JSR Publications

Available Online

J. Sci. Res. **14** (2), 593-599 (2022)

JOURNAL OF SCIENTIFIC RESEARCH

www.banglajol.info/index.php/JSR

A 20-Gene Expression Diagnostic Signature of Bovine Respiratory Disease in Cattle

A. Giwa[1*], R. Giwa[2]

[1]Department of Zoology and Environmental Biology, Lagos State University, Lagos, Nigeria

[2]Department of Animal Production and Health, Federal University of Agriculture, Abeokuta, Nigeria

Received 16 August 2021, accepted in final revised form 4 February 2022

Abstract

Bovine Respiratory Disease (BRD) is a prevalent disease in cattle rearing systems globally with significant health and economic costs. Current diagnostic methods of BRD rely on subjective visual signs and physical examination, which are suboptimal. This study, therefore, aims to find a blood-based gene expression signature for the diagnostic identification of BRD in cattle. The Gene Expression Omnibus dataset, GSE152959, was downloaded and used for analysis. The analyses performed included differential gene expression (DGE), clustering and machine learning prediction. Ninety genes were differentially expressed in BRD samples compared to controls. The GSE150706 dataset was used as the test dataset for machine learning prediction. The DEGs identified clustered the GSE150706 samples with good accuracy. For the machine learning prediction, 92 % of correctly predicted samples were obtained using twenty genes as features. Therefore, the identified 20-gene expression signature has BRD diagnostic utility in cattle. This signature could potentially be used to develop standardized and reliable diagnostic tests of BRD in cattle. Improved diagnostics will lead to early detection and treatment, reducing the health and economic costs associated with the disease. Further validation in larger cattle cohorts is required.

Keywords: Bovine respiratory disease; Gene expression profiling; Machine learning; Cattle; Biomarker.

doi: http://dx.doi.org/10.3329/jsr.v14i2.55193 J. Sci. Res. **14** (2), 593-599 (2022)

1. Introduction

Bovine Respiratory Disease (BRD) is a prevalent disease in cattle rearing systems and is responsible for a huge amount of the ill health and mortality occurring in feed yards globally [1–3]. The disease causes high economic costs as a result of the decreased production and treatment costs increase [4,5]. BRD is caused by multiple factors, including host, pathogen, and environmental factors [6]. However, bacterial and viral members of the BRD Complex (BRDC) are the primary cause of the disease [7]. The major viral causes of BRD include bovine respiratory syncytial virus (BRSV), bovine parainfluenza virus type 3 (BPIV3), and bovine herpesvirus type 1 (BHV-1) [8].

* *Corresponding author*: abdulazeez.giwa@lasu.edu.ng

Investigating the underlying biology of the disease in cattle is crucial to reducing the significant health and economic costs associated with the disease.

Performing gene expression profiling to characterize cattle biology and disease is well-established [9–12]. RNA sequencing (RNA-Seq) is the current technology for measuring and quantifying gene expression. This is because it enables a more comprehensive mapping and transcriptome quantification method to identify genes differentially expressed between multiple conditions [13,14]. Transcriptome characterization by RNA-Seq is important for identifying biomarkers [15]. Gene expression profiling of bovine blood and bronchial lymph node tissues has been done to characterize gene expression changes due to BRD [7,16–19]. These reports have increased our understanding of the transcriptomic changes associated with BRD. However, current diagnostic methods of BRD in cattle are still suboptimal.

Common general clinical signs produce change in feed intake, body temperature increase, nasal discharge, decreased milk production, cough, and increased respiratory rate [20]. Diagnosis of BRD is by physical examination of the animal for visual signs of sickness [20,21]. This is a subjective method dependent on the accuracy of the evaluation done by the individual attempting the diagnosis. Indeed, many cases at the subclinical level are not detected until the animals are slaughtered [22]. Improved diagnostic methods would allow for the early detection of the disease, which would improve treatment outcomes. Therefore, the search for better methods of diagnosis of BRD is necessary. This study, therefore, aims to find a blood-based gene expression signature for the diagnostic identification of BRD in cattle.

2. Materials and Methods

2.1. *Dataset*

The GSE152959 [7] BRD dataset was downloaded from the Gene Expression Omnibus (GEO) [23]. This dataset is a gene expression dataset derived from the whole blood of cattle challenged with BRSV and control cattle. This dataset is comprised of eighteen (18) samples, including twelve (12) BRD (BRSV challenged) samples and six (6) control samples. The gene expression values were in raw counts format.

2.2. *Differential gene expression analysis*

Low expressed count values were removed using the filterByExpr function in the edgeR package in R [24]. Differential gene expression (DGE) analysis was then performed between the BRD and control groups with the DESEq2 R package [25]. An adjusted *p*-value < 0.05 and a 1.5 fold change threshold were set in the DGE analysis to identify differentially expressed genes. Thereafter, a heatmap of the genes differentially expressed was plotted with the pheatmap package in R. The corresponding gene symbols of the Ensembl IDs were obtained using the biomaRt R package [26,27].

2.3. *Machine learning*

The differentially expressed genes were assessed for their use as biomarkers of BRD in cattle. The GSE152959 dataset was used as the training set, while the GSE150706 [19] dataset was used as the testing set. The GSE150706 dataset is a gene expression dataset derived from the blood of 24 samples at three different stages; Entry (at arrival in feedlot), Pulled (with identified sickness), and Close-out (At recovery, healthy animal). These three stages represented subclinical, clinical, and healthy states, respectively [19]. All samples were used as the testing set for machine learning prediction. The close-out stage was the healthy state, while the entry and pulled stages were the diseased stages.

Firstly, the DEGs identified in the DGE analysis were used to cluster the GSE150706 samples to assess their possible use for machine learning classification tasks. The machine learning feature selection technique was then performed. Log2 transformed counts values were used for constructing the training and test set. Thereafter, we performed recursive feature elimination (RFE) to obtain the features with better classification performance. For the RFE, a linear support vector machine (SVM) algorithm was used for building a model, which was assessed with repeated stratified 5-fold cross-validation to find the best features and parameters for classification. The differentially expressed genes obtained from RFE were used to build an SVM model. LibSVM [28] was used to create the SVM model using the polynomial kernel and a cost value of 2. A built-in LibSVM python script was used to scale the training and test sets' feature values. Thereafter, the 72 samples of the GSE150706 dataset were used to test the SVM model. The metrics for evaluating the model were accuracy, f1-score, recall, and precision.

3. Results and Discussion

The filtration step of the DGE analysis removed 11717 genes with low expression and kept 12879 genes for further analysis. The DGE analysis identified 90 differentially expressed genes (DEGs) between the two comparison groups. Among these 90 DEGs, 68 were upregulated while 22 were downregulated. This result demonstrates a strong host transcriptional response reflected in the blood tissue of cattle with BRD. Similar results were reported by Scott *et al.* [18], who also noted transcriptional differences between cattle that developed BRD and those that did not develop BRD (healthy) within the same herd.

The DEGs accurately clustered the samples into their respective groups (Fig 1A), differentiating diseased samples from controls. Clustering the GSE150706 samples using the DEGs identified demonstrated that the DEGs could potentially be used for machine learning classification (Fig. 1B). The entry and pulled samples generally clustered together and were separated from the close-out samples (Fig. 1B). These results demonstrated the classification ability of the DEGs as shown in the clustering results of the GSE150706 test dataset, where the entry (subclinical) and pulled (clinical) samples were generally clustered together. This is consistent and in line with the fact that the Entry and Pulled stages are BRD stages but different disease time points [19].

Twenty genes were selected from the 90 DEGs to be the best features for classification from the RFE performed (Table 1). These 20 genes were used in constructing the machine learning training and test sets (Table 1). The repeated stratified 5-fold cross-validation resulted in an accuracy of 100%. Assessment of the SVM model with the GSE150706 test set resulted in an accuracy of 92% (66/72) of correctly predicted samples (Table 2). Four healthy state samples and two diseased samples were misclassified. The high accuracy results from the SVM model (Table 2) built on the 20 DEGs demonstrate that they could be used as diagnostic indicators of BRD in cattle. This 20-gene expression signature, therefore, has BRD diagnostic utility. Many of these genes are associated with inflammatory response and immune regulation [29–38]. Activation of the immune response is crucial to resolving the disease.

Improvement in diagnosis of BRD will lead to early detection, eliminate non-detection of subclinical cases, early treatment, ultimately reducing the significant morbidity, mortality, and economic costs associated with the disease. However, further validation in larger cattle cohorts is required.

Table 1. Information on 20 DEGs selected by RFE used for machine learning prediction.

Ensembl ID	Gene symbol	log2FC	*p*-adj	Status
ENSBTAG00000014113	CCL8	4.54692	9.57953e-11	Up
ENSBTAG00000015596	SLCO2B1	4.49062	1.0740e-06	Up
ENSBTAG00000003152	IFI27	3.99460	2.10357e-05	Up
ENSBTAG00000013167	SIGLEC1	3.74138	3.27498e-11	Up
ENSBTAG00000001826	SASH1	2.84225	2.24283e-07	Up
ENSBTAG00000014529	GBP4	2.64954	5.32350e-15	Up
ENSBTAG00000011467	BATF2	2.55206	2.24283e-07	Up
ENSBTAG00000016042	TM6SF2	2.41732	0.00903	Up
ENSBTAG00000038737	P2RY6	2.38926	3.04185e-08	Up
ENSBTAG00000014046	BPI	2.31362	0.00557	Up
ENSBTAG00000010057	GZMB	2.18215	0.00707	Up
ENSBTAG00000019018	LOC112441484	1.72261	0.03756	Up
ENSBTAG00000019015	IFITM3	1.71952	3.26177e-14	Up
ENSBTAG00000008612	C1R	1.58201	0.01466	Up
ENSBTAG00000008248	DMD	-2.97592	2.24283e-07	Down
ENSBTAG00000007073	CPB2	-2.34423	0.00056	Down
ENSBTAG00000013305*		-2.24336	0.00606	Down
ENSBTAG00000007490	SULF2	-1.85160	9.59336e-05	Down
ENSBTAG00000008945	SDSL	-1.82149	0.00902	Down
ENSBTAG00000013078	DNAH11	-1.57412	0.00735	Down

Shown are Ensembl IDs, gene symbols, fold change (log2FC), *p*-adjusted value and expression status. *gene name/symbol not available.

Table 2. Prediction results of the GSE150706 test set of BRD and healthy cattle.

Class	Precision	Recall	F1 score
Healthy	0.91	0.83	0.87
BRD	0.92	0.96	0.94
Accuracy	92%		

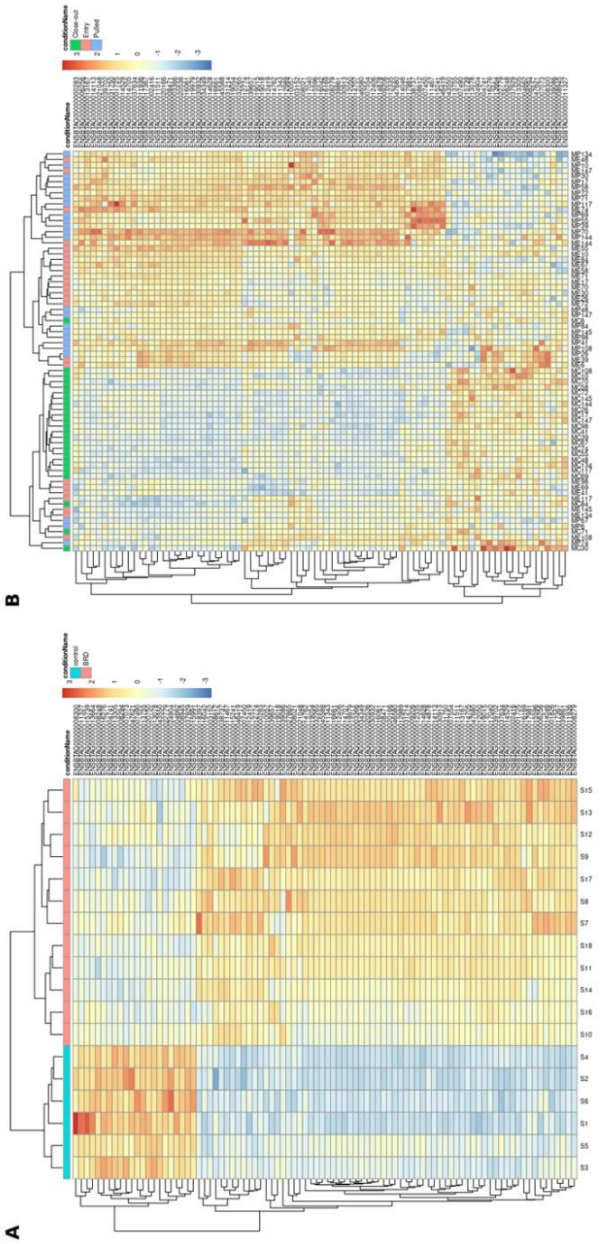

Fig. 1. Heatmap and clustering of BRD and control/healthy samples of the (A) GSE152959 dataset (B) GSE150706 test dataset

4. Conclusion

In this study, we identified a 20-gene expression signature capable of diagnosing Bovine Respiratory Disease in cattle. This signature could potentially be used to develop standardized and reliable diagnostic tests of Bovine Respiratory Disease in cattle. This would be an improvement over the current subjective and visual diagnosis of the disease.

References

1. R. A. Smith, J. Anim. Sci. **76**, 272 (1998). https://doi.org/10.2527/1998.761272x
2. D. G. Miles, Anim. Health Res. Rev. **10**, 101 (2009). https://doi.org/10.1017/S1466252309990090
3. A. L. Baptista, A. L. Rezende, P. A. Fonseca, R. P. Massi, G. M. Nogueira, L. Q. Magalhães, S. A. Headley, G. L. Menezes, A. A. Alfieri, and J. P. E. Saut, J. Infect. Dev. Ctries. **11**, 791 (2017). https://doi.org/10.3855/jidc.9296
4. N. K. Chirase and L. W. Greene, Anim. Feed Sci. Technol. **93**, 217 (2001). https://doi.org/10.1016/S0377-8401(01)00277-2
5. C. Blakebrough-Hall, J. P. McMeniman, and L. A. González, J. Anim. Sci. **98,** 1 (2020). https://doi.org/10.1093/jas/skaa005
6. M. Stokstad, T. B. Klem, M. Myrmel, V. S. Oma, I. Toftaker, O. Østerås, and A. Nødtvedt, Front. Vet. Sci. **7**, 167 (2020). https://doi.org/10.3389/fvets.2020.00167
7. D. Johnston, B. Earley, M. S. McCabe, J. Kim, J. F. Taylor, K. Lemon, C. Duffy, M. McMenamy, S. L. Cosby, and S. M. Waters, Sci. Rep. **11**, 9392 (2021). https://doi.org/10.1038/s41598-021-88878-1
8. J. Kirchhoff, S. Uhlenbruck, K. Goris, G. M. Keil, and G. Herrler, Vet. Res. **45**, 20 (2014). https://doi.org/10.1186/1297-9716-45-20
9. S. A. Lehnert, K. A. Byrne, A. Reverter, G. S. Nattrass, P. L. Greenwood, Y. H. Wang, N. J. Hudson, and G. S. Harper, J. Anim. Sci. **84**, 3239 (2006). https://doi.org/10.2527/jas.2006-192
10. D. Streyl, R. Kenngott, N. Herbach, R. Wanke, H. Blum, F. Sinowatz, E. Wolf, H. Zerbe, and S. Bauersachs, Reproduction **143**, 85 (2012). https://doi.org/10.1530/REP-11-0204
11. Y. Gao, X. Lin, K. Shi, Z. Yan, and Z. Wang, PLoS One **8**, ID e70393 (2013). https://doi.org/10.1371/journal.pone.0070393
12. M. Hillreiner, C. Schmautz, I. Ballweg, V. Korenkova, M. W. Pfaffl, and H. Kliem, BMC Vet. Res. **13**, 369 (2017). https://doi.org/10.1186/s12917-017-1293-z
13. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, Nat. Methods **5**, 621 (2008). https://doi.org/10.1038/nmeth.1226
14. C. W. Law, M. Alhamdoosh, S. Su, X. Dong, L. Tian, G. K. Smyth, and M. E. Ritchie, F1000Res. **5**, 1408 (2018). https://doi.org/10.12688/f1000research.9005.3
15. Z. Akond, M. Alam and M. N. Mollah, Bioinformation **14**, 153 (2018). https://doi.org/10.6026/97320630014153
16. P. C. Tizioto, J. Kim, C. M. Seabury, R. D. Schnabel, L. J. Gershwin, A. L. Van Eenennaam, R. Toaff-Rosenstein, H. L. Neibergs, and J. F. Taylor, PLoS One **10**, ID e0131459 (2015). https://doi.org/10.1371/journal.pone.0131459
17. D. Johnston, B. Earley, M. S. McCabe, K. Lemon, C. Duffy, M. McMenamy, S. L. Cosby, J. Kim, G. Blackshields, J. F. Taylor, and S. M. Waters, Sci. Rep. **9**, 14736 (2019). https://doi.org/10.1038/s41598-019-51094-z
18. M. A. Scott, A. R. Woolums, C. E. Swiderski, A. D. Perkins, B. Nanduri, D. R. Smith, B. B. Karisch, W. B. Epperson, and J. R. Blanton, PLoS One **15**, ID e0227507 (2020). https://doi.org/10.1371/journal.pone.0227507
19. H. Z. Sun, V. Srithayakumar, J. Jiminez, W. Jin, A. Hosseini, M. Raszek, K. Orsel, L. L. Guan, and G. Plastow, Genomics **112**, 3968 (2020). https://doi.org/10.1016/j.ygeno.2020.07.014
20. S. Ferraro, G. Fecteau, J. Dubuc, D. Francoz, M. Rousseau, J. P. Roy, and S. Buczinski, J. Dairy Sci. **104**, 7095 (2021). https://doi.org/10.3168/jds.2020-19471
21. B. Wolfger, E. Timsit, B. J. White, and K. Orsel, Vet. Clin. North Am. Food Anim. Pract. **31**, 351 (2015). https://doi.org/10.1016/j.cvfa.2015.05.005
22. D. Griffin. Anim. Health Res. Rev. **15**, 138 (2014). https://doi.org/10.1017/S1466252314000255
23. T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L.

Robertson, N. Serova, S. Davis, and A. Soboleva, Nucleic Acids Res. **41**, D991 (2013). https://doi.org/10.1093/nar/gks1193
24. M. D. Robinson, D. J. McCarthy, and G. K. Smyth, Bioinformatics **26**, 139 (2010). https://doi.org/10.1093/bioinformatics/btp616
25. M. I. Love, W. Huber, and S. Anders, Genome Biol. **15**, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
26. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber, Bioinformatics **21**, 3439 (2005). https://doi.org/10.1093/bioinformatics/bti525
27. S. Durinck, P. T. Spellman, E. Birney, and W. Huber, Nat. Protoc. **4**, 1184 (2009). https://doi.org/10.1038/nprot.2009.97
28. C. C. Chang and C. J. Lin, ACM Trans. Intell. Syst. Technol. **2**, 27 (2011). https://doi.org/10.1145/1961189.1961199
29. D. J. Vestal, J. Interferon Cytokine Res. **25**, 435 (2005). https://doi.org/10.1089/jir.2005.25.435
30. P. R. Crocker, J. C. Paulson, and A. Varki, Nat. Rev. Immunol. **7**, 255 (2007). https://doi.org/10.1038/nri2056
31. S. M. Dauphinee, A. Clayton, A. Hussainkhel, C. Yang, Y. J. Park, M. E. Fuller, J. Blonder, T. D. Veenstra, and A. Karsan. J. Immunol. **191**, 892 (2013). https://doi.org/10.4049/jimmunol.1200583
32. L. Fan, G. Fu, Y. Ding, P. Lv, and H. Li, Int. J. Colorectal Dis. **32**, 433 (2017). https://doi.org/10.1007/s00384-016-2740-1
33. B. M. Tang, M. Shojaei, G. P. Parnell, S. Huang, M. Nalos, S. Teoh, K. O'Connor, S. Schibeci, A. L. Phu, A. Kumar, J. Ho, A. F. A. Meyers, Y. Keynan, T. Ball, A. Pisipati, A. Kumar, E. Moore, D. Eisen, K. Lai, M. Gillett, R. Geffers, H. Luo, F. Gul, J. Schreiber, S. Riedel, D. Booth, A. McLean, and K. Schughart, Eur. Respir. J. **49**, 1602098 (2017). https://doi.org/10.1183/13993003.02098-2016
34. X. Yang, Y. Lou, G. Liu, X. Wang, Y. Qian, J. Ding, S. Chen, and Q. Xiao, J. Neuroinflammation **14**, 38 (2017). https://doi.org/10.1186/s12974-017-0795-8
35. E. Salem, S. Hägglund, H. Cassard, T. Corre, K. Näslund, C. Foret, D. Gauthier, A. Pinard, M. Delverdier, S. Zohari, J. F. Valarcher, M. Ducatez, and G. Meyer, J. Virol. **93**, ID e01853-18 (2019). https://doi.org/10.1128/JVI.01853-18
36. F. A. Cimini, I. Barchetta, V. Ceccarelli, C. Chiappetta, A. Di Biasio, L. Bertoccini, F. Sentinelli, F. Leonetti, G. Silecchia, C. Di Cristofano, M. G. Baroni, F. Velotti, and M. G. Cavallo, Front. Immunol. **11**, ID 589188 (2020). https://doi.org/10.3389/fimmu.2020.589188
37. J. Alghamdi, M. Alaamery, T. Barhoumi, M. Rashid, H. Alajmi, N. Aljasser, Y. Alhendi, H. Alkhalaf, H. Alqahtani, O. Algablan, A. I. Alshaya, N. Tashkandi, S. Massadeh, B. Almuzzaini, S. N. Ehaideb, M. Bosaeed, K. Ayoub, S. Yezli, A. Khan, A. Alaskar, and A. Bouchama, Genomics, **113**, 1733 (2021). https://doi.org/10.1016/j.ygeno.2021.04.002
38. A. S. Bedin, A. Makinson, M. C. Picot, F. Mennechet, F. Malergue, A. Pisoni, E. Nyiramigisha, L. Montagnier, K. Bollore, S. Debiesse, D. Morquin, N. Veyrenche, C. Renault, V. Foulongne, C. Bret, A. Bourdin, V. Le Moing, P. Van de Perre, and E. Tuaillon, J. Infect. Dis. **223**, 562 (2021). https://doi.org/10.1093/infdis/jiaa724