the level of mean connectivity among all genes within a module (i.e., network density) across the two data sets [24,42]. Higher Zsummary values indicate a stronger level of module preservation between data sets but is dependent on the number of genes within the module (i.e., module size) [42]. To further evaluate preservation in a module size-independent manner, medianRank scores are calculated from the mean connectivity and density measurements observed from each module

and assigned a rank score [42]. Lower medianRank values indicate a stronger level of module preservation between data sets. For this study, any module possessing Zsummary ≥ 10 and medianRank ≤ 5 was considered highly preserved.

Functional enrichment analysis of preserved modules

WebGestalt 2019 [44] (WEB-based Gene SeT AnaLysis Toolkit; accessed September 13, 2021) was utilized for over-representation analysis to identify enriched Gene Ontology (GO) biological processes, cellular components, molecular functions, and pathways from genes found in each module considered well preserved. Pathway enrichment analysis was performed with the pathway database Reactome [45]. Human (*Homo sapiens*) gene orthologs and functional databases were utilized for GO term and pathway enrichment analyses. Over-representation analysis parameters within WebGestalt 2019 included between 3 and 3000 genes per category, Benjamini-Hochberg (BH) procedure for multiple hypothesis correction, adjusted p-value (FDR) cutoff of 0.05 for significance, and a total of 10 expected reduced sets of the weighted set cover algorithm for redundancy reduction.

BRD-associated hub gene identification and network analyses

Hub genes are those genes found within a module (eigengenes) that possess high connectivity which may exhibit a greater degree of biological significance in respect to significantly associated clinical traits, when compared to all other eigengenes [38,46,47]. Here, we sought to identify hub genes found from modules which are significantly associated with any of the clinical BRD categories (BRD, Treat_Freq, and Risk_Days). This was performed in the WGCNA R package with two procedures. First, Pearson correlation between gene expression and module eigengenes was calculated, resulting in the level of module membership (k_{ME}) for each gene. Second, the Pearson correlation between individual gene expression level and clinical trait was calculated, resulting in the level of gene significance (GS) for each gene. Any gene possessing k_{ME} and GS values ≥ 0.7 and ≥ 0.3, respectively, were considered hub genes for clinical traits [36]. All BRD-associated hub genes were used for network construction of known and predicted protein-protein interactions with the Search Tool for the Retrieval of Interacting

Genes (STRING) database v11.5 [48], utilizing bovine (*Bos taurus*) annotations. STRING analysis was performed with the physical subnetwork setting, where edges only display protein interactions that have evidence of binding to or forming a physical complex. Any interaction above a combined score (confidence) of 0.200 was incorporated into the complete network prior to network clustering; disconnected nodes were removed from the network. The Markov Cluster (MCL) algorithm was utilized for network clustering due to its superior performance in complex extraction without the need of additional parameter tuning [49]. Hub genes within the interaction network were placed into distinct clusters based on MCL clustering of the distance matrix acquired from the combined interaction scores, using a MCL inflation parameter of 1.4.

Statistical analysis

Clinical and hematology data (described in animal enrollment and hematology analysis) were compared between cattle treated for naturally-acquired clinical BRD within the first 28 days following facility arrival (BRD) and those never being diagnosed nor treated (Healthy). Residual normality was assessed in R v4.0.4 with the Shapiro-Wilk test [50], with an *a priori* level of significance set at 0.10; neutrophil percentage (Neu%), eosinophil percentage (Eos%), basophil percentage (Baso%), lymphocyte percentage (Lymph%), neutrophil-to-lymphocyte ratio (NL ratio), FEC-d0, MCHC, RDW, and Sex were considered non-normally distributed. Differences in normally distributed variables between BRD and Healthy cattle were assessed with the Student's *t*-test. Differences in non-normally distributed variables were assessed with the Welch's *t*-test; differences between the two groups with respect to Sex was assessed with Pearson's chi-square test with Yates' continuity correction. Differences between BRD and Healthy cattle were considered significant having a p-value ≤ 0.05.

Results

Statistical analysis of clinical and hematological parameters

Descriptive statistics for the clinical and hematological data are provided in Table 1. Regarding the hematological parameters, average values of Lymph%, RDW, and PLT were

outside of the internal reference intervals for both BRD and Healthy cattle. In this study, RBC was considered significantly higher at arrival in BRD cattle compared to Healthy cattle; no other parameter was considered significantly different between the two groups. Regarding clinical data, BRD cattle possessed significantly lower weight gain by end of study (ADG-d82; 2.273 lbs/day in BRD and 2.946 lbs/day in Healthy) and lower calculated slopes of weight gain over time (Growth Rate; 2.370 in BRD and 2.995 in Healthy); no other clinical parameter was considered significantly different between the two groups.

Table 1. Statistical analysis of hematological and clinical traits between BRD and Healthy groups.

Variable	Internal Reference	BRD mean (s.d.)	Healthy mean (s.d.)	p-value
Neu%	37.000 – 80.000	35.917 (5.547)	37.213 (9.748)	0.717
Eos%	0.000 – 12.000	3.944 (3.237)	2.635 (1.616)	0.251
Baso%	0.000 – 2.500	0.193 (0.213)	0.151 (0.218)	0.658
Mono%	0.000 – 12.000	8.862 (4.603)	8.363 (4.507)	0.805
Lymph%	10.000 – 50.000	51.083 (4.756)	51.635 (11.928)	0.893
NL Ratio	N/A	0.711 (0.141)	0.859 (0.660)	0.504
WBC (K/µL)	4.000 – 12.000	7.430 (2.722)	7.320 (1.292)	0.913
RBC (M/µL)	5.000 – 9.990	9.605 (0.568)	9.032 (0.676)	0.047
HGB (g/dL)	7.700 – 15.000	13.075 (1.071)	12.491 (0.906)	0.194
HCT (%)	25.000 – 45.000	36.125 (3.269)	35.000 (2.534)	0.391
MCV (fL)	36.000 – 55.000	37.725 (3.843)	38.845 (2.851)	0.460
MCH (pg)	12.000 – 22.000	13.625 (1.112)	13.855 (0.806)	0.597
MCHC (g/dL)	32.000 – 40.000	36.225 (1.190)	35.691 (0.977)	0.272
RDW (%)	11.600 – 14.800	29.258 (2.362)	27.564 (3.023)	0.171
PLT (K/µL)	200.000 – 900.000	1413.083 (506.885)	1149.000 (401.516)	0.203
FEC-d0	N/A	761.250 (768.795)	618.364 (408.492)	0.597

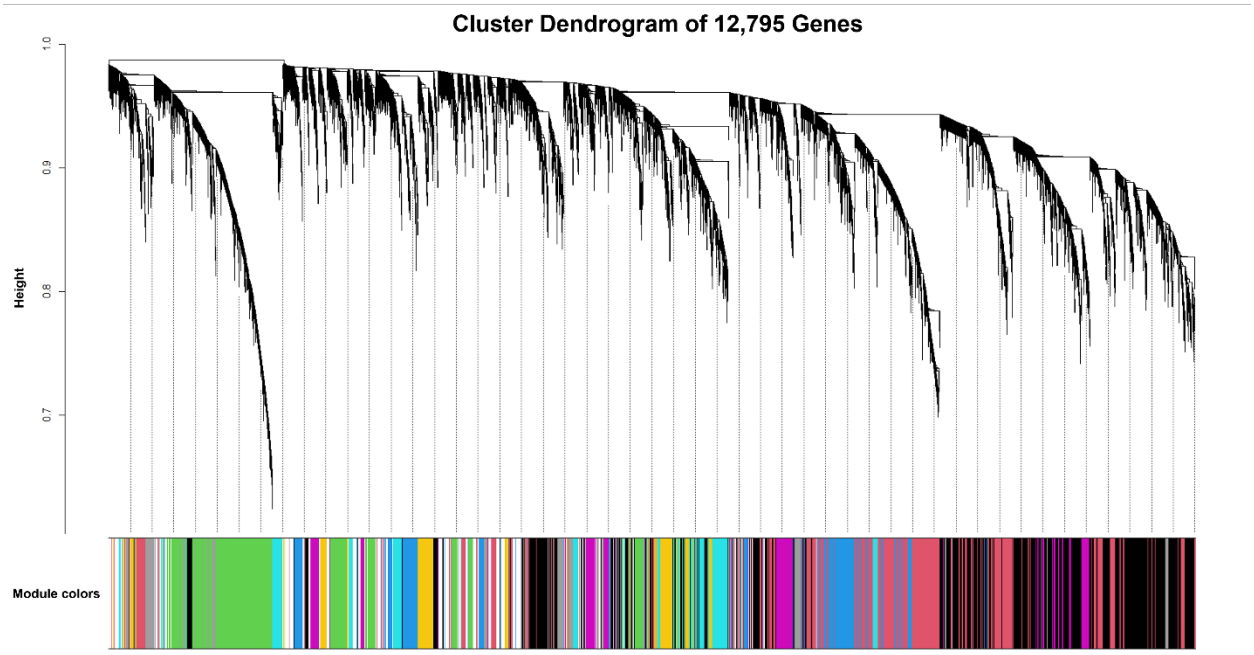
ADG-d12	N/A	0.667 (1.604)	2.167 (1.838)	0.059
ADG-d26	N/A	1.917 (1.204)	2.710 (0.948)	0.110
ADG-d82	N/A	2.273 (0.599)	2.946 (0.432)	0.008
Growth Rate	N/A	2.370 (0.554)	2.995 (0.435)	0.009
Temp-d0 (F°)	N/A	103.333 (0.712)	103.291 (0.667)	0.890
Sex	N/A	10 bulls, 2 steers	10 bulls, 1 steer	1.000

Means, standard deviations (in parentheses), and statistical probability values of differences in hematological and clinical parameters between BRD (n=12) and Healthy (n=11) cattle. Parameters were considered significantly different with p-values ≤ 0.05.

Weighted gene co-expression network construction

The remaining filtered genes (n=12,795) were used for WGCNA network and module construction. The resulting network identified a total of 41 color-coded modules of co-expressed genes, excluding the grey module which incorporates uncorrelated genes (n=1,235) (Figure 1). Across the 41 assigned modules, the turquoise module possessed the largest number of co-expressed genes (n=2,503) and the lightsteelblue1 module possessed the smallest number of co-expressed genes (n=38); the average size of each module was approximately 282 genes. The complete list of genes and module assignment is found in Supplemental Table S3.

Figure 1. Cluster dendrogram of 12,795 genes generated through dissimilarity metrics (1-TOM) and hierarchical clustering.



Automated block-wise module detection of interconnected genes were grouped into 41 unique color-coded modules, excluding the grey module (uncorrelated genes). The x-axis corresponds to the gene-module assignment and the y-axis (Height) depicts the calculated distance between co-expressed genes from hierarchical average linkage clustering.

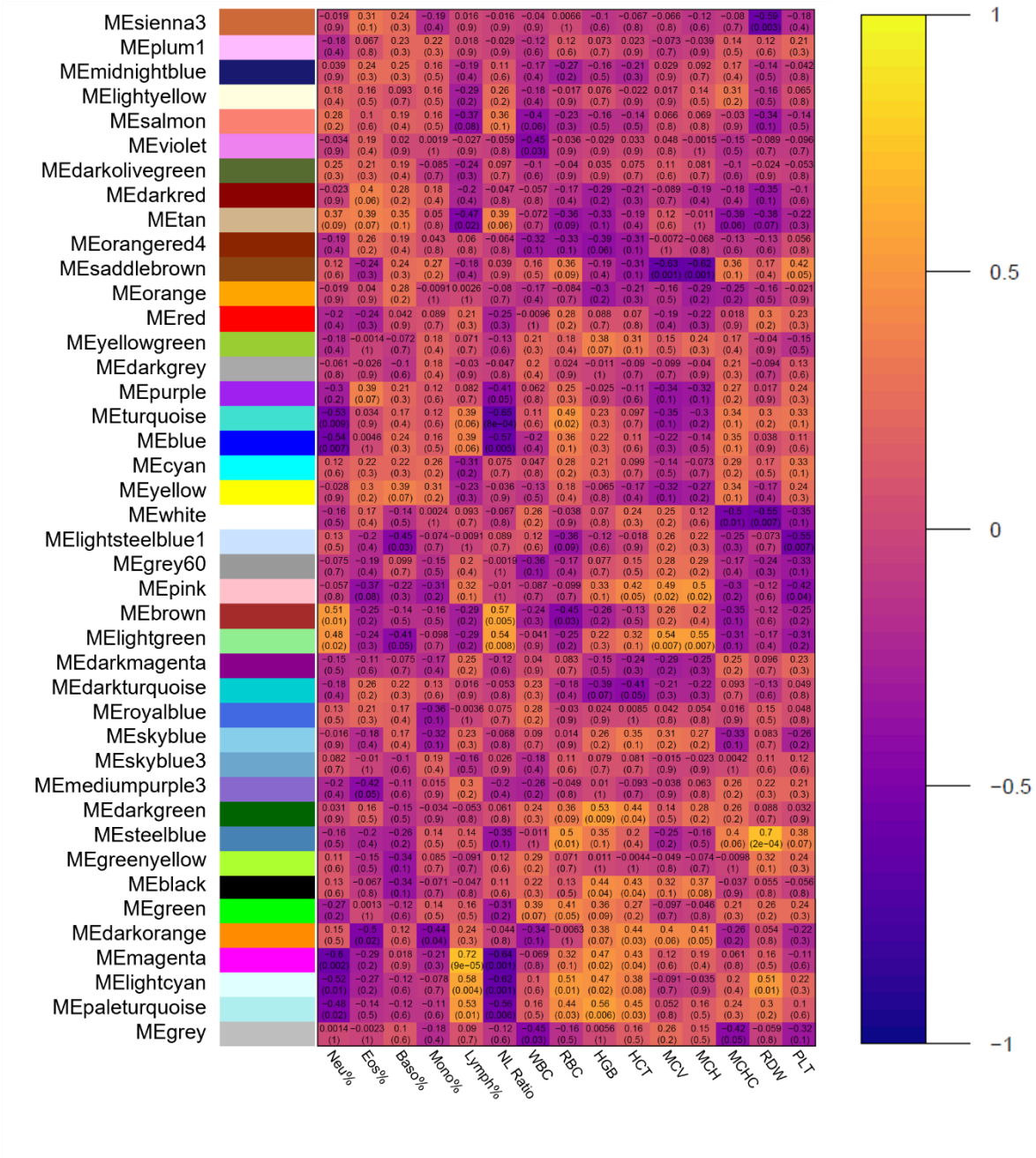
Module-trait relationship with hematological and clinical datasets

Pearson correlation heatmaps were generated to assess the relationship between all modules and hematological clinical datasets. Regarding hematological data, several significant relationships of interest exist (Figure 2). The tan module possessed the highest number of significant correlations with the hematological data (8), followed by turquoise, pink, lightgreen, and lightcyan modules (7). With respect to RBC, considered significantly higher at arrival in BRD cattle compared to Healthy cattle, six modules were strongly correlated: paleturquoise (R = 0.44, p = 0.03), lightcyan (R = 0.51, p = 0.01), green (R = 0.41, p = 0.05), steelblue (R = 0.50, p = 0.01), brown (R = -0.45, p = 0.03), turquoise (R = 0.49, p = 0.02). Additionally, seven modules were considered weakly correlated with RBC: magenta (R = 0.32, p = 0.10), darkgreen (R = 0.36, p = 0.09), lightsteelblue1 (R = -0.36, p = 0.09), blue (R = 0.36, p = 0.10), saddlebrown (R =

329 0.36, p = 0.09), orangered4 (R = -0.33, p = 0.10), tan (R = -0.36, p = 0.09). Regarding modules
330 correlating with RBC, three modules possessed significant associations with multiple related red
331 cell indices (HGB, HCT, MCV, MCH, MCHC, and RDW): saddlebrown, steelblue, and
332 lightcyan. Saddlebrown was strongly associated with MCV (R = -0.63, p =0.001) and MCH (R =
333 -0.62, p =0.001), and weakly associated with HCT (R = -0.31, p = 0.10) and MCHC (R = 0.36, p
334 = 0.10). Steelblue was strongly associated with RDW (R = 0.70, p = 2e-04) and weakly
335 associated with HGB (R = 0.35, p = 0.10) and MCHC (R = 0.40, p = 0.06). Lightcyan was
336 strongly associated with HGB (R = 0.47, p = 0.02) and RDW (R = 0.51, p = 0.01) and weakly
337 associated with HCT (R = 0.38, p = 0.08).

338

339 **Figure 2. Module-trait relationships between co-expression modules and hematological**
340 **traits.**



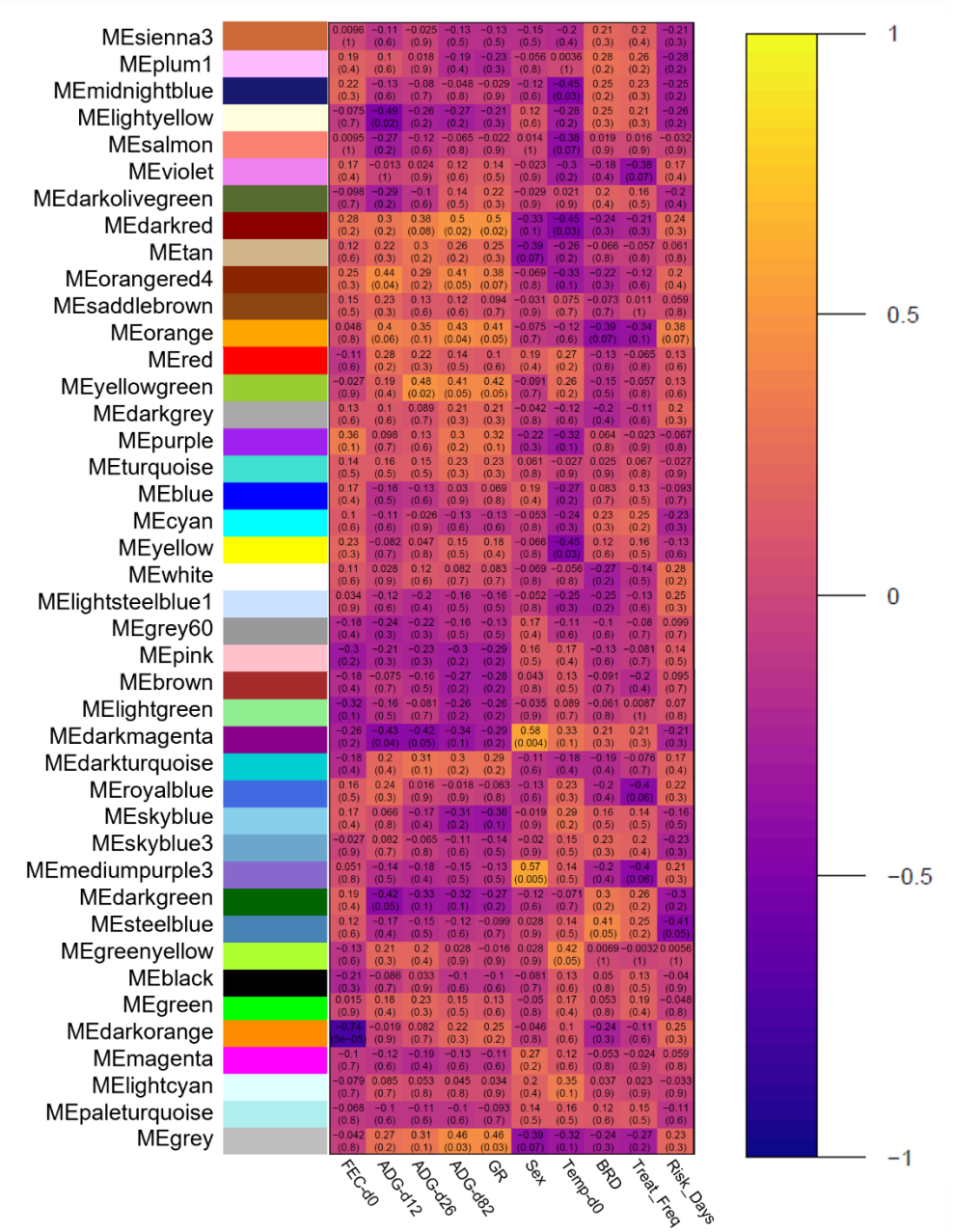
Pearson correlations between each of the unique color-coordinated modules and hematological traits are visualized and represented as a heatmap. Each row represents a distinct co-expression module, and each column represents hematological traits as follows: white blood cells (WBC; K/μL), erythrocytes (RBC; M/μL), hemoglobin (HGB; g/dL), hematocrit (HCT; %), mean corpuscular volume (MCV; fL), mean corpuscular hemoglobin (MCH; pg), mean corpuscular hemoglobin concentration (MCHC; g/dL), red blood cell distribution width (RDW; %), and

348 platelets (PLT; K/μL). Cells are represented by how positive (yellow/white) or negative
349 (purple/black) the correlation is between module and hematological trait, respectively.

350 The relationships between modules and clinical data are found in Figure 3. With respect
351 to all clinical disease associations (BRD, Treat_Freq, and Risk_Days), five modules possessed
352 significant correlations: steelblue, mediumpurple3, royalblue, orange, and violet. Steelblue was
353 strongly associated with BRD (R = 0.41, p = 0.05) and Risk_Days (R = -0.41, p = 0.05).
354 Mediumpurple3 was weakly associated with Treat_Freq (R = -0.40, p = 0.06). Royalblue was
355 weakly associated with Treat_Freq (R = -0.40, p = 0.06). Orange was weakly associated with
356 BRD (R = -0.39, p = 0.07), Treat_Freq (R = -0.34, p = 0.10), and Risk_Days (R = 0.38, p = 0.07).
357 Violet was weakly associated with Treat_Freq (R = -0.38, p = 0.07). Regarding production traits
358 (ADG-d12, ADG-d26, ADG-d82, and GR), ten modules possessed significant correlations:
359 darkgreen, skyblue, darkturquoise, darkmagenta, purple, yellowgreen, orange, orangered4,
360 darkred, and lightyellow. However, to mitigate unexplained variation which may confound
361 differences in ADG-d12 and ADG-d26, coupled with the lack of significance between disease
362 cohorts, eight modules correlating with ADG-d82 and GR were prioritized. Darkred was strongly
363 associated with ADG-d82 (R = 0.50, p = 0.02) and GR (R = 0.50, p = 0.02). Orangered4 was
364 strongly associated with ADG-d82 (R = 0.41, p = 0.05) and weakly associated with GR (R =
365 0.38, p = 0.07). Orange was strongly associated with ADG-d82 (R = 0.43, p = 0.04) and GR (R =
366 0,41, p = 0.05). Yellowgreen was strongly associated with ADG-d82 (R = 0.41, p = 0.05) and
367 GR (R = 0.42, p = 0.05). Purple was weakly associated with GR (R = 0.32, p = 0.10).
368 Darkmagenta was weakly associated with ADG-d82 (R = -0.34, p = 0.10). Skyblue was weakly
369 associated with GR (R = -0.36, p = 0.10). Darkgreen was weakly associated with ADG-d82 (R =
370 -0.32, p = 0.10). Notably, orange was the only module which possessed significant correlations
371 with both disease-associated and weight gain traits. However, orange did not possess any
372 significant correlations with hematological traits.

373

374 **Figure 3. Module-trait relationships between co-expression modules and clinical traits.**



375

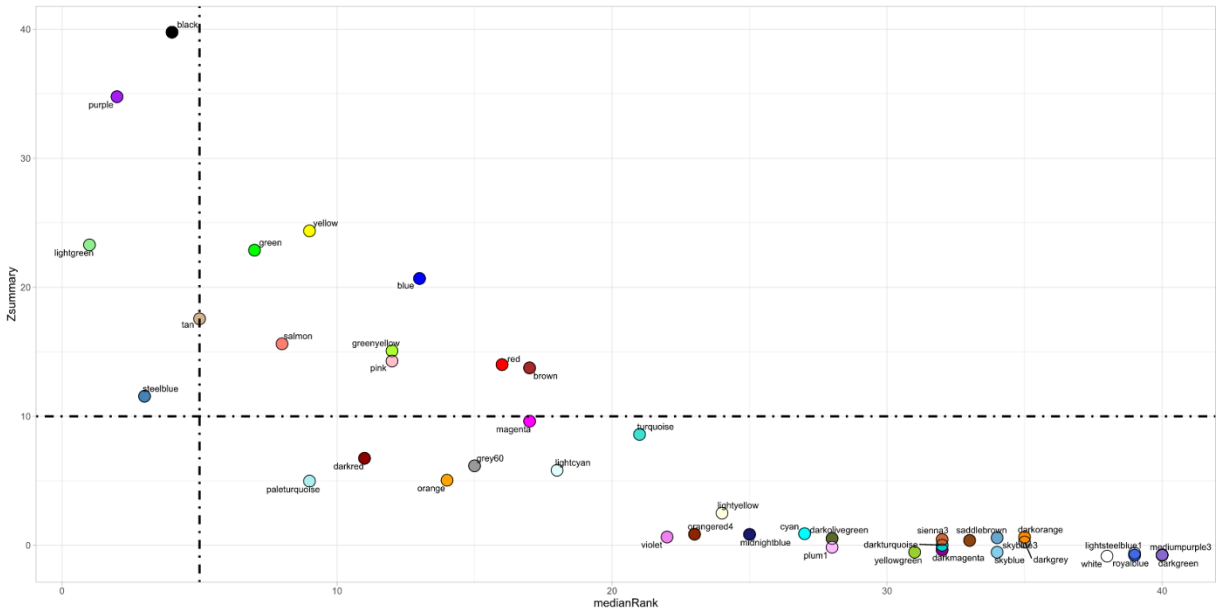
376 Pearson correlations between each of the unique color-coordinated modules and clinical traits are
377 visualized and represented as a heatmap. Each row represents a distinct co-expression module,
378 and each column represents clinical traits as follows: at-arrival fecal egg counts per gram via
379 modified-Wisconsin procedure (FEC-d0), body weight in pounds (WT) at arrival, Day 12, Day
380 26, and Day 82, calculated average daily weight gain at each time point (ADG), growth rate
381 (slope of weight over days recorded; GR), at-arrival castration status (Sex), at-arrival rectal

temperature (Temp-d0), development of clinical BRD within 28 days post-arrival (BRD), number of clinical BRD treatments (Treat_Freq), and timing to first BRD treatment (Risk_Days). Cells are represented by how positive (yellow/white) or negative (purple/black) the correlation is between module and clinical trait, respectively.

Cross-populational network preservation analysis

Module preservation analysis identified five modules considered well preserved across the 2017 and 2019 populations: black (size = 432; Zsummary = 39.772; medianRank = 4), purple (size = 296; Zsummary = 34.773; medianRank = 2), lightgreen (size = 123; Zsummary = 23.291; medianRank = 1), tan (size = 222; Zsummary = 17.559; medianRank = 5), and steelblue (size = 59; Zsummary = 11.555; medianRank = 3) (Figure 4). Notably, steelblue was the only well-preserved module which possessed significant association with BRD-related clinical traits.

Figure 4. Cross-populational module preservation analysis.



The medianRank and Zsummary values across all modules are depicted through the scatterplot x- and y-axes, respectively. Zsummary values $\geq$ 10.0 and medianRank values $\leq$ 5.0, indicated by dashed lines, denote that a module identified with the 2017 gene expression data is well preserved across the 2019 gene expression data.

400

401 Functional enrichment analysis of well-preserved modules

402 To explore the functionality and biological relevance of the five well preserved modules,
403 we performed over-representation analysis with all genes from each module (black, purple,
404 lightgreen, tan, and steelblue; Supplemental Table S4). Analysis of genes from the black module
405 revealed 47 biological process terms, 49 cellular component terms, 17 molecular function terms,
406 and five significantly enriched pathways. Biological processes identified from genes within the
407 black module were related to neutrophil activity and degranulation, aldehyde metabolism,
408 nitrogen compound response and catabolism, and cellular transport. Cellular components
409 identified from genes within the black module involved intracellular and extracellular vesicles,
410 secretory granules, cellular junctions, and lysosomes. Molecular functions identified from genes
411 within the black module involve cytokine, enzyme, and calcium-dependent protein binding,
412 aldehyde dehydrogenase (NAD) activity, and interleukin-1 receptor activity. Enriched pathways
413 identified from genes within the black module involved neutrophil degranulation, metabolic
414 disease, and signaling via tyrosine kinase receptor.

415 Analysis of genes from the purple module revealed 54 biological process terms, 46
416 cellular component terms, 16 molecular function terms, and 40 significantly enriched pathways.
417 Biological processes identified from genes within the purple module involved mitochondrial
418 processes (cristae formation, respiratory chain complex assembly), non-coding RNA processing
419 and maturation, cellular protein transport, and metabolic processes and biosynthesis. Cellular
420 components identified from genes within the purple module involved cell substrate and adhesion
421 junction, ribosomes, cytoplasmic side of endoplasmic reticulum, mitochondrial inner membrane
422 and envelope, and the 48S preinitiation complex. Molecular functions identified from genes
423 within the purple module involved mRNA/rRNA binding, ubiquitin ligase inhibition, ATP
424 synthase activity, and NADH dehydrogenase. Enriched pathways identified from genes within
425 the purple module involved infectious disease/viral infection, amino acid metabolism, translation
426 initiation/termination, rRNA processing, and ATP synthesis and respiratory electron transport.

427 Analysis of genes from the lightgreen module revealed 38 biological process terms, 49
428 cellular component terms, three molecular function terms, and one significantly enriched
429 pathway. Biological processes identified from genes within the lightgreen module involved

leukocyte/neutrophil differentiation, activation, and degranulation, tissue remodeling, cell secretion and exocytosis, phagocytosis and micropinocytosis, dendritic cell activation, and interleukin-8 secretion. Cellular components identified from genes within the lightgreen module involved lysosome, secretory/azurophil granule, vesicular/vacuolar membrane, granule lumen, and macropinosome. Molecular functions identified from genes within the lightgreen module involved symporter activity, potassium-chloride symporter activity, and phosphatidylinositol binding. The single enriched pathway identified from genes within the lightgreen module was neutrophil degranulation.

Analysis of genes from the tan module revealed 35 biological process terms, 32 cellular component terms, four molecular function terms, and two significantly enriched pathways. Biological processes identified from genes within the tan module involved B-cell activation, receptor signaling, and regulation, immunoglobulin production, cytokine production, positive regulation of interferon-gamma production, and mononuclear cell proliferation. Cellular components identified from genes within the tan module involved MHC class II protein complex, lytic vacuole membrane, clathrin-coated endocytic vesicle, endosomal membrane, and B-cell receptor complex. Molecular functions identified from genes within the tan module involved MHC class II receptor activity, MHC class II protein complex binding, and peptide antigen binding. Enriched pathways identified from genes within the tan module were antigen activates B-cell receptor (BCR) leading to generation of second messengers and CD22-mediated BCR regulation.

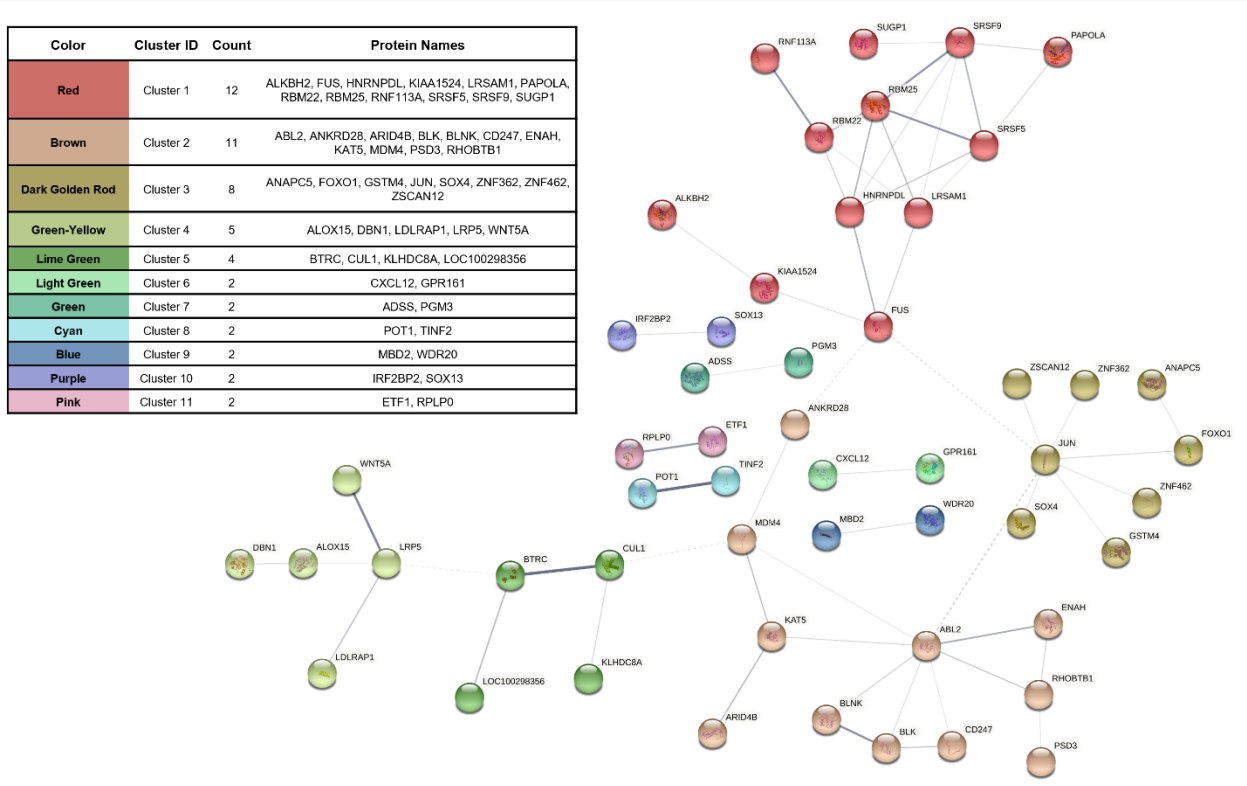
Analysis of genes from the steelblue module revealed three biological process terms, three cellular component terms, no molecular function terms, and no significantly enriched pathways. Biological processes identified from genes within the steelblue module were cell surface receptor signaling pathway, negative regulation of fibroblast growth factor receptor signaling pathway, and antigen receptor-mediated signaling pathway. Cellular components identified from genes within the steelblue module involved side of membrane, plasma membrane part, and alpha-beta T cell receptor complex.

BRD-associated hub gene identification and *in silico* protein-protein interaction and clustering analyses

Hub gene identification analysis included co-expressed genes from the following modules: violet (54), orange (68), royalblue (100), mediumpurple3 (41), and steelblue (59). The k_{ME} and GS value cutoffs within each module resulted in 24, 46, 30, 22, and 32 BRD-associated hub genes from the violet, orange, royalblue, mediumpurple3, and steelblue modules, respectively (Supplemental Table S5). These resulting hub genes were further utilized for physical subnetwork protein-protein interactions and network clustering. After removal of all disconnected nodes, the interaction network demonstrated significant connectivity between 52 proteins across 11 distinct clusters with high inter-nodal connectivity (Figure 5); these gene products and their combined interaction scores are found in Supplemental Table S6. These connected gene products demonstrate possible at-arrival biomolecular complexes associated with BRD development and severity.

Figure 5. Protein-protein interaction network of interconnected BRD-associated hub genes.

Color	Cluster ID	Count	Protein Names
Red	Cluster 1	12	ALKBH2, FUS, HNRNPDL, KIAA1524, LRSAM1, PAPOLA, RBM22, RBM25, RNF113A, SRSF5, SRSF9, SUGP1
Brown	Cluster 2	11	ABL2, ANKRD28, ARID4B, BLK, BLNK, CD247, ENAH, KAT5, MDM4, PSD3, RHOBTB1
Dark Golden Rod	Cluster 3	8	ANAPC5, FOXO1, GSTM4, JUN, SOX4, ZNF362, ZNF462, ZSCAN12
Green-Yellow	Cluster 4	5	ALOX15, DBN1, LDLRAP1, LRP5, WNT5A
Lime Green	Cluster 5	4	BTRC, CUL1, KLHDC8A, LOC100298356
Light Green	Cluster 6	2	CXCL12, GPR161
Green	Cluster 7	2	ADSS, PGM3
Cyan	Cluster 8	2	POT1, TINF2
Blue	Cluster 9	2	MBD2, WDR20
Purple	Cluster 10	2	IRF2BP2, SOX13
Pink	Cluster 11	2	ETF1, RPLP0



Interaction score analysis reveals 52 genes, with high intramodular and BRD-trait relationship, which possess high connectivity. Interconnected gene products (nodes) were further grouped into distinct clusters based on their interaction scores (edges). Edge thickness represents the level of interaction confidence between nodes.

Discussion

While at-arrival management practices are somewhat dependent upon anticipated risk of BRD development, both inter- and intra-herd level disease prevalence is highly variable [5,51]. To counter this variability, beef production systems will often administer antimicrobials and/or immunostimulants at arrival to reduce the risk of clinical BRD development and associated production losses [52,53]. However, immunostimulant administration alone as a metaphylactic protocol for controlling BRD appears to have minimal impact on rates of morbidity [54,55,56]. Metaphylactic use of antimicrobials at arrival reduces risk of morbidity and mortality across beef production systems, however this management practice is variable in efficacy, in both rates of disease across cattle populations and in pharmacological choice, and the practice is suspected to drive expansion of antimicrobial resistance, a growing societal concern [52,57,58]. Given this background, our research group and others have focused on evaluating host transcriptomes at arrival, to better characterize host-driven mechanisms and develop candidate mRNA biomarkers associated with clinical BRD outcomes [18,19,20]. These studies have provided valuable information regarding cattle treated based on clinical signs of BRD, but these studies heavily rely on semi-objective evaluation of BRD cases and may miss underlying subclinical or misdiagnosed disease. As such, the underlying host mechanisms involved in BRD development remain disputed. Therefore, to identify at facility arrival genes and mechanisms which may represent the variable development of BRD cases and leverage the total expression profile of individual cattle, we employed a systems biology approach with weighted co-expression network analysis. This methodology allows us to identify networks of genes exclusively co-expressed, and to evaluate said networks in a reduced manner in order to identify molecules and mechanisms of interest for future BRD prediction studies. Importantly, co-expression network analysis serves as a complementary, yet distinct, approach to identifying genes and mechanisms associated with disease status, when compared to differential expression analyses. The network

approach performed in this study evaluates and identifies genes that are strongly coordinated in terms of expression, and determines correlation with overlapping metadata (clinical data), whereas differential expression analyses typically follow a pairwise approach to determine level of effect and probability of gene differences between groups. Co-expression network analyses consider greater biological context when evaluating gene expression differences, compared to more traditional pairwise approaches. Additionally, through utilization of hematological parameters, we could capture changes in the cellular composition of whole blood as they may relate to cellular and gross pathophysiology across individuals.

While we recognize that dynamic changes captured in whole blood may not completely encompass biomolecular characteristics seen within lung tissue, whole blood serves as a practical and easily obtainable sample for respiratory and inflammatory disease diagnostics [59,60]. After initial statistical assessment of CBC data, we identified that both BRD and Healthy cattle possessed comparable lymphocytosis, thrombocytosis, and erythrocytic macrocytosis; the distribution of these values were not considered significantly different between the two groups. Notably, mild to moderate levels of dehydration, a common condition in newly arrived post-weaned beef animals, may cause elevated changes in hematocrit levels and lymphocytes [61,62]. Lymphocytosis and thrombocytosis may also result from host responses to infection or inflammation. Additionally, reticulocytosis (i.e., immature erythrocytosis) is the most common cause of erythrocytic macrocytosis [61] and was noted as a common feature found across all blood samples submitted for analysis. While these cattle did not possess physiological nor hematological evidence of hemolysis or blood loss upon facility arrival, this finding may be associated with early regenerative anemia, systemic inflammation, or mineral deficiencies [61,62,63]. Furthermore, blood-borne pathogens were not reported from blood smear assessment. Nevertheless, it does not rule out the possibility of mild/subclinical intraerythrocytic pathology or asymptomatic convalescence that may result in these increased hematological changes. Such pathology is often caused by parasitic diseases such as anaplasmosis, a common infectious disease of cattle across the United States [64,65]. It is plausible that these findings indicate that cattle at facility entry are undergoing similar physiological changes as it relates to stressful and/or pathogenic events (long-distance transportation, co-mingling, etc.) and underlying genomic mechanisms serve to resolve or prolong deleterious physiological conditions that result in BRD.

With respect to distributions, we identified that RBCs were significantly increased in cattle that would go on to develop BRD versus those that did not. Although this result was identified in a relatively small number of cattle, it corresponds with the work of Richeson and colleagues [16]. As discussed within their prior research, elevated RBCs may indicate dehydration and subsequent predisposition with BRD development [5,16]. Interestingly, we were able to identify one well-preserved co-expression module which possessed significant correlations with RBCs, RDW, PLT, BRD, and Risk_Days (i.e., shorter time to first treatment): steelblue. Upon further investigation, we discovered that the genes within this module were related to antigen receptor-mediated signaling (*BLK*, *CD247*, *CD276*, *CD3G*, *GATA3*, and *PLEKHA1*) and negative regulation of fibroblast growth factor receptor signaling (*CREB3L1*, *GATA3*, and *WNT5A*), and specifically components of alpha-beta T-cell receptor complexes (*CD247* and *CD3G*). The upregulation of IL-7R and associated signaling molecules, which include *CD3G* and *CD247*, initiate NOTCH-dependent proliferation of T-cell precursors [66]. Furthermore, elevated levels of *BLK* and *GATA3* tend to skew the immune response towards Th2-type immunity [67,68,69]. In terms of RBC relationship, previous research has demonstrated that Th2-stimulated bone marrow T-cells promote erythroid differentiation and lead to the development of erythroblasts [70]. Additionally, *CXCL12*, also identified within the steelblue module and previously identified as a differentially expressed gene associated with BRD development [20], is involved in Th2-cell migration and immune response [70,71]. *HNRNPH3*, found within the steelblue module, has previously been identified as a key transcription factor associated with clinical BRD [18]. Lastly, several genes identified in the steelblue module were also found in the "turquoise" module identified by Hasankhani and colleagues [24], which enriched positive regulation of activated T-cell proliferation and Th1/Th2-cell differentiation pathways. While this study cannot elucidate the exact mechanistic components nor temporality of molecular events, it suggests that promotion of Th2-mediated T-cells at arrival shares a common mechanism with RBC elevation and risk of BRD development. Our previous research has indicated that genes elevated at arrival in cattle that eventually develop BRD interact, and may enhance, TLR-4 and IL-6 responses [20,40,72], which may contribute to the co-expressed pattern related to Th2-mediated T-cell development [73]. Overall, this pattern of Th2-mediated immunity is strongly associated with clinical BRD development and

565 timing to first treatment, and may further strengthen the depiction that early Th2 responses
566 indicate clinical disease development and lung pathology [74,75].

567 While steelblue was the only well-preserved BRD-associated module detected, four other
568 modules were determined to be well-preserved across populations and warranted specific
569 functional enrichment investigation: black, purple, lightgreen, and tan. Genes within the black
570 module, largely involved with neutrophil activation and degranulation, IL-1 activity, and
571 metabolic disease, was only significantly associated with hemoglobin and erythrocyte parameters
572 (HGB, HCT, MCV, and MCH); notably, the black module did not possess any significant
573 associations with clinical variables. This may indicate that neutrophilic and IL-1 activity was not
574 indicative of BRD within this population of cattle, and/or additional disease-associated variables
575 were not recorded in this study. Genes within the purple module, associated with increased
576 eosinophil percentage, decreased neutrophil-lymphocyte ratio, decreased MCV and MCH,
577 increased at-arrival fecal parasitic egg count, and increased growth rate (weight gain over 82
578 days), largely enriched for mitochondrial function and aerobic metabolism and RNA processing.
579 Importantly, this module possessed positive association to weight gain independent of BRD
580 development. Previous research has investigated many of these ribosomal protein-encoding
581 genes for their potential for immune effector capacity [76] and cell regulation [77], however this
582 marks the first time, to our knowledge, that they have been implicated in contributing to weight
583 gain potential in high-risk cattle. Notably, one gene (*RPS26*) has been previously identified as a
584 differentially decreased marker in the diseased lungs of cattle experimentally challenged with
585 BRD-associated pathogens [78,79]. Similar to the black module, genes identified within the
586 lightgreen module were associated with hemoglobin and erythrocyte parameters, but additionally
587 positively correlated with neutrophil percentage and neutrophil-lymphocyte ratio, and negatively
588 correlated with basophil percentage; likewise, the lightgreen module did not possess significant
589 associations with clinical variables. Lastly, the tan module, possessing several significant
590 hematological associations, and was negatively correlated with castration status at arrival,
591 possessed genes which enriched for B-cell receptor complexes and regulation and interferon-
592 gamma production. Unfortunately, the underlying physiological impact of co-expressed genes
593 identified within the black, lightgreen, and tan modules were not captured in this study. As this
594 study was primarily focused on BRD development and severity, the genes within these three

595 modules may possess a role in other disease complexes or immune-mediated events, such as
596 gastrointestinal or apoptotic/necrotic diseases.

597 Utilizing hub gene and interaction network analyses, we further identified genes related
598 to BRD development and severity. Here, we detected and mapped 52 genes into a protein-protein
599 interaction network, further stratified into 11 distinct clusters based on their combined interaction
600 scores. This procedure helps describe the physical relationship that multiple BRD-associated
601 gene products possess with one another in a more holistic approach. Here, we may infer that
602 these interactions possess accompanying transient functions involved in BRD development not
603 previously described in literature. As such, these predicted protein-protein network interactions
604 may infer potential modular units which participate in BRD development or resistance [80,81].
605 Further evidence of the associative importance related to BRD development exists with these
606 genes, as *CXCL12* [20], *TLL2* [20], *ALOX15* [18,20,40], and *LOC100298356* [72,78,79,82] have
607 been previously identified as differentially expressed when comparing cattle with and without
608 BRD development. Proteomic approaches have detailed that proteins infrequently operate as
609 single biological entities and, when involved in similar biological functions, interact in dynamic,
610 yet organized complexes [83,84,85,86]. As such, these findings provide candidate protein
611 complexes related to BRD development and severity, which warrants further investigation for
612 avenues of confirmation in larger populations of cattle and novel therapeutic target development.

613

614 Conclusions

615 This study was conducted to utilize systems biology methodology to further establish
616 genes, mechanisms, and coordinated biological complexes associated with dynamic
617 hematological changes and BRD development. Utilizing our previously published RNA-Seq data
618 and WGCNA, we identified five well-preserved modules of highly co-expressed genes with
619 significant associations with hematological and clinical traits in cattle at facility arrival. The
620 “steelblue” module, containing genes involved in alpha-beta T cell receptor complex and
621 negative regulation of fibroblast growth factor receptor signaling, possessed significant positive
622 correlations with erythrocyte count, platelet count, red cell width, and BRD diagnosis, and
623 negative correlation with days at risk for BRD. The “purple” module, containing genes involved

in mitochondrial processes and non-coding RNA processing and maturation, possessed significant correlation with increased eosinophil percentage, decreased neutrophil-lymphocyte ratio, and increased growth rate (weight gain over time). Protein-protein interaction network and clustering analyses of BRD-related hub genes identified possible at-arrival biological complexes strongly associated with BRD development; many of these hub genes have been described as differentially expressed genes in previous BRD research. Through this holistic molecular approach, we provide genes, mechanisms, and predicted protein complexes associated with BRD development and performance which are warranted for future analyses targeted in predicting BRD at facility arrival.

Acknowledgements

This work was supported by the Animal Health and Disease Program [Grant no. 2020-67016-31469] from the USDA National Institute of Food and Agriculture. The authors thank William Crosby, Kirsten Midkiff, Alexis Thompson, Joseph Gerlach, Merrilee Thoresen, and the supporting staff at the Mississippi State University College of Veterinary Medicine Clinical Pathology Laboratory for their technical assistance and insights throughout this study. The authors would also thank students and staff of the Mississippi Agricultural and Forestry Experiment Station (MAFES) and Mississippi State University Animal and Diary Science Department for their assistance in animal care and sample collection.

Data availability

The data utilized in this study are found in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI-GEO), accession number GSE161396. All data and code used for running experimental analyses are found on a GitHub repository at https://github.com/mscott16/2022-BRD-WGCNA and archived on Zenodo (DOI:

649 10.5072/zenodo.1015612). All remaining relevant data are found within the paper and its
650 additional files.

651

652 Additional Files

653 Supplemental Table S1: **Clinical metadata of cattle selected for WGCNA analysis**

654 Supplemental Table S2: **CBC and leukocyte distribution data of cattle selected for WGCNA**
655 **analysis**

656 Supplemental Table S3: **Full gene list and weighted module assignment**

657 Supplemental Table S4: **Functional enrichment analysis of the five well-preserved modules**

658 Supplemental Table S5: **Hub gene analysis of all five BRD-associated modules**

659 Supplemental Table S6: **STRING identifiers and physical interaction scores**
660 **(combined_score ≥ 0.200)**

661

662 Supplemental Figure S1: **Heatmap and hierarchical clustering of clinical and hematological**
663 **data across the 23 cattle utilized for transcriptome network analysis**

664 Standardized connectivity was calculated from network adjacency matrices and used to classify
665 potential outliers (Z.k < -5); no animal was identified as an outlier. The remaining rows represent
666 the numerical values of all clinical and hematological traits across each animal. Colors indicate
667 an increase (yellow/white) or decreased (purple/black) value for each trait; Sex and BRD are
668 both represented as a value of 1 for bulls and Yes, and 0 for steers and No, respectively.

669

670 Supplemental Figure S2: **Soft threshold (β) selection for signed weighed correlation network**
671 **construction through scale free topology (SFT) plot analysis**

672 A) SFT index R^2 (y-axis) at increasing soft threshold powers (β; x-axis). The value β=8 was
673 selected, seen where the saturation curve is above 0.8 (orange horizontal line). B) Increasing soft
674 threshold powers (β; x-axis) with respect to decreasing mean connectivity (y-axis). The goal of

675 selecting a value β is to maximize scale independence (i.e., suppress low correlations) while
676 simultaneously minimizing loss in mean connectivity.

677

678 Author Contributions

679 Conceptualization: MAS, ARW, ADP, BN; Methodology: MAS, ARW, CES, ADP, BN; Sample
680 and data curation: MAS, ARW, AF, BBK; Computational and statistical analysis: MAS; Animal
681 maintenance and treatment supervision: MAS, ARW, BBK; Project supervision: MAS, ARW,
682 AF, BBK; Writing – original draft preparation: MAS; Writing – Review & Editing: All authors
683 read and approved the final manuscript.

684

685 Competing Interests

686 The author declares that he has no conflict of interest.

687

688 References

689 1. Loneragan GH, Dargatz DA, Morley PS, Smith MA. Trends in mortality ratios among cattle in
690 US feedlots. Journal of the American Veterinary Medical Association. 2001 Oct; 219(8):1122–7.
691
692 2. Brooks KR, Raper KC, Ward CE, Holland BP, Krehbiel CR, Step DL. Economic effects of
693 bovine respiratory disease on feedlot cattle during backgrounding and finishing phases. The
694 Professional Animal Scientist. 2011 Jun; 27(3):195–203.
695
696 3. Griffin D. Economic Impact Associated with Respiratory Disease in Beef Cattle. Veterinary
697 Clinics of North America: Food Animal Practice. 1997 Nov; 13(3):367–77.
698
699 4. Cortes JA, Hendrick S, Janzen E, Pajor EA, Orsel K. Economic impact of digital dermatitis,
700 foot rot, and bovine respiratory disease in feedlot cattle. Translational Animal Science. 2021 Apr
701 1; 5(2):txab076.
702
703 5. Taylor JD, Fulton RW, Lehenbauer TW, Step DL, Confer AW. The epidemiology of bovine
704 respiratory disease: What is the evidence for predisposing factors? Can Vet J. 2010 Oct;
705 51(10):1095–102.
706

6. Grissett GP, White BJ, Larson RL. Structured literature review of responses of cattle to viral and bacterial pathogens causing bovine respiratory disease complex. J Vet Intern Med. 2015 May; 29(3):770–80.

7. McGill JL, Sacco RE. The immunology of bovine respiratory disease. Veterinary Clinics of North America: Food Animal Practice. 2020 Jul; 36(2):333–48.

8. Nobrega D, Andres-Lasheras S, Zaheer R, McAllister T, Homerosky E, Anholt RM, et al. Prevalence, risk factors, and antimicrobial resistance profile of respiratory pathogens isolated from suckling beef calves to reprocessing at the feedlot: a longitudinal study. Front Vet Sci. 2021 Nov 2; 8:764701.

9. Mulliniks T, Funston R. Developmental Programming in Livestock Production, An Issue of Veterinary Clinics of North America: Food Animal Practice - Ebook. 2019.

10. USDA. Part IV: Health and Health Management on U.S. Feedlots with a Capacity of 1,000 or More Head. USDA-APHIS-VS-CEAH-NAHMS; 2013. (Feedlot 2011).

11. White BJ, Renter DG. Bayesian Estimation of the Performance of Using Clinical Observations and Harvest Lung Lesions for Diagnosing Bovine Respiratory Disease in Post-weaned Beef Calves. J VET Diagn Invest. 2009 Jul; 21(4):446–53.

12. Wolfger B, Timsit E, White BJ, Orsel K. A Systematic Review of Bovine Respiratory Disease Diagnosis Focused on Diagnostic Confirmation, Early Detection, and Prediction of Unfavorable Outcomes in Feedlot Cattle. Veterinary Clinics of North America: Food Animal Practice. 2015 Nov; 31(3):351–65.

13. Gifford CA, Holland BP, Mills RL, Maxwell CL, Farney JK, Terrill SJ, et al. GROWTH AND DEVELOPMENT SYMPOSIUM: Impacts of inflammation on cattle growth and carcass merit. Journal of Animal Science. 2012 May 1; 90(5):1438–51.

14. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. Journal of Laboratory and Clinical Medicine. 2006 Mar; 147(3):126–32.

15. Tan L-L, Lyon AR. Role of Biomarkers in Prediction of Cardiotoxicity During Cancer Treatment. Curr Treat Options Cardio Med. 2018 Jul; 20(7):55.

16. Richeson JT, Pinedo PJ, Kegley EB, Powell JG, Gadberry MS, Beck PA, et al. Association of hematologic variables and castration status at the time of arrival at a research facility with the risk of bovine respiratory disease in beef calves. Journal of the American Veterinary Medical Association. 2013 Oct; 243(7):1035–41.

17. Lindholm-Perry AK, Kuehn LA, McDaneld TG, Miles JR, Workman AM, Chitko-McKown CG, et al. Complete blood count data and leukocyte expression of cytokine genes and cytokine

752 receptor genes associated with bovine respiratory disease in calves. BMC Res Notes. 2018 Dec;
753 11(1):786.
754
755 18. Sun H-Z, Srithayakumar V, Jiminez J, Jin W, Hosseini A, Raszek M, et al. Longitudinal
756 blood transcriptomic analysis to identify molecular regulatory patterns of bovine respiratory
757 disease in beef cattle. Genomics. 2020 Nov; 112(6):3968–77.
758
759 19. Jiminez J, Timsit E, Orsel K, van der Meer F, Guan LL, Plastow G. Whole-Blood
760 Transcriptome Analysis of Feedlot Cattle With and Without Bovine Respiratory Disease. Front
761 Genet. 2021 Mar 8; 12:627623.
762
763 20. Scott MA, Woolums AR, Swiderski CE, Perkins AD, Nanduri B, Smith DR, et al.
764 Multipopulational transcriptome analysis of post-weaned beef cattle at arrival further validates
765 candidate biomarkers for predicting clinical bovine respiratory disease. Sci Rep. 2021 Dec;
766 11(1):23877.
767
768 21. Tieri P, Farina L, Petti M, Astolfi L, Paci P, Castiglione F. Network Inference and
769 Reconstruction in Bioinformatics. In: Encyclopedia of Bioinformatics and Computational
770 Biology. Elsevier; 2019. pp. 805–13.
771
772 22. Farber CR, Mesner LD. A Systems-Level Understanding of Cardiovascular Disease through
773 Networks. In: Translational Cardiometabolic Genomic Medicine. Elsevier; 2016. pp. 59–81.
774
775 23. Kadarmideen HN, Watson-haigh NS. Building gene co-expression networks using
776 transcriptomics data for systems biology investigations: Comparison of methods using
777 microarray data. Bioinformation. 2012 Sep 21; 8(18):855–61.
778
779 24. Hasankhani A, Bahrami A, Sheybani N, Fatehi F, Abadeh R, Ghaem Maghami Farahani H,
780 et al. Integrated Network Analysis to Identify Key Modules and Potential Hub Genes Involved in
781 Bovine Respiratory Disease: A Systems Biology Approach. Front Genet. 2021 Oct 18;
782 12:753839.
783
784 25. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2019.
785 Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
786
787 26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
788 data. Bioinformatics. 2014 Aug 1; 30(15):2114–20.
789
790 27. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
791 requirements. Nat Methods. 2015 Apr; 12(4):357–60.
792
793 28. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and
794 genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug; 37(8):907–15.
795

29. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015 Mar; 33(3):290–5.

30. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019 Dec; 20(1):278.

31. Pertea G. prepDE.py. 2019. Available from: https://github.com/gpertea/stringtie/blob/master/prepDE.py

32. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline . F1000Research; 2016 Aug . Report No.: 5:1438.

33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1; 26(1):139–40.

34. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research. 2012 May 1; 40(10):4288–97.

35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec; 9(1):559.

36. Horvath S. Weighted network analysis: application in genomics and systems biology. New York, NY: Springer; 2011.

37. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. BMC Genomics. 2009 Dec; 10(1):327.

38. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol. 2007 Dec; 1(1):54.

39. Garnier S, Ross N, Rudis B, Filipovic-Pierucci A, Galili T, Greenwell B, et al. viridis. 2021. Available from: https://github.com/sjmgarnier/viridis/tree/v0.6.1CRAN

40. Scott MA, Woolums AR, Swiderski CE, Perkins AD, Nanduri B, Smith DR, et al. Whole blood transcriptomic analysis of beef cattle at arrival identifies potential predictive molecules

and mechanisms that indicate animals that naturally resist bovine respiratory disease. PLoS ONE. 2020 Jan 13; 15(1):e0227507.

41. Scott M, Woolums A, Swiderski C, Thompson A, Perkins A, Nanduri B, et al. Use of nCounter mRNA Profiling to Identify at-arrival Gene Expression Patterns for Predicting Bovine Respiratory Disease in Beef Cattle. In Review; 2021 Sep. Available from: https://www.researchsquare.com/article/rs-923361/v1

42. Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible? Bourne PE, editor. PLoS Comput Biol. 2011 Jan 20; 7(1):e1001057.

43. Horvath S, Dong J. Geometric Interpretation of Gene Coexpression Network Analysis. Miyano S, editor. PLoS Comput Biol. 2008 Aug 15; 4(8):e1000117.

44. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Research. 2019 Jul 2; 47(W1):W199–205.

45. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Research. 2019 Nov 6; gkz1031.

46. Bakhtiarizadeh MR, Mirzaei S, Norouzi M, Sheybani N, Vafaei Sadi MS. Identification of Gene Modules and Hub Genes Involved in Mastitis Development Using a Systems Biology Approach. Front Genet. 2020 Jul 13; 11:722.

47. Uribe-Querol E, Rosales C. Phagocytosis: Our Current Understanding of a Universal Biological Process. Front Immunol. 2020 Jun 2; 11:1066.

48. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research. 2021 Jan 8; 49(D1):D605–12.

49. Brohée S, van Helden J. Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics. 2006 Dec; 7(1):488.

50. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965 Dec 1; 52(3–4):591–611.

51. Amrine DE, McLellan JG, White BJ, Larson RL, Renter DG, Sanderson M. Evaluation of three classification models to predict risk class of cattle cohorts developing bovine respiratory disease within the first 14 days on feed using on-arrival and/or pre-arrival information. Computers and Electronics in Agriculture. 2019 Jan; 156:439–46.

52. Abell KM, Theurer ME, Larson RL, White BJ, Apley M. A mixed treatment comparison meta-analysis of metaphylaxis treatments for bovine respiratory disease in beef cattle1,2. Journal of Animal Science. 2017 Feb 1; 95(2):626–35.

53. Zygmuntowicz A, Burmańczuk A, Markiewicz W. Selected Biological Medicinal Products and Their Veterinary Use. Animals. 2020 Dec 9; 10(12):2343.

54. Martin B. Effect of Zelnate Administered as a Metaphylactic upon Initial Processing of High-Risk, Newly Received Beef Calves on Performance and Morbidity [Animal Science Undergraduate Honors Theses]. University of Arkansas; 2020. Available from: https://scholarworks.uark.edu/anscuht/41/

55. Gaspers J, Swanson K, Keomanivong F, Fontoura A, Ward A, Knutson E, et al. Evaluation of response to vaccination with a bacterial-produced plasmid DNA, Zelnate, on feedlot performance of weaned calves . North Dakota State University; 2016 p. 9–11. (2016 North Dakota Beef Report). Report No.: AS1815. Available from: https://www.ag.ndsu.edu/publications/livestock/2016-north-dakota-beef-report

56. Lippolis K, Cooke R, Marques R, Brandão A, Schubach K, Bohnert D. Feeding immunostimulant ingredients to optimize health and performance of receiving cattle . Oregon State University; 2017 p. 9–13. (Oregon Beef Council Report). Report No.: BEEF153. Available from: https://blogs.oregonstate.edu/beefcattle/research-reports/

57. Coetzee JF, Magstadt DR, Sidhu PK, Follett L, Schuler AM, Krull AC, et al. Association between antimicrobial drug class for treatment and retreatment of bovine respiratory disease (BRD) and frequency of resistant BRD pathogen isolation from veterinary diagnostic laboratory samples. Munderloh UG, editor. PLoS ONE. 2019 Dec 13; 14(12):e0219104.

58. Coetzee JF, Cernicchiaro N, Sidhu PK, Kleinhenz MD. Association between antimicrobial drug class selection for treatment and retreatment of bovine respiratory disease and health, performance, and carcass quality outcomes in feedlot cattle. Journal of Animal Science. 2020 Apr 1; 98(4):skaa109.

59. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. Journal of Laboratory and Clinical Medicine. 2006 Mar; 147(3):126–32.

60. Obeidat M, Nie Y, Chen V, Shannon CP, Andiappan AK, Lee B, et al. Network-based analysis reveals novel gene signatures in peripheral blood of patients with chronic obstructive pulmonary disease. Respir Res. 2017 Dec; 18(1):72.

61. Latimer KS, Duncan JR, Prasse KW. Duncan & Prasse's veterinary laboratory medicine: clinical pathology. 5th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2011. pp. 509.

62. Roland L, Drillich M, Iwersen M. Hematology as a diagnostic tool in bovine medicine. J VET Diagn Invest. 2014 Sep; 26(5):592–8.

63. Jones M. Interpreting ruminant bloodwork. In: Proceedings of VetFest 2020 . Australian Veterinary Association; p. 184–9. Available from: https://vetfest.ava.com.au/client_uploads/jones-m-interpreting-ruminant-bloodwork-24a65b87-8843-4ccf-b309-cfe519183589/download

64. Spare MR, Hanzlicek GA, Wootten KL, Anderson GA, Thomson DU, Sanderson MW, et al. Bovine anaplasmosis herd prevalence and management practices as risk-factors associated with herd disease status. Veterinary Parasitology. 2020; 277:100021.

65. Okafor CC, Collins SL, Daniel JA, Coetzee JF, Whitlock BK. Seroprevalence of bovine Anaplasmosis in Georgia. Veterinary Parasitology: Regional Studies and Reports. 2019 Jan; 15:100258.

66. Hosokawa H, Rothenberg EV. How transcription factors drive choice of the T cell fate. Nat Rev Immunol. 2021 Mar; 21(3):162–76.

67. Pasman Y, Merico D, Kaushik AK. Preferential expression of IGHV and IGHD encoding antibodies with exceptionally long CDR3H and a rapid global shift in transcriptome characterizes development of bovine neonatal immunity. Developmental & Comparative Immunology. 2017 Feb; 67:495–507.

68. Tindemans I, Serafini N, Di Santo JP, Hendriks RW. GATA-3 Function in Innate and Adaptive Immunity. Immunity. 2014 Aug; 41(2):191–206.

69. Hosoya T, Maillard I, Engel JD. From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation: GATA-3 regulates multiple stages of T-cell development. Immunological Reviews. 2010 Nov; 238(1):110–25.

70. Li P, Zheng S, Jiang C, Zhou S, Tian H, Zhang G, et al. Th2 lymphocytes migrating to the bone marrow under high-altitude hypoxia promote erythropoiesis via activin A and interleukin-9. Experimental Hematology. 2014 Sep; 42(9):804–15.

957 71. Guo Z, González JF, Hernandez JN, McNeilly TN, Corripio-Miyar Y, Frew D, et al. Possible
958 mechanisms of host resistance to Haemonchus contortus infection in sheep breeds native to the
959 Canary Islands. Sci Rep. 2016 Sep; 6(1):26200.
960
961 72. Scott MA, Woolums AR, Swiderski CE, Perkins AD, Nanduri B, Smith DR, et al.
962 Comprehensive at-arrival transcriptomic analysis of post-weaned beef cattle uncovers type I
963 interferon and antiviral mechanisms associated with bovine respiratory disease mortality. Ortega-
964 Villaizan M del M, editor. PLoS ONE. 2021 Apr 26; 16(4):e0250758.
965
966 73. Diehl S, Rincón M. The two faces of IL-6 on Th1/Th2 differentiation. Molecular
967 Immunology. 2002 Dec; 39(9):531–6.
968
969 74. Gershwin LJ, Berghaus LJ, Arnold K, Anderson ML, Corbeil LB. Immune mechanisms of
970 pathogenetic synergy in concurrent bovine pulmonary infection with Haemophilus somnus and
971 bovine respiratory syncytial virus. Veterinary Immunology and Immunopathology. 2005 Aug;
972 107(1–2):119–30.
973
974 75. McGinley J, Thwaites R, Brebner W, Greenan-Barrett L, Aerssens J, Öner D, et al. A
975 Systematic Review and Meta-analysis of Animal Studies Investigating the Relationship Between
976 Serum Antibody, T Lymphocytes, and Respiratory Syncytial Virus Disease. The Journal of
977 Infectious Diseases. 2021 Sep 15; jiab370.
978
979 76. Chen C, Yuan J, Ji G, Zhang S, Gao Z. Amphioxus ribosomal proteins RPS15, RPS18,
980 RPS19 and RPS30-precursor act as immune effectors via killing or agglutinating bacteria. Fish &
981 Shellfish Immunology. 2021 Nov; 118:147–54.
982
983 77. Xu X, Xiong X, Sun Y. The role of ribosomal proteins in the regulation of cell proliferation,
984 tumorigenesis, and genomic integrity. Sci China Life Sci. 2016 Jul; 59(7):656–72.
985
986 78. Behura SK, Tizioto PC, Kim J, Grupioni NV, Seabury CM, Schnabel RD, et al. Tissue
987 Tropism in Host Transcriptional Response to Members of the Bovine Respiratory Disease
988 Complex. Sci Rep. 2017 Dec; 7(1):17938.
989
990 79. Scott MA, Woolums AR, Swiderski CE, Perkins AD, Nanduri B. Genes and regulatory
991 mechanisms associated with experimentally-induced bovine respiratory disease identified using
992 supervised machine learning methodology. Sci Rep. 2021 Dec; 11(1):22916.
993
994 80. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks.
995 Proceedings of the National Academy of Sciences. 2003 Oct 14; 100(21):12123–8.
996
997 81. Maulik U, Basu S, Ray S. Identifying protein complexes in PPI network using non-
998 cooperative sequential game. Sci Rep. 2017 Dec; 7(1):8410.
999

82. Tizioto PC, Kim J, Seabury CM, Schnabel RD, Gershwin LJ, Van Eenennaam AL, et al. Immunological Response to Single Pathogen Challenge with Agents of the Bovine Respiratory Disease Complex: An RNA-Sequence Analysis of the Bronchial Lymph Node Transcriptome. Harrod K, editor. PLoS ONE. 2015 Jun 29; 10(6):e0131459.

83. Lei X, Wang F, Wu F-X, Zhang A, Pedrycz W. Protein complex identification through Markov clustering with firefly algorithm on dynamic protein–protein interaction networks. Information Sciences. 2016 Feb; 329:303–16.

84. Zhang A. Protein interaction networks: computational analysis. Cambridge ; New York: Cambridge University Press; 2009. Pp. 278.

85. Tu S, Chen R, Xu L. A binary matrix factorization algorithm for protein complex prediction. Proteome Sci. 2011 Dec; 9(S1):S18.

86. Poyatos JF, Hurst LD. How biologically relevant are interaction-based modules in protein networks? Genome Biol. 2004; 5(11):R93.

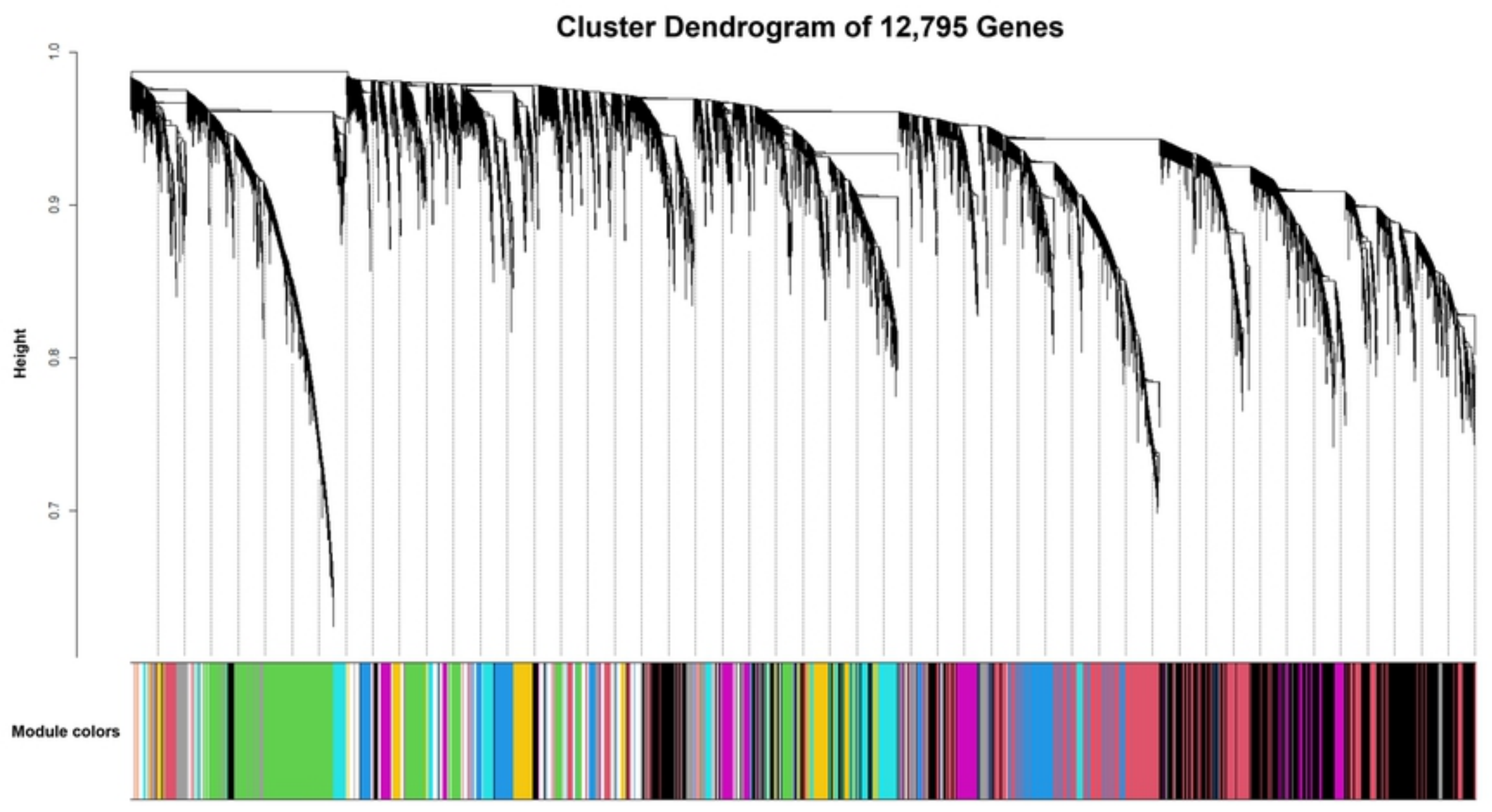

Cluster Dendrogram of 12,795 Genes
Height
1.0
0.9
0.8
0.7
Module colors

Figure 1

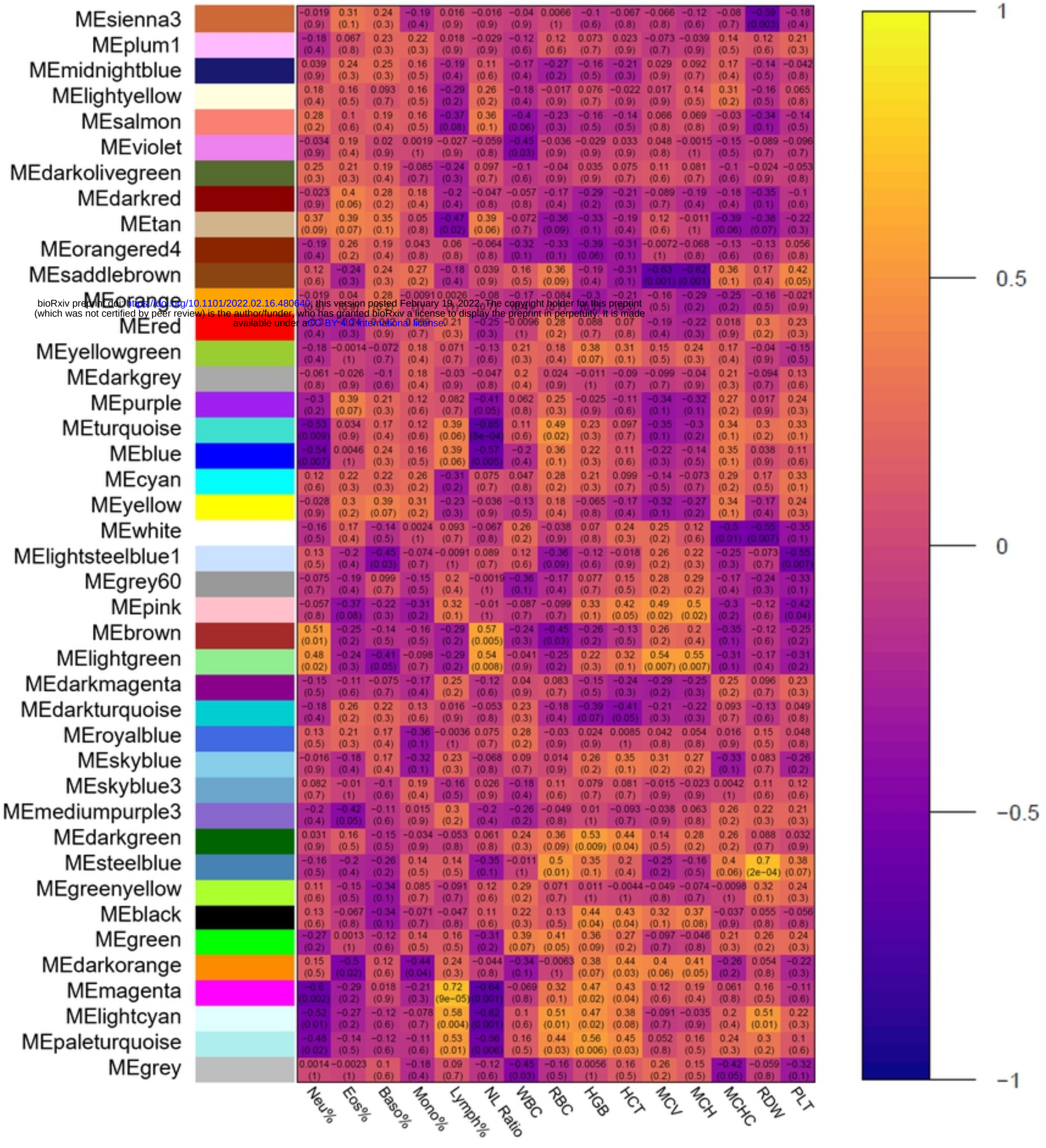


Figure 2

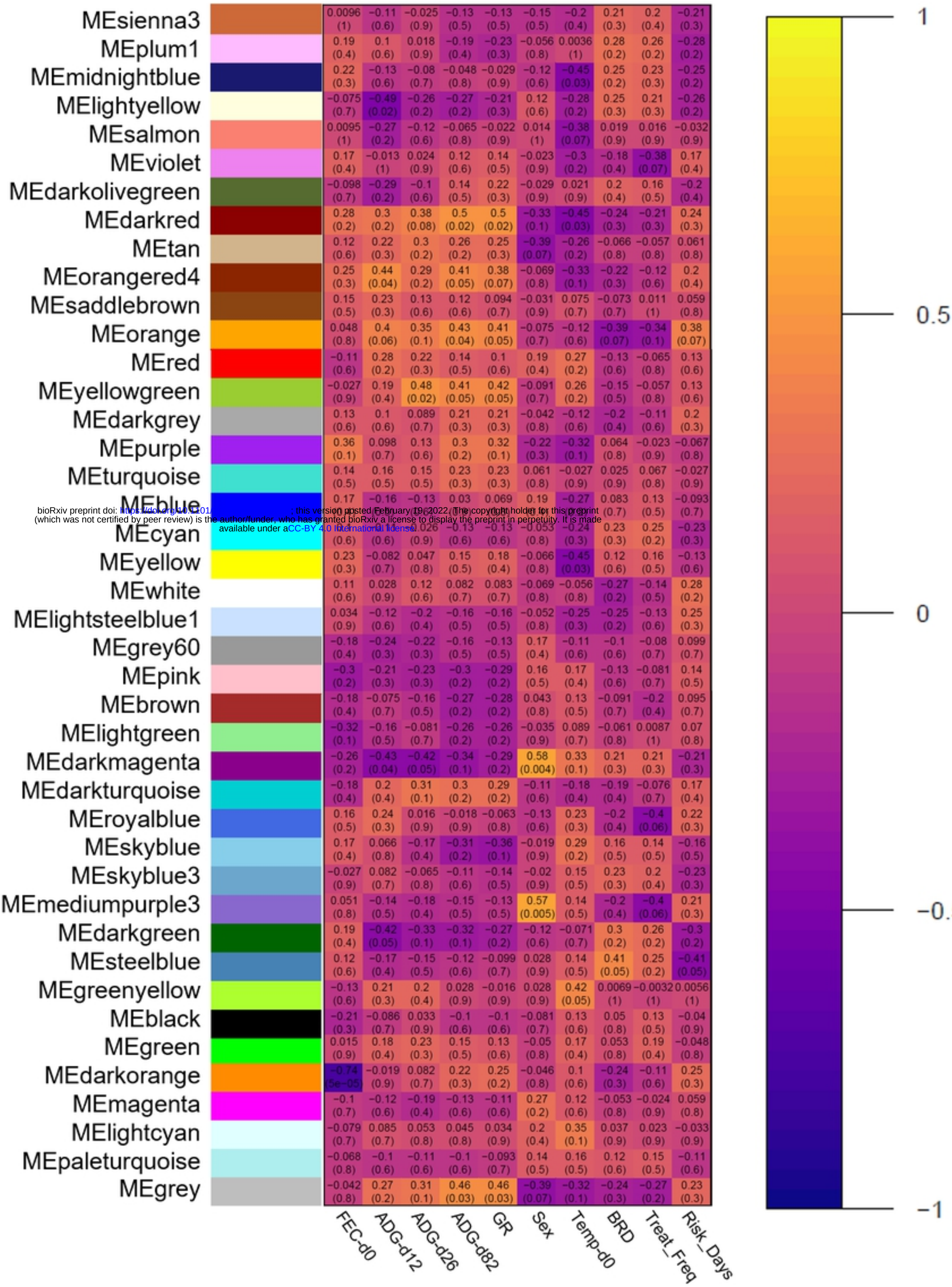


Figure 3

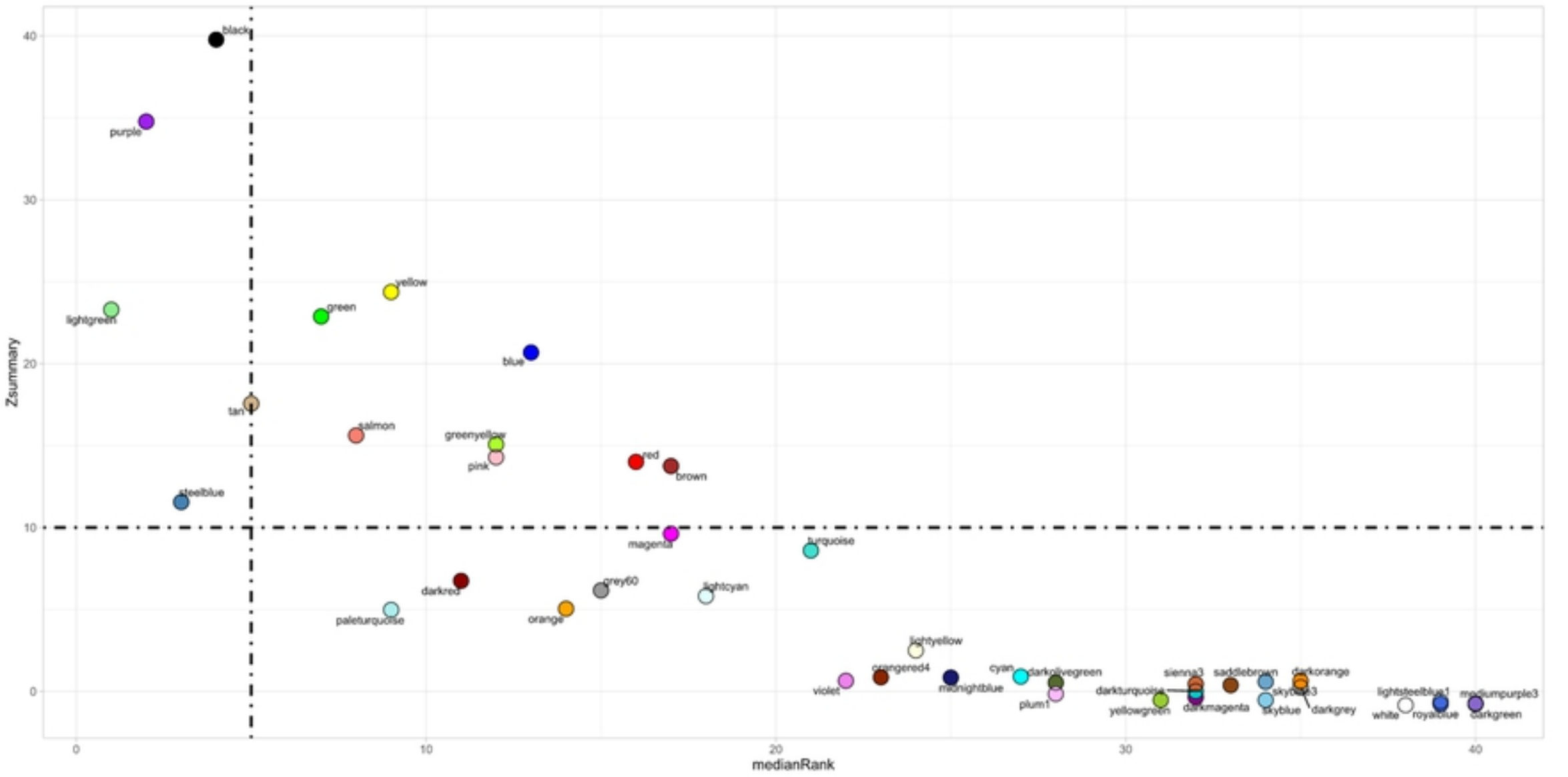

Zsummary
medianRank
black
purple
lightgreen
yellow
green
blue
tan
salmon
greenyellow
pink
red
brown
steelblue
magenta
turquoise
darkred
grey60
lightcyan
paleturquoise
orange
lightyellow
orangered4
violet
midnightblue
cyan
darkolivegreen
plum1
darkturquoise
yellowgreen
sienna3
darkmagenta
saddlebrown
skyblue3
skyblue
darkorange
darkgrey
lightsteelblue1
white
royalblue
mediumpurple3
darkgreen

Figure 4

Color	Cluster ID	Count	Protein Names
Red	Cluster 1	12	ALKBH2, FUS, HNRNPDL, KIAA1524, LRSAM1, PAPOLA, RBM22, RBM25, RNF113A, SRSF5, SRSF9, SUGP1
Brown	Cluster 2	11	ABL2, ANKRD28, ARID4B, BLK, BLNK, CD247, ENAH, KAT5, MDM4, PSD3, RHOBTB1
Dark Golden Rod	Cluster 3	8	ANAPC5, FOXO1, GSTM4, JUN, SOX4, ZNF362, ZNF462, ZSCAN12
Green-Yellow	Cluster 4	5	ALOX15, DBN1, LDLRAP1, LRP5, WNT5A
Lime Green	Cluster 5	4	BTRC, CUL1, KLHDC8A, LOC100298356
Light Green	Cluster 6	2	CXCL12, GPR161
Green	Cluster 7	2	ADSS, PGM3
Cyan	Cluster 8	2	POT1, TINF2
Blue	Cluster 9	2	MBD2, WDR20
Purple	Cluster 10	2	IRF2BP2, SOX13
Pink	Cluster 11	2	ETF1, RPLP0

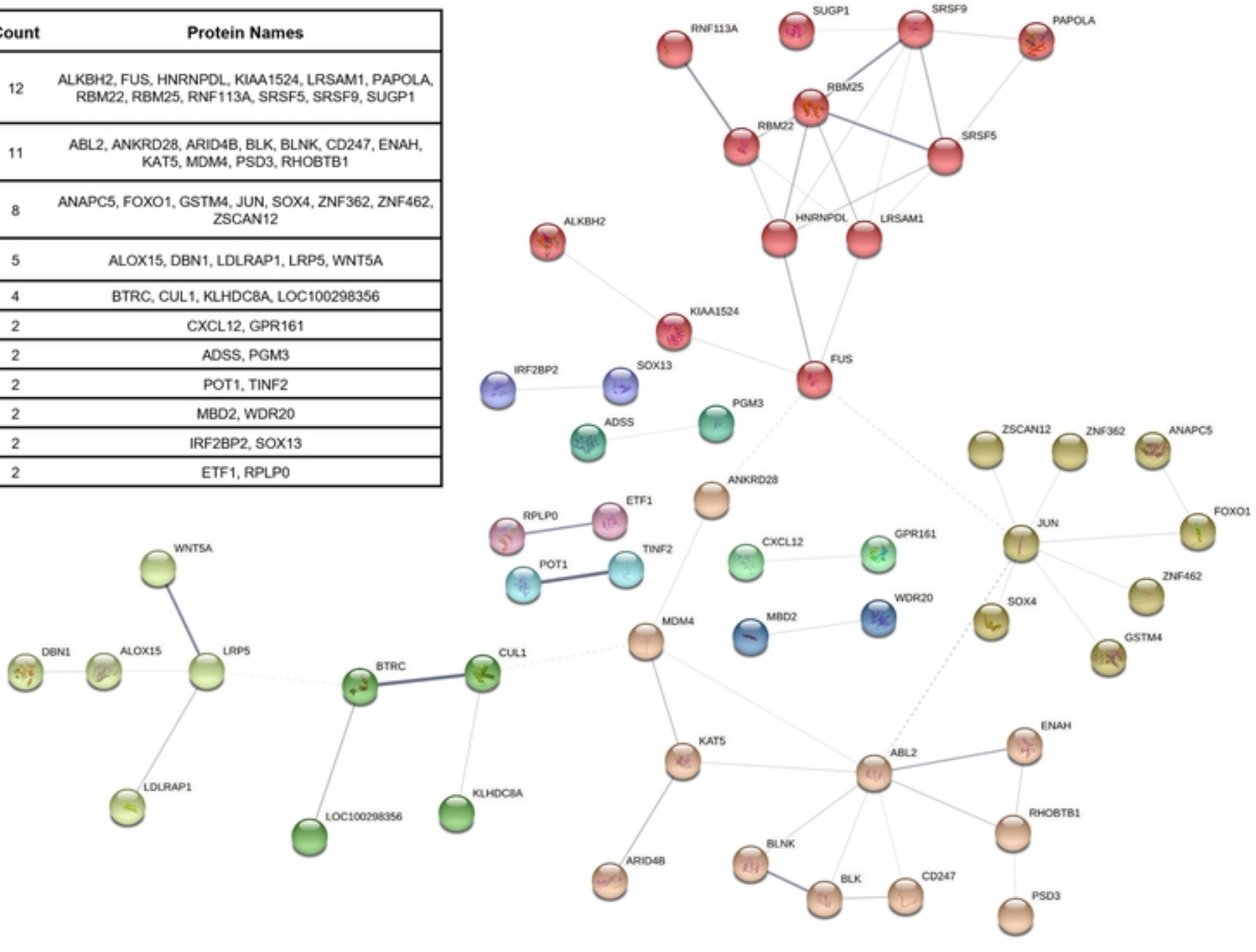
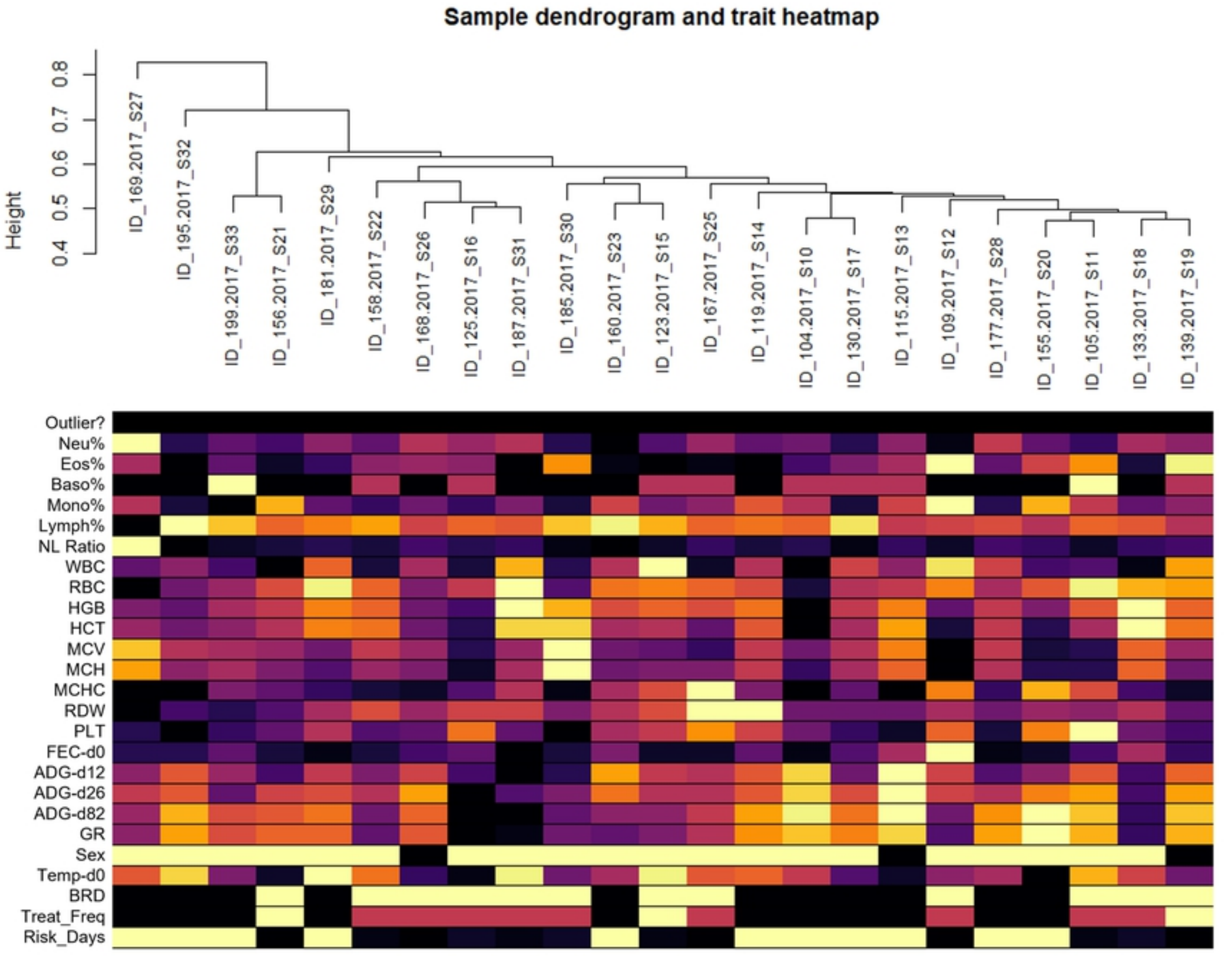
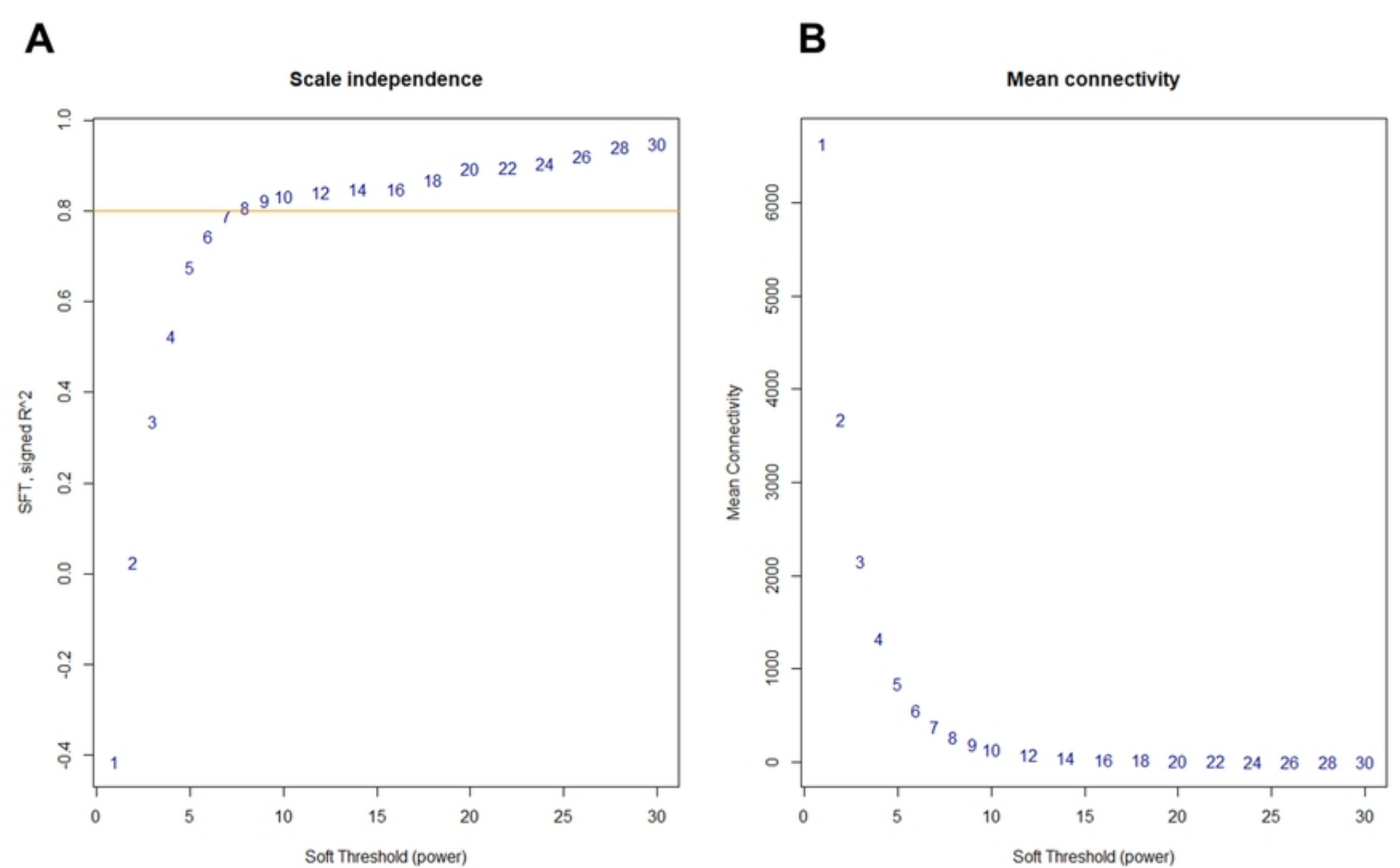


Figure 5



Supplementary Figure S1



Supplementary Figure S2