Minimum Redundancy Maximal Relevance Gene Selection of Apoptosis Pathway Genes in Peripheral Blood Mononuclear Cells of HIV-infected Patients with Antiretroviral Therapy-associated Mitochondrial Toxicity

Eliezer Bose
Massachusetts General Hospital Institute of Health Professions
Elijah Paintsil
Yale University
Musie Ghebremichael (musie_ghebremichael@dfci.harvard.edu)
Harvard Medical School, The Ragon Institute of MGH, MIT and Harvard

Research Article

Keywords: HIV, apoptosis, antiretroviral therapy, mitochondrial toxicity, machine learning, maximum relevance minimum redundancy (mRMR)

Posted Date: May 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-487036/v1

Version of Record: A version of this preprint was published at BMC Medical Genomics on December 1st, 2021. See the published version at https://doi.org/10.1186/s12920-021-01136-1.

Abstract

Background: We previously identified differentially expressed genes on the basis of false discovery rate adjusted P value using empirical Bayes moderated tests. However, that approach yielded a subset of differentially expressed genes without accounting for redundancy between the selected genes.

Methods: This study is a secondary analysis of a case-control study of the effect of antiretroviral therapy (ART) on apoptosis pathway genes comprising of 16 cases (HIV infected with mitochondrial toxicity) and 16 controls (uninfected). We applied the maximum relevance minimum redundancy (mRMR) algorithm on the genes that were differentially expressed between the cases and controls. The mRMR algorithm iteratively selects features (genes) that are maximally relevant for class prediction and minimally redundant. We implemented m-fold cross validation using machine learning classifiers and tested the prediction accuracy of the two mRMR genes. We next used network analysis to estimate the association between the differentially expressed genes using partial correlations, relative importance estimates, and centrality measures of each item. The Spinglass algorithm was used to identify clusters of gene communities.

Results: The mRMR algorithm ranked DFFA and TNFRSF1A, two of the upregulated proapoptotic genes, on the top. The overall prediction accuracy was 90%, which clearly show that the mRMR gene sets outperforms the performance of the gene sets based on gene expression analyses. FASLG had highest centrality in cases with ABL1 in controls.

Conclusion: The mRMR algorithm and network analysis revealed a new correlation of genes associated with mitochondrial toxicity.

Background

Although current antiretroviral therapy (ART) has reduced HIV-associated morbidity and mortality (1-4), ART-associated toxicity is still pervasive in people living with HIV (PLWH) (5-7). A recent study from Italy showed that the 1-year probability of discontinuation of ART due to toxicity was 19% for patients who initiated ART between 2008 and 2014 (8). All classes of antiretroviral drugs are associated with toxicity. Nucleoside reverse transcriptase inhibitors (NRTIs), the first class to show anti-HIV activity, are associated with toxicities such as skeletal muscle myopathy, lactic acidosis, lipodystrophy, peripheral neuropathies, cardiomyopathies, and pancytopenia (9-12). These toxicities are due to NRTI-induced mitochondrial dysfunction through the inhibition of mitochondrial DNA (mtDNA) polymerase gamma (Pol-γ) (13). Recently, Pol-γ independent mitochondrial dysfunction has been associated with several components of ART (14-16). For example, protease inhibitors (PIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) do not inhibit Pol-γ and yet they cause toxicities commensurate with mitochondrial dysfunction (17, 18). Although the underlying mechanisms are not well understood, most classes of ART can cause apoptosis, a mitochondrion function (19, 20). Thus, apoptosis biomarkers could be used potentially to diagnose and monitor ART-associated toxicity. We recently reported that in a case-control study (HIV+

with mitochondrial toxicity vs HIV uninfected controls) a total of 26 of 84 genes of the apoptosis pathway were differentially expressed (21).

In the current study, a secondary data analysis, we sought to select the most relevant and least redundant genes in the differential expression profile of the apoptosis pathway in HIV-infected patients with ART-associated mitochondrial toxicity (cases) versus HIV-uninfected individuals (controls). We employed the maximum relevance minimum redundancy (mRMR) algorithm on the 26 genes that were differentially expressed between the cases and controls. This algorithm performs better than the differential gene expression analyses we had previously conducted for the latter fails to account for redundancy between the selected genes. We implemented m-fold cross-validation using two machine learning classifiers (Elastic-Net Regularized Generalized Linear Model and K-nearest neighbors (KNN)). We tested the prediction accuracy of the mRMR gene sets.

Methods

Study Design and Participants:

This study is a secondary analysis of data obtained from a previous case control study comprising of HIV infected individuals with mitochondrial toxicity (cases, n=16) and HIV uninfected individuals (controls, n=16). The rationale, organization, and recruitment of the subjects, biological procedures used have been described previously by Foli et al. (21). In brief, 32 individuals were enrolled from April 2011 to March 2013 at the Yale-New Haven Hospital.. Cases were matched for age, race, and gender to HIV-negative controls. At enrollment, participant's past medical history and demographic information were obtained. For the cases, we reviewed their medical records for medication history, HIV RNA copy number, and CD4+ T-cell count. The Human Apoptosis RT2 Profiler PCR Array kit (SuperArray Biosciences) was used to investigate the impact of ART on apoptosis pathway-specific genes according to manufacturer's instructions. The institutional review board of the Yale School of Medicine approved the study protocol.

Statistical Analysis:

We previously analyzed the data and identified 26 out of the 84 genes differentially expressed between the cases and controls (21). We identified the 26 differentially expressed genes based on the false discovery rate (FDR) adjusted p-value using empirical Bayes moderated tests. In this secondary analysis, we sought to rank further the critical genes which contributed to profiling differences, using minimum redundancy and maximal relevance (mRMR) method. In addition to the mRMR algorithm, network analysis was used to estimate the interactions between the 26 genes using partial correlations, relative importance estimates, and centrality measures of each item. We performed the Network Comparison Test (NCT) to compare the correlation of variables among networks. Finally, we explored clusters of communities within the genes between cases and controls.

Minimum Redundancy and Maximum Relevance (mRMR)

The mRMR algorithm chooses a subset of genes (features) having the most correlation with a class (relevance, the outcome), and the least correlation between themselves (redundancy), ranking features according to the minimal-redundancy-maximal-relevance criteria (22). For continuous features, such as gene expression, the F-statistic was used to calculate correlation with the class (relevance). For correlation between genes (redundancy), the Pearson correlation coefficient was used. Next, genes were selected one by one by applying a greedy search to maximize the objective function, a function that integrates relevance and redundancy information of each gene into a single scoring mechanism (22). Once computed, the algorithm ranks the variables according to their importance score. We estimated the features or genes' predictive accuracy in distinguishing class membership (case or control) using two different machine learning algorithms (elastic-net and KNN).

Elastic-net is a penalization algorithm that uses a combination of ridge and least absolute shrinkage and selection operator (LASSO) regression for model building. Ridge regression creates a regression model penalized with the L2-norm, which shrinks the coefficient values allowing coefficients with a minor contribution to the target variable to get close to zero. On the other hand, LASSO creates a regression model penalized with the L1-norm, which affects shrinking coefficient values allowing some with a minor impact on the target variable to become zero (23, 24). Elastic net penalization creates a regression model with both the L1-norm and L2-norm. Using a cross-validation procedure, elastic-net models pick an optimal value for the penalization parameter (λ). Discrete estimates of the coefficients (βs) are made along the way. KNN makes predictions for new instances by searching through the entire training set for the K most similar instances (the neighbors) and summarizing the output variable for those K instances, for our study, classifying the mode (or the most common) class value. To determine which of the K instances in the training dataset are most like new input, we used a Euclidean distance measure (24, 25).

Network analysis

Network analysis (26) involved four steps: 1) estimate a statistical model using network analysis; 2) use graph theory to analyze network and compute centrality indices; 3) evaluate the accuracy of the network and 4) community detection using Spinglass algorithm. Furthermore, we employed Network Comparison Test (NCT) to compare the estimated networks.

***Estimating networks*:**

A regularized Gaussian graphical model (27) which utilizes graphical LASSO (28) in combination with tuning parameter selected by minimizing the Extended Bayesian Information Criterion (29) was used to estimate the networks. The network plot structures were generated with nodes representing each of the individual genes and edges, their corresponding partial correlation coefficient values. Further, in the

network plots, the width of the edges represented the strength of the connections; meanwhile, the green or red edges illustrate positive or negative partial correlation values, respectively.

***Computing centrality indices*:**

We computed three different centrality indices of closeness, betweenness and strength. Closeness is the mean length of connected edges, suggesting how likely the given node influenced the whole network structure. Betweenness was the value of how many times a node lies on the edges of two other nodes, indicating the effect of the given node on the information flow of the whole network. The strength of the nodes was the sum of the weights of the connected edges (30, 31).Accordingly, perturbations to the nodes with the highest closeness and betweenness may affect large parts of the network structure, and perturbations to the node with the highest strength might influence many other nodes and are therefore considered most important within the network structure.

***Accuracy test*:**

The accuracy test assesses the stability of the networks and centrality indices by performing bootstrapped difference tests between edge-weights and the stability of the centrality indices in the network structure. These methods include a) estimation of the accuracy of the edge-weights, by drawing bootstrapped confidence intervals; b) investigating the stability of (the order of) centrality indices after observing only portions of the data; and c) performing bootstrapped difference tests between edge-weights and centrality indices to test whether these differ significantly from each other (26).

***Network comparison test*:**

The overall connectivity (or global strength) of the networks, defined as the weighted sum of the absolute connections (32), was determined for cases and controls. We performed statistical assessment of the difference in overall connectivity between networks of both groups network comparison test (NCT) (33). The NCT is a two-tailed permutation test in which the difference between two groups (cases and controls) is calculated repeatedly (100,000 times) for randomly regrouped genes. This results in a distribution under the null hypothesis (if both groups are equal), which can be used to test the observed difference between the empirical groups. The observed difference is considered significant at the threshold of .05.

***Community detection between cases and controls*:**

We used the spin-glass model to find communities in the network graphs between cases and controls. Spinglass algorithm is a community detection method based on the Potts model (34), which maps a network onto a zero-temperature *K*-Potts model with nearest neighbor interactions. The Potts model is a system of spins that can be in *K* different states, optimizing an energy function. For the problem of

community detection, the spin states are the group labels of the nodes, and the energy of the spin system is the quality function of the communities (35). The quality function tries to find communities in the network graph. A community is a set of nodes with many edges inside the community and few edges outside it (i.e., between the community itself and the rest of the network graph) (36).

Results

The study included a total of 32 HIV-infected and HIV-uninfected participants. Seventy-eight percent of the participants were whites (n=25), and the majority of them were males (n=22, 69%) (21). The median age of the study participants was 49.5 years (IQR=33-66). In this study, we applied a maximum relevance minimum redundancy method to rank the importance of the 26 genes which were differentially expressed between the two groups. DFFA was the most relevant (positive score) and TNFRSF1A (redundant, least negative score), as shown in Figure 1A. DFFA is a proapoptotic gene in the executioner pathway, and TNFRSF1A is a proapoptotic gene in the extrinsic pathway. To assess the discriminatory power of DFFA and TNFRSF1A, we then tested two different classifier models (Elastic-Net and KNN) to classify study participants based on these two selected genes into groups. Once we fit the elastic-net model, we used it to predict the outcome (case or control). Figure 1B shows that using all the 26 genes, we could predict the outcome with 80% accuracy. However, even with a reduced set of genes (either 2 (DFFA and TNFRSF1A or 5 (DFFA, TNFRSF1A, BCL2, FASLG, TNF) or 10 (DFFA, TNFRSF1A, BCL2, FASLG, TNF, CASP14, CASP7, TRAF2, CYCS, LTBR) or 20 (Figure 1B)), we still achieved 60-80% accuracy. For the KNN algorithm predictive accuracy, Figure 1B also shows that using all the 26 genes, we could predict the outcome (case or control) with 80% accuracy. Like elastic-net testing, we tested a reduced number of features and achieved 90% accuracy with the KNN classifier using only two genes–the minimum redundant (TNFRSF1A) and maximal relevant (DFFA) genes. The KNN classifier model using the two-top ranked mRMR genes correctly classified 90% of the participants into their respective groups.

Estimating Networks

We estimated a GLASSO network plot of all the genes, with edges connected by partial correlation values as shown in Figure 2. Of the 26 genes, 18 were proapoptotic (TNFRSF1A, CYCS, DFFA, ABL1, LTBR, CASP7, FASLG, BAD, TRAF2, BAK1, CIDEA, TNFRSF11B, CASP14, BIK, GADD45A, CASP5, CD70, and TNFRSF9), 5 were antiapoptotic (BCL2, BRAF, BIRC5, IL-10, and NOL3), and 3 had dual functions (CD27, HRK, and TNF). We also estimated the networks using GLASSO separately for cases and controls (Figure 3). We obtained centrality measures and assessed the stability of networks for cases and controls.

Cases

The case network structure (Figure 3) showed the strongest positive edge-weights between DFFA and TNFRSF1A, CYCS and BCL2 and negative edge-weights between BCL2 and TNFRSF1A, BCL2 and FASLG, and FASLG and CIDEA. The accuracy of connections was evaluated by bootstrapped CIs analysis. The

bootstrapped CIs revealed large CIs for the estimated edge-weights, suggesting that many of the edge-weights did not differ significantly from one another. However, CIs for the edges of CD27 and FASLG, and FASLG and TNFRSF1A did not overlap with bootstrapped CIs of other edges and were likely the strongest edges. As we decreased the sample size, stability was reduced. Centrality indices results revealed that FASLG had the highest strength, betweenness, and closeness (Figure 4, red) among all 26 genes analyzed, suggesting that FASLG had most interactions with other genes.

Controls

As shown, Figure 3 is the controls network. LTBR and TNF, DFFA and ABL1, HRK and CASP7, BCL2 and BAD, CYCS and BCL2 had strong edge-weights, with weak negative edge-weights found between TNF and ABL1, LTBR and CD70, CYCS and TRAF2 and CYCS and LTBR. The edge-weight accuracy results revealed that most of the edge-weights did not differ significantly from one another. With a decrease in sample size, strength was unstable. Centrality indices plot (Figure 4, teal) and centrality scores showed that ABL1 was the most central variable in the controls network.

Network comparison

To further analyze the overall differences between the two networks, a network comparison test was performed to examine the differences in the weights of connection. Ninety-two out of 325 connections differed significantly between networks. In addition, the highest strength centrality in controls, the ABL1 gene linked to TNFRAS1A, CASP14, TRAF2, CASP5, TNFRSF9 and DFFA were significantly different ($p<0.005$) in the two networks. The highest strength centrality in cases, the FASLG linked to TNFRS1A was the only significant edge ($p<0.001$) between the two networks. The paired t-test revealed that the global strength was significantly different between the two networks ($p < 0.05$), and the controls network had more significant edge-weights between nodes compared to the cases.

Community Detection between Cases and Controls

We explored a network model-based clustering using the Spinglass algorithm separately for cases and controls, as shown in Figure 5. The algorithm identified 3 clusters in cases and 5 clusters in controls. For cases, DFFA, TNFRSF1A, BCL2, CYCS, and ABL1 were in cluster one (blue in Fig 5), FASLG, BRAF, NOL3, IL10, CD70, TNFRSF9, CASP5, CD27, LTBR, TRAF2, CASP7, and TNF belonged to cluster two (green in Fig 5) and CASP14, HRK, GADD45A, BIRC5, BAK1, BIK, BAD, TNFRS11B, and CIDEA belonged to cluster three (light red in Fig 5). For controls, DFFA, FASLG, TRAF, and ABL1 belonged to cluster one (orange in Fig 5), CASP14, IL10, CD70, BIRC5, TNFRSF9, BIK, TNFRSF11B, CIDEA, CASP5, and CD27 were in cluster two (light red in Figure 5), TNF, BRAF, HRK, BAK1, LTBR, and CASP7 were in cluster three (green in Figure 5), TNFRSF1A, BAD, CYCS, and BCL2 were in cluster four (blue in Figure 5) and NOL3 and GADD45A were in cluster five (yellow in Figure 5).

Discussion

In this secondary analysis, we used 26 genes that we previously identified to be differentially expressed genes of the apoptosis pathway among HIV infected with toxicity and HIV uninfected participants. We applied a maximum relevance minimum redundancy algorithm on the 26 genes to rank gene importance. We implemented two different machine learning classifiers and tested the accuracy of prediction of cases or controls. DFFA and TNFRSF1A, two of the upregulated proapoptotic genes, classified 90% of study participants correctly. We previously used a penalized regression analysis to select the best subset of genes which contributed to profile differences. We previously also developed a classifier model to classify study participants into groups based on these two selected genes. The classifier model correctly classified 75% of the participants into their respective groups (21). However, using machine learning classifier algorithms (elastic net and KNN) we achieved 90% prediction accuracy using top two mRMR genes—DFFA and TNFRSF1A.

Since apoptosis is associated with almost all classes of ART, genes of the apoptotic pathway could serve as biomarkers for identifying and monitoring HIV treatment-experienced with ART-associated toxicity. Currently, there is no gold standard for diagnosing ART-induced mitochondrial toxicity. Diagnosis is based on a combination of clinical symptoms, laboratory testing, imaging studies, and if available a tissue biopsy to confirm mitochondrial damage (35). Confirmatory tissue biopsies are expensive, invasive and not readily available. The use of the differentially expressed apoptotic genes could provide accurate diagnosis of toxicity and eliminate the “trial and error” approach of switching around medications to relieve toxicity. Trial and error approach is expensive in the long run, as it favors the emergence of drug-resistant strains of HIV (37). There is a need for a non-invasive, cost-effective biomarker for ART-induced mitochondrial toxicity to prevent unnecessary interruptions in ART and to guide the use of second-line regimens. If we could validate our findings in a larger cohort, quantitative PCR assay of these apoptotic genes could serve as biomarkers for ART-induced toxicity.

Our network analysis findings of the proapoptotic FASLG gene being highly influential, due to its high centrality in cases, concur with several other physiologic studies that have found increased expression of FASLG in cases (37-39). However, instead of FASLG, ABL1 had highest centrality in controls. We also observed that the number of stronger edges (higher edge-weights) were lower in cases compared to the controls, suggesting that perturbations to the genes in cases network structure incapable of affecting multiple nodes (genes), unlike in control individuals. Whether this suggests that HIV alters the protective dependent network structure of genes in control requires further testing. We explored how genes were related to each other within the clusters of the Spinglass algorithms separately for cases and controls. For cases, the two most important proapoptotic genes (DFFA, TNFRSF1A) selected by mRMR belonged to cluster 1, while they were separated in the controls network structure. That the two most important genes belonged to the same community (cluster 1) in cases along with a strong positive edge between two other proapoptic genes (CYCS and BCL2) warrants further exploration, if these two genes should also be evaluated as the next step in an algorithm evaluating HIV case status.

In homeostasis, genes of the apoptotic pathway (pro- and anti-apoptotic) work in tandem (40). It is therefore interesting that we found a deferential clustering of pro- and anti-apoptotic genes between cases and control. Moreover, among the cases, we found less clustering compared to controls. This might suggest that cases with mitochondrial toxicity have perturbation of the apoptotic pathway favoring apoptosis. FASLG and ABL1, both proapoptotic genes had the highest strength, betweenness, and closeness in cases and controls, respectively. FASLG is a member of a family of proteins that signals the initiation of caspase cascade—a series of steps that result in apoptosis. This signaling is common to both extrinsic and intrinsic apoptotic pathway. Thus, the high degree of apoptosis in cases may be due to deploying of both extrinsic and intrinsic pathways. In response to oxidative stress, ABL1 targets to the mitochondria and mediates mitochondrial dysfunction and apoptosis. The role of FASLG in apoptosis may be more global than the role of ABL1 in apoptosis. The study has several limitations. First, as a cross-sectional study, we do not know the dynamic changes of these genes before and during ART. Second, we did not have a control group of HIV treatment-naive or HIV treatment-experienced individuals without toxicity. Third, the small sample size of study participants did not allow us to obtain patient-specific risks for the upregulation of these genes.

Conclusions

This is a case-control study of the effect of antiretroviral therapy (ART) on apoptosis pathway genes comprising of participants who were HIV infected with mitochondrial toxicity and uninfected controls. We applied the maximum relevance minimum redundancy (mRMR) algorithm on the genes that were differentially expressed between the cases and controls. The mRMR algorithm ranked DFFA and TNFRSF1A, two of the upregulated proapoptotic genes, on the top. The overall prediction accuracy was 90%, which clearly show that the mRMR gene sets outperforms the performance of the gene sets based on gene expression analyses. Network analysis revealed that FASLG had highest centrality in cases with ABL1 in controls, with a new correlation of genes associated with mitochondrial toxicity. Our findings are consistent with other studies that suggest apoptosis may be a critical step in ART-associated toxicity. Future studies should validate the use of apoptotic genes, particularly DFFA and TNFRSF1A, as biomarkers of ART-associated toxicity.

Abbreviations

HIV: Human Immunodeficiency Virus; IQR: interquartile range; mRMR: maximum relevance minimum redundancy; ART: antiretroviral therapy; PLWH: people living with HIV; NRTIs: nucleoside reverse transcriptase inhibitors; mtDNA: mitochondrial DNA; Pol-γ: polymerase gamma; PIs: protease inhibitors; NNRTIs: non-nucleoside reverse transcriptase inhibitors; KNN: k-nearest neighbors; NCT: network comparison test; LASSO: least absolute shrinkage and selection operator.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Yale University Human Investigation Committee. All participants gave their written informed consent before participation in the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are available in the BMC Genomics Dataset

repository, https://ragon.partners.org/musiebiostats/publications.html.

Competing interests

Elijah Paintsil and Musie Ghebremichael are editorial board members of BMC Infectious Diseases journal. Other authors do not have a commercial or other association that might pose a conflict of interest, i.e., authors declare that they have no competing interests to disclose.

Funding

The study was supported by grants from Harvard University Center for AIDS Research (HU CFAR NIH/NAIDS P30-AI 060354) and the Ragon Institute of MGH, MIT and Harvard. The funding body had no role in the design of the study; the collection, analyses, and interpretation of data, and in writing manuscript.

Acknowledgements

The authors would like to thank the subjects, all the providers and staff who participated in the study and for making the study possible.

Authors' contributions

E.B. analyzed data and was a major contributor in manuscript writing. E.P. collected data and wrote manuscript. M.G. conceived the study; analyzed data; wrote manuscript and provided guidance on data analyses and interpretation of the findings. All authors read and approved the final manuscript.

References

1. Palella Jr FJ, Baker RK, Moorman AC, Chmiel JS, Wood KC, Brooks JT, et al. Mortality in the highly active antiretroviral therapy era: changing causes of death and disease in the HIV outpatient study. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2006;43(1):27-34.
2. Palella Jr FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. New England Journal of Medicine. 1998;338(13):853-60.
3. Fang C, Chang Y, Hsu H, Twu S, Chen K-T, Lin C, et al. Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. Journal of the Association of Physicians. 2007;100(2):97-105.
4. Egger M, May M, Chêne G, Phillips AN, Ledergerber B, Dabis F, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. The Lancet. 2002;360(9327):119-29.
5. Gonzalez-Serna A, Chan K, Yip B, Chau W, McGovern R, Samji H, et al. Temporal trends in the discontinuation of first-line antiretroviral therapy. Journal of Antimicrobial Chemotherapy. 2014;69(8):2202-9.
6. Cicconi P, Cozzi-Lepri A, Castagna A, Trecarichi E, Antinori A, Gatti F, et al. Insights into reasons for discontinuation according to year of starting first regimen of highly active antiretroviral therapy in a cohort of antiretroviral-naive patients. HIV medicine. 2010;11(2):104-13.
7. d'Arminio Monforte A, Lorenzini P, Cozzi-Lepri A, Mussini C, Castagna A, Baldelli F, et al. Durability and tolerability of first-line regimens including two nucleoside reverse transcriptase inhibitors and raltegravir or ritonavir boosted-atazanavir or-darunavir: data from the ICONA Cohort. HIV Clinical Trials. 2018;19(2):52-60.
8. Di Biagio A, Cozzi-Lepri A, Prinapori R, Angarano G, Gori A, Quirino T, et al. Discontinuation of initial antiretroviral therapy in clinical practice: moving toward individualized therapy. Journal of acquired immune deficiency syndromes (1999). 2016;71(3):263.
9. Gertner E, Thurn JR, Williams DN, Simpson M, Balfour HH, Rhame F, et al. Zidovudine-associated myopathy. The American journal of medicine. 1989;86(6):814-8.
10. Brinkman K, ter Hofstede HJ, Burger DM, Smeitink JA, Koopmans PP. Adverse effects of reverse transcriptase inhibitors: mitochondrial toxicity as common pathway. Aids. 1998;12(14):1735-44.
11. Montaner JS, Côté HC, Harris M, Hogg RS, Yip B, Harrigan PR, et al. Nucleoside-related mitochondrial toxicity among HIV-infected patients receiving antiretroviral therapy: insights from the evaluation of venous lactic acid and peripheral blood mitochondrial DNA. Clinical infectious diseases. 2004;38(Supplement_2):S73-S9.
12. Moyle G. Clinical manifestations and management of antiretroviral nucleoside analog-related mitochondrial toxicity. Clinical therapeutics. 2000;22(8):911-36.
13. Lewis W, Dalakas MC. Mitochondrial toxicity of antiviral drugs. Nature medicine. 1995;1(5):417-22.

14. Apostolova N, Gomez-Sucerquia LJ, Gortat A, Blas-Garcia A, Esplugues JV. Autophagy as a rescue mechanism in efavirenz-induced mitochondrial dysfunction: a lesson from hepatic cells. Autophagy. 2011;7(11):1402-4.
15. Apostolova N, Blas-García A, Esplugues JV. Mitochondrial interference by anti-HIV drugs: mechanisms beyond Pol-γ inhibition. Trends in pharmacological sciences. 2011;32(12):715-25.
16. Blas-Garcia A, V Esplugues J, Apostolova N. Twenty years of HIV-1 non-nucleoside reverse transcriptase inhibitors: time to reevaluate their toxicity. Current medicinal chemistry. 2011;18(14):2186-95.
17. Blas-García A, Apostolova N, Ballesteros D, Monleon D, Morales JM, Rocha M, et al. Inhibition of mitochondrial function by efavirenz increases lipid content in hepatic cells. Hepatology. 2010;52(1):115-25.
18. Deng W, Baki L, Yin J, Zhou H, Baumgarten CM. HIV protease inhibitors elicit volume-sensitive Cl– current in cardiac myocytes via mitochondrial ROS. Journal of molecular and cellular cardiology. 2010;49(5):746-52.
19. Vlahakis SR, Bennett SA, Whitehead SN, Badley AD. HIV protease inhibitors modulate apoptosis signaling in vitro and in vivo. Apoptosis. 2007;12(5):969-77.
20. Karamchand L, Dawood H, Chuturgoon AA. Lymphocyte mitochondrial depolarization and apoptosis in HIV-1-infected HAART patients. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2008;48(4):381-8.
21. Foli Y, Ghebremichael M, Li M, Paintsil E. Upregulation of apoptosis pathway genes in peripheral blood mononuclear cells of HIV-infected individuals with antiretroviral therapy-associated mitochondrial toxicity. Antimicrobial agents and chemotherapy. 2017;61(8).
22. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC bioinformatics. 2017;18(1):1-14.
23. Kuhn M, Johnson K. Applied predictive modeling: Springer; 2013.
24. Lee JS, Paintsil E, Gopalakrishnan V, Ghebremichael M. A comparison of machine learning techniques for classification of HIV patients with antiretroviral therapy-induced mitochondrial toxicity from those without mitochondrial toxicity. BMC medical research methodology. 2019;19(1):1-10.
25. Chomboon K, Chujai P, Teerarassamee P, Kerdprasop K, Kerdprasop N, editors. An empirical study of distance metrics for k-nearest neighbor algorithm. Proceedings of the 3rd international conference on industrial application engineering; 2015.
26. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behavior Research Methods. 2018;50(1):195-212.
27. Costantini G, Epskamp S, Borsboom D, Perugini M, Mõttus R, Waldorp LJ, et al. State of the aRt personality research: A tutorial on network analysis of personality data in R. Journal of Research in Personality. 2015;54:13-29.
28. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9(3):432-41.

29. Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. 2008;95(3):759-71.
30. Dalege J, Borsboom D, van Harreveld F, van der Maas HL. Network analysis on attitudes: A brief tutorial. Social psychological and personality science. 2017;8(5):528-37.
31. Robinaugh DJ, Millner AJ, McNally RJ. Identifying highly influential nodes in the complicated grief network. Journal of abnormal psychology. 2016;125(6):747.
32. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proceedings of the national academy of sciences. 2004;101(11):3747-52.
33. Van Borkulo C, Epskamp S, Millner A. Network comparison test: permutation-based test of differences in strength of networks. 2015.
34. Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical review E. 2006;74(1):016110.
35. Hoffman M, Steinley D, Gates KM, Prinstein MJ, Brusco MJ. Detecting clusters/communities in social networks. Multivariate behavioral research. 2018;53(1):57-73.
36. Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical review E. 2004;69(2):026113.
37. Gehri R, Hahn S, Rothen M, Steuerwald M, Nuesch R, Erb P. The Fas receptor in HIV infection: expression on peripheral blood lymphocytes and role in the depletion of T cells. Aids. 1996;10(1):9-16.
38. Sloand EM, Young NS, Kumar P, Weichold FF, Sato T, Maciejewski JP. Role of Fas ligand and receptor in the mechanism of T-cell depletion in acquired immunodeficiency syndrome: effect on CD4+ lymphocyte depletion and human immunodeficiency virus replication. Blood, The Journal of the American Society of Hematology. 1997;89(4):1357-63.
39. Badley AD, McElhinny JA, Leibson PJ, Lynch DH, Alderson MR, Paya CV. Upregulation of Fas ligand expression by human immunodeficiency virus in human macrophages mediates apoptosis of uninfected T lymphocytes. Journal of virology. 1996;70(1):199-206.
40. Packham G, Stevenson FK. Bodyguards and assassins: Bcl-2 family proteins and apoptosis control in chronic lymphocytic leukaemia. Immunology. 2005;114(4):441-9.

Figures

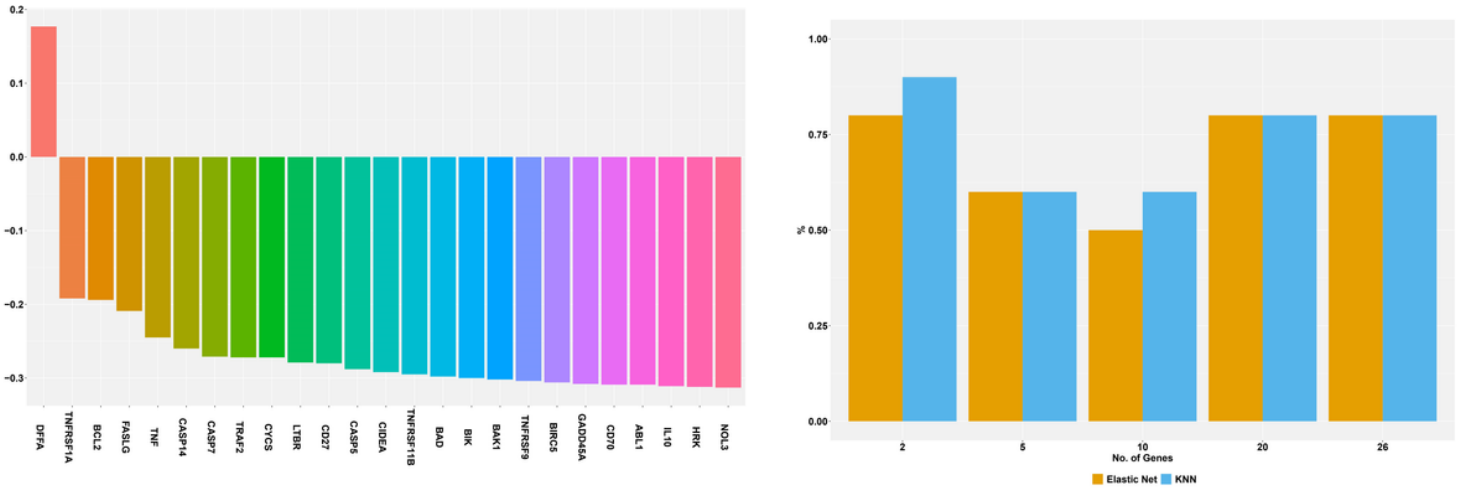

Figure 1

A: Gene Importance Scores. DFFA (most relevant) and TNFRS1A (redundant) were the top two genes. The rest are shown in decreasing order of gene importance B: Classification accuracy using two classifiers for the top genes

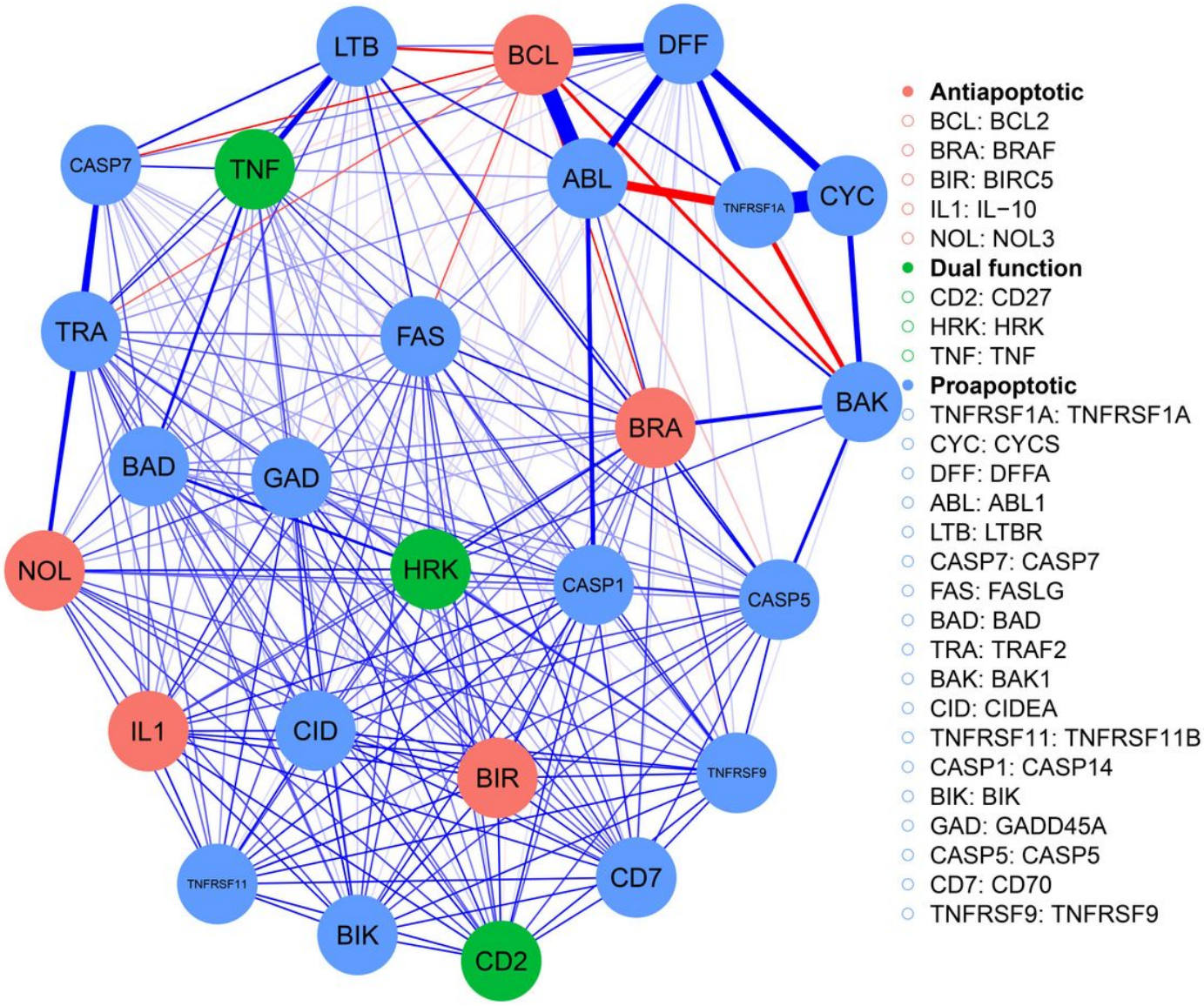

Figure 2

Network plot on the entire dataset. The circles represent nodes and the lines connecting them indicate edges, which are the partial correlation values between the nodes. Blue indicates positive and red indicates negative correlation value. The nodes were specified to show which were antiapoptotic (red), proapoptotic (blue) and dual function (green).

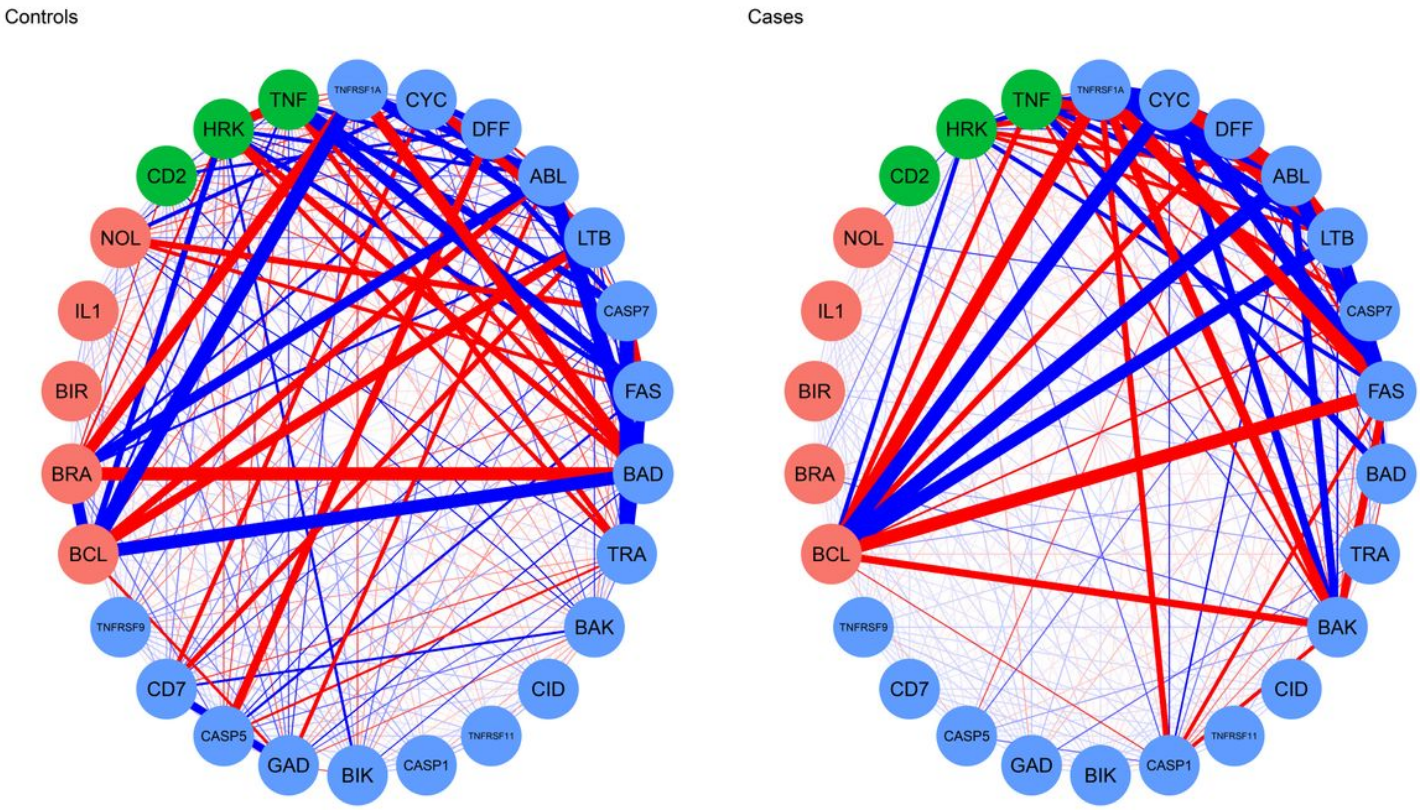

Figure 3

Network plots for cases and controls separately. Green lines indicate positive partial correlation values and red lines indicate negative partial correlation values. The thickness of the edges indicates the strength of the correlation.

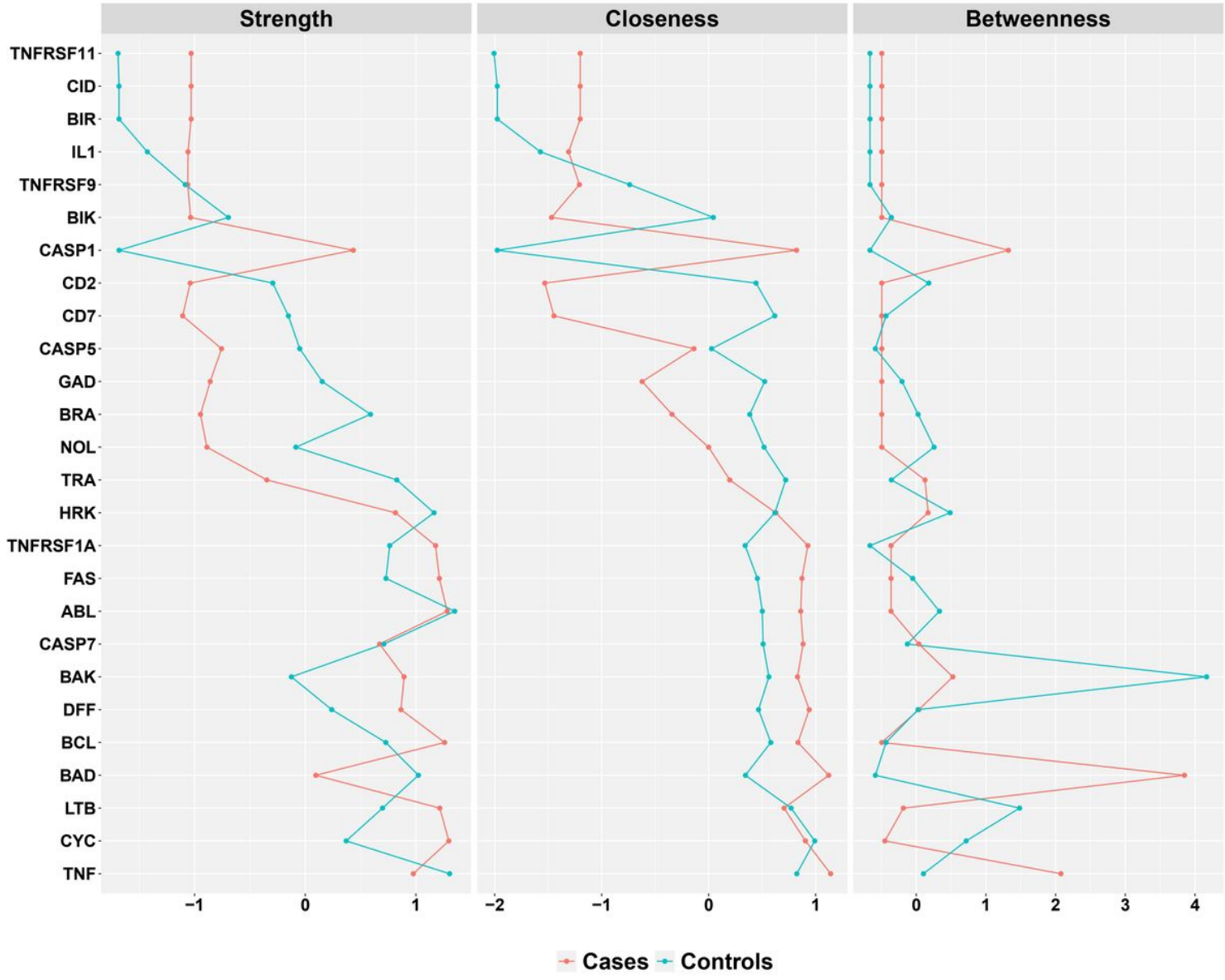

Figure 4

Centrality Plots. Cases are shown in red and controls in teal. FASLG had the highest strength, betweenness and closeness in cases while ABL had the highest strength, betweenness and closeness in controls

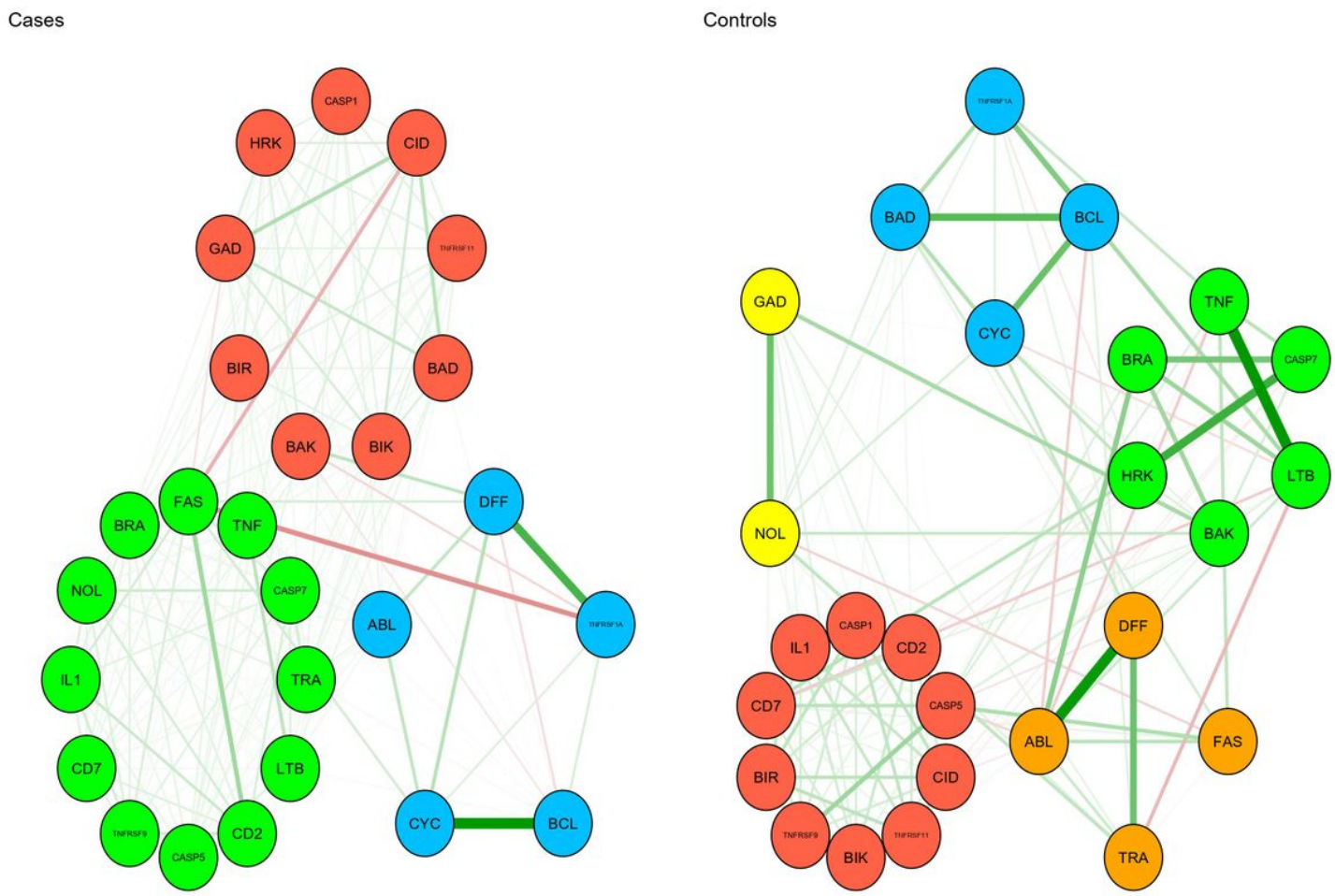

Figure 5

Spinglass algorithms showing clusters for cases and controls. There were 3 clusters for cases and 5 clusters in controls