2020

HIV-Associated Neurocognitive Disorder (HAND) Biomarker Identification: Significance Analysis of Microarrays and Two Persuasive Approaches with Random Forest

Hansapani Rodrigo
The University of Texas Rio Grande Valley

Bryan Martinez
The University of Texas Rio Grande Valley

Roberto De La Garza
The University of Texas Rio Grande Valley

Upal Roy
The University of Texas Rio Grande Valley

Recommended Citation

Hansapani Rodrigo, Bryan Martinez, Roberto De La Garza et al. HIV-Associated Neurocognitive Disorder (HAND) Biomarker Identification: Significance Analysis of Microarrays and Two Persuasive Approaches with Random Forest, 26 March 2020, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-19489/v1]

HIV-Associated Neurocognitive Disorder (HAND) Biomarker Identification: Significance Analysis of Microarrays and Two Persuasive Approaches with Random Forest

CURRENT STATUS: POSTED

Research Square

Hansapani Rodrigo
University of Texas Rio Grande Valley

hansapani.rodrigo@utrgv.edu*Corresponding Author*
ORCiD: https://orcid.org/0000-0002-2122-2150

Bryan Martinez
University of Texas Rio Grande Valley

Roberto De La Garza
University of Texas Rio Grande Valley

Upal Roy
University of Texas Rio Grande Valley

DOI:
10.21203/rs.3.rs-19489/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
HIV-Associated Neurocognitive Disorder, Biomarker Identification Approaches, Random Forest Analysis, Microarray Data

Abstract

Background: HIV Associated Neurological Disorders (HAND) is relatively common among people with HIV-1 infection, even those taking combined antiretroviral treatment (cART). Genome-wide screening of transcription regulation in brain tissue helps in identifying substantial abnormalities present in patients' gene transcripts and to discover possible biomarkers for HAND. This study explores the possibility of identifying differentially expressed (DE) genes, which can serve as potential biomarkers to detect HAND. In this study, we have investigated the gene expression levels of three subject groups with different impairment levels of HAND along with a control group in three distinct brain sectors: white matter, frontal cortex, and basal ganglia. Methods: Linear models with weighted least squares along with Benjamini-Hochberg multiple corrections were used to identify DE genes in each brain region. Genes with an adjusted p-value of less than 0.01 were identified as differentially expressed. Principal component analyses (PCA) were performed to detect any groupings among the subject groups. Significance Analysis of Microarrays (SAM) and random forests (RF) methods with two distinct approaches were used to identify DE genes. Results: A total of 710 genes in basal ganglia, 794 genes in the frontal cortex, and 1481 genes in white matter were screened. The highest proportion of DE genes was observed within the two brain regions, frontal neocortex, and basal ganglia. PCA analyses do not exhibit clear groupings among four subject groups. SAM and RF models reveal the genes, CIRBP, RBM3, GPNMB, ISG15, IFIT6, IFI6, and IFIT3, to have DE genes in the frontal cortex or basal ganglia among the subject groups. The gene, GADD45A, a protein-coding gene whose transcript levels tend to increase with stressful growth arrest conditions, was consistently ranked among the top genes by both RF models within the frontal cortex. Conclusions: Our study contributes to a comprehensive understanding of the gene expression levels of the subject with different severity levels of HAND. Several genes that appear to play critical roles in the inflammatory response have been found, and they have an excellent potential to be used as biomarkers to detect HAND under further investigations.

Introduction

Human Immunodeficiency Virus type 1 (HIV-1) is a potent pathogen that has affected many

individuals worldwide and is considered one of the deadliest diseases in the world (Karlsson, Kwiatkowski and Sabeti, 2014). The mode of action of HIV-1 is through the infection of helper T-cells (CD4+) by the use of glycoproteins gp120 and gp41 that aid in the fusion of HIV-1 with the cells. Helper T-cells are responsible for many immunological functions, one being the activation of the adaptive immune response. Once HIV-1 has infected and replicated within most of the CD4 + cells, the immune system becomes vulnerable, and the individual faces a substantial degree of infection (Campbell and Hope, 2015).

It has been observed that individuals with HIV-1 have a higher possibility of developing HIV-associated neurocognitive disorders (HAND) at later stages of HIV-1 infection. When the virus has made its way to the brain, it begins replicating within the microglia of the brain and eventually causing neuronal damage that gives the onset of the symptoms associated with HAND (Garcia-Mesa and others, 2017). The most severe form of this neurocognitive impairment is known as HIV-associated dementia (HAD). The pathogenesis of HAD has been known to elicit a neurological inflammatory response, which has also been recognized as HIV encephalitis (HIVE) (Fischer and others, 2014). Studies have suggested that the HIV-1 replication alone itself does not cause neuronal damage, but the inflammatory response HIVE is responsible for this, based on the elevated levels of inflammatory cytokines and chemokines (Shityakov, Dandekar and Forster, 2015).

There has been a significant advancement in the diagnosis and treatment of individuals infected with HIV-1, along with a decrease in the cost of treatment, making it more affordable for individuals living with this infection (Danforth and others, 2017). The cART has shown effective outcomes against the prevalence of HAD and HIVE. Nevertheless, individuals receiving cART show signs of mild forms of HAND, including HIVE (Ghafouri and others, 2006, Spudich, 2016).

Many studies have focused on investigating brain tissue samples in order to identify the key pathophysiological features of the patients with different severity levels of HAND (HAD and HIVE). These studies have found a significant amount of abnormal genomic differences within the frontal neocortex, mostly among the patients with HIVE (Everall and others, 2005, Gelman and others, 2012, Masliah and others, 2004, Roberts, Masliah and Fox, 2004, Roberts and others, 2003). Interferon

response genes (IFRG), which are related to the cell functions in the immune system and autoimmunity, were observed to be differentially expressed (Everall and others, 2005, Gelman and others, 2012, Masliah and others, 2004). Despite the fact that many of these studies have focused on the people with HIVE and a single brain region (mostly neocortex), Gelman and others, have explored the differences in the gene expression values within three brain regions among healthy controls, HIV-1 infected subjects and two other subjects' groups with different levels of HAND, namely HAD and HIVE.

We have used the gene expression data derived from the same brain specimen samples as Gelman and others, 2012. Two of our main objectives were similar to theirs; (1) identifying the genomic differences among the subjects with HIVE and HAD and (2) evaluating the functional behavior changes within the three brain regions are similar in both our studies. In addition to those two, we also want to (3) identify the genomic differences between HAD and HIV subjects and (4) evaluate the effectiveness between the two random forest models which we have used for our analysis. In order to accomplish our goals, we have utilized the Significance Analysis of Microarrays (SAM) and two random forest-based biomarker identification methods.

SAM is a popular nonparametric method introduced by Tusher et al. (Tusher, Tibshirani and Chu, 2001) in order to identify DE genes among different subject groups. This has the capacity to identify DE genes after controlling for false discovery rates in large as well as in small samples (Chu and others, 2005).

Data mining is an effective technique that has been used in the field of gene expression analysis in recent years (Aouf, Liyanage and Hansen, 2008, Chen and Ishwaran, 2012). A typical microarray experiment includes only tens or hundreds of patients' samples and explores thousands of genes leading to a small 'n' a large 'p' problem (Slonim, 2002). Usually, random forest (RF) models are well capable of handling these high dimensional gene expression data (Chen and Ishwaran, 2012, Qi, 2012). One of the key challenges in gene expression data modeling is how to account for their natural interactions (Aschard and others, 2012). It is important to weigh the co-occurrence between genes that are correlated with each other. One of the effective solutions to this problem is gene clustering.

In fact, clustering has been extensively used as an approach for detecting interesting patterns and potential biomarkers in genetic data. A previous study has explored the efficiency of a cluster ensemble approach using RF-derived proximity measures in determining the underlying structure of genetic data (Karpievitch and others, 2009). They have introduced a new random forest-based algorithm called “RF++” using a subject-level bootstrapping resampling.

The use of gene clusters that preserve their co-interaction would be an appealing approach. Therefore, we have focused on developing an RF-based classification and biomarker identification models based on the gene expression network analysis (GXNA) introduced by Nacu et al. (Nacu and others, 2007) GXNA forms gene interaction subnetworks and ranked them according to an efficient scoring function. These subnetworks are created based on the known pathways and hence employ prior biological information. In our work, we present details of two RF methods, one with regular RF (using single genes as features) and a cluster-based RF approach using high scoring gene interaction subnetworks as clusters.

Materials And Methods

Subjects

The gene expression data for our study was taken from the Gene Expression Omnibus (GEO) database, which is maintained by the National Center for Biotechnology Information (Clough and Barrett, 2016). The original gene expression data were collected by Gelman and others, 2012 using an Affymetrix array platform and saved in GEO under the ID GSE35864. This contains the gene expression data related to 24 subjects where we had six controls without HIV-1 infection; another 6 HIV-1 infected subjects with none or slight neurocognitive impairment (HIV); 7 subjects with HIV-Associated dementia (HAD); and five subjects with both HIV-associated dementia and HIV encephalitis (HIVE). Each subject has brain specimen samples related to three different brain regions; white matter, basal ganglia, and frontal neocortex (72 brain specimen samples in total). As expected, subjects with HIVE had the lowest survival time with a mean age at death being 42.8 years. The mean age at death for subjects with HAD was 43.7 years, and for subjects with HIV, it was 49.5 years. The mean age at death for the control group was of 50.0 years. More details about the patients’

demographic characteristics can be found in Gelman and others, 2012.

Gene expression data were first normalized using the GCRMA (Wu and others, 2004). GCRMA adjusts for background intensities in Affymetrix array data, including optical noise and non-specific binding (Z.J and R, 2010). There were 20,419 genes after the removal of Affy control genes. A non-specific gene filtering technique was applied to each gene expression data derived from the three brain regions. More specifically, we filtered out the genes which had an unlogged normalized intensity greater than 100 in at least 20% of the samples and the genes which had their coefficient of variation (standard deviation/mean) between 0.7 and 10. This helps to retain the genes with enough variance to be informative in classification. After the filtration, we have screened 710 genes in basal ganglia, 794 genes in the frontal cortex, and 1481 genes in white matter. These filtered genes were used in the rest of the analyses. For the evaluation purposes of RF models, we have divided the 24 subjects into two subsets, a training set (70% of data) with 19 subjects and the remaining to a testing set (30% of data).

Linear Models and Principal Component Analysis

Filtered log2 gene expression data was utilized to fit linear models per each brain region using weighted least squares with empirical Bayes moderation of the standard errors (Smyth, 2004). This approach is well suited to identify DE genes that are non-normally distributed and when the variability of the expression values differs between genes. After applying the Benjamini-Hechberg correction for multiple testing comparisons, the genes with an adjusted p-value of less than 0.01 were identified as differentially expressed. Moreover, principal component analysis (PCA) was performed on the three brain sectors to identify the subjects with similar gene profiles.

Significance Analysis of Microarrays (SAM)

SAM (Tusher, Tibshirani and Chu, 2001) as used to identify the differentially expressed genes among the four subject groups; Control, HIV, HAD, and HIVE within each brain region. SAM helps to identify expression patterns that have little difference between the control and other subject groups, yet significant. This test has the advantage over regular t-test as it is a robust test that can be applied even when the samples are not independent and are not normally distributed. We used the “samr”

package in R (version 3.6.1) to perform the SAM analysis and report the SAM score (d), fold chang and the q-value of each gene. The q-value of a gene is the false discovery rate for the gene list that includes that gene and all other genes that are more significant (Chu and others, 2005). It is like the familiar “p-value,” adapted to the analysis of many genes. The q-value measures how significant the gene is: as $d > 0$ increases, the corresponding q-value decreases. In all our SAM analyses, differentially expressed genes with at least a two-fold change were identified within a 10% false discovery rate.

Random Forest Methods

Random Forest has been first introduced by Leo Breiman in 2001 and has emerged since then as an efficient algorithm capable of handling high-dimensional data (Breiman, 2001). This method utilized the two concepts called bagging (Breiman, 1996) and random subspace (Tin Kam, 1998). Bagging is the process of aggregating the results of multiple trees, where each tree is grown on a bootstrap sample of the subjects. Random subspace refers to the selection of a random subset of variables as candidates for splitting at each node. Rather than considering all variables as candidates for splitting, RFs considers only a subset of variables, thus reducing the correlation between trees. These features make RF models less vulnerable to overfitting problems (Slonim, 2002) compared to other decision tree models (Tin Kam, Hull and Srihari, 1994).

As mentioned earlier, a bootstrap sample (Efron and Tibshirani, 1993) of a specified size is drawn with replacement from the original data. Approximately one-third of the subjects are left out of the bootstrap sample and not used in the construction of the tree and this is used to calculate the “out-of-bag” (OOB) error. The final classification of a subject is determined by the majority of the voting over all trees in the forest.

In this study, we have utilized two different RF-based approaches to detect genomic differences among different subject groups. In the first method (hereafter, referred to as RF method 1), we directly utilized the filtered genes in creating RF models for each brain region. Each RF model created in this approach has 10,001 number of trees and the best tree model with the lowest OOB error was selected out of several RF trees created with varying number of genes at each split.

In the second approach, a new random forest model (hereafter, referred to as RF method 2) was developed to identify the potential biomarkers by utilizing the gene clusters, which is a collection of interrelated genes. We have identified the most informative clusters of genes using gene expression network analysis (GXNA) tool (Nacu and others, 2007). GXNA identifies optimal gene interaction networks based on a new scoring function called $T\Sigma$, which accounts for gene intercorrelations across subjects. This is an intuitive idea that was extended from Ideker et al. original work (Ideker and others, 2002). As real genes function in concert rather than acting alone. Each top-ranking subnetworks or clusters have been used to create RF models. Each RF model created in this approach also has a 10,001 number of trees.

Figure 1 summarizes the steps of both RF methods.

The best RF models per each brain region were selected from the two approaches and hence in total, we had 9 RF models (for all 3 brain regions; one RF model from method 1 and 2 RF models from method 2 using the top two gene clusters). Different evaluation criteria including accuracy, precision, and recall rates, were considered for the comparisons of RF methods. Accuracy is the ratio between the number of correct predictions made for each subject group to the total number of subjects. Precision is the number of correct subject group predictions divided by the number of total predictions made by the model for each subject group. A recall is the number of correct subject group predictions divided by the total number of subjects present in that group. We have also calculated two overall measurements called the macro precision and macro average for each RF model. This is the average of precision and the average recall rates among all four subject groups. These measurements were calculated using the testing set. However, due to the fewer number of subjects we had in both training and testing data sets, these model summary measurements are not that appealing. All statistical analyses were performed using R (version 3.6.1) and the GXNA tool (Nacu and others, 2007).

Results

Results of Linear Models and Principal Component Analyses

Table S1 (supplementary materials) has details related to all comparisons that we have made among

four subject groups (HAD vs. control, HIV vs. control, HIVE vs. control, HAD vs. HIV, HIVE vs. HAD, HIVE vs. HIV) across all three brain sectors; basal ganglia, frontal cortex, and white matter identified using linear models. The up-regulated genes among a subject group compared to the other group in Table S1 are denoted by “1”, down-regulated genes are denoted by “-1” and genes with no difference are denoted by “0”.

Figure 2 shows a representation of the DE genes among HIV, HAD and HIVE groups compared to controls within the three brain sectors. According to Fig. 2, subjects with HIVE had the most DE genes in all three brain sectors. Many of the DE genes among the HIVE group within the frontal cortex were downregulated, while many of them were upregulated in the white matter and basal ganglia. Subjects with HIV, but without either HAD or HIVE, had 11 upregulated genes common with HIVE subjects in the frontal cortex. The details of these 11 genes can be found in Table S1. Those are, IFI44L *, DTX3L, GBP1 *, IFIT3 *, AQP1, MX1 *, PLSCR1, RARRES3, SPARC, TIMP1, and SERPING1. Genes marked with * are interferon response genes (Boehm and others, 1997), which are related to the cell functions in the immune system and autoimmunity. As per above mentioned DE gene selection criterion, subjects with HAD did not exhibit any DE genes over control subjects across all three brain sectors. In fact, this makes it harder to detect genomic differences among HAD patients compared to healthy subjects. As per Table S1, CIRBP was the only gene which was down-regulated among HIVE compared to HAD subjects within basal ganglia, while BTN3A3 *, BTN3A2, IFIT1, IFIT2, IFIT3 *, MX1, PARP14, HERC5, STAT1, and ISG15 * were upregulated among HIVE compared to HAD within the same brain region (Refer Table S1 for a full list of upregulated genes). There were no down-regulated genes among HIVE compared to HIV in basal ganglia while BTN3A3 *, GPNMB, BTN3A2, HERC5, CKLF, PARP14, IFIH1, GUSBP11 were up-regulated among them. There were no up or down-regulated genes among HAD compared to HIV in basal ganglia CDH13, CDK7, CITED2, TBR1, CIRBP, GCA, CMAS, RBM3, CTXN3, SST were there among the 145 down-regulated genes while IRF9, BTN3A3 *, CPQ, GBP1, IFIT2, IFIT3 *, MX1, HERC5, PARP14, ISG15 * were there among 63 upregulated genes within HIVE subjects compared to HAD in frontal cortex. CDH13, CITED2, TBR1, ORC6, PARM1, C3orf80, RBM3, CTXN3,

SNRPB, SST are some of the genes which were down-regulated and RFT2, GAREM2, ATP10B, IGKC, LINC00320, FOXO4, LIN28A are some of the genes which were upregulated among HIVE compared HIV in the same brain region. None of the genes were up or downregulated among HAD compared to HIV.

The genes BTN3A3 *, GPNMB, BTN3A2 *, IFIT3 *, LY6E, ISG15 * were upregulated whereas LRRCC1 and OPALIN were downregulated within HIVE subjects compared to HAD within the white matter. Surprisingly, there were no up or down-regulated genes among HIVE or HAD compared to HIV within the white matter.

Figure 3 (a) represents the distribution of subjects within each brain region with respect to the corresponding top three principal components. No clear groupings were observed among the four subject groups. Varimax rotation of PCA (Kaiser, 1958) did not add many changes to the original PCA patterns among subjects, as seen in Fig. 3.b.

Results of SAM Analysis

Table S2 (supplementary materials) contains DE genes identified through the SAM analysis. Similar to the linear model analysis, six comparisons were made among four subject groups within each brain sector, and the corresponding SAM score (*d*), fold change, and the q-value of each DE gene were reported. Here, we focused on the DE genes among HIVE compared to both HAD and HIV and HAD compared to HIV subjects with the intention of finding possible biomarkers to detect HAD (and hence HAND) in its early stages.

There were no down-regulated genes among HIVE compared to HAD subjects within basal ganglia. In contrast, 58 genes were upregulated among HIVE compared to HAD with in the same brain sector. Some of those genes are, ISG15, BTN3A2, BTN3A3, GPNMB, IFIT3, IFIT1, IFI44, IFIH1, IFI44L, GBP1, STAT1, IFIT2, HERC5, HERC6, GBP2, MX1, GBP3, and IRF9. The only downregulated gene among HIVE and HIV was ZNF808. Some of the upregulated genes within those two subjects' groups were; IGKC, GUSBP11, GPNMB, RGS1, RCC2, GBP3, BTN3A3, BTN3A2, IFIH1, IFI44, IFIT1, ISG15, CD14, STAT1, HERC5, GBP1. The genes ZNF808, IFI6, IFIT3, ISG15, were upregulated between HAD compared to HIV within the same brain region while no down-regulated genes were found.

Within the frontal cortex, 143 genes were down-regulated while 45 genes were up-regulated among HIVE compared to HAD subjects. LOC105370687, PCYOX1L, WDR54, PCSK1, CDH13, DHRS11, ENPP5, NPTX2, ORC6, STAT4, SST, PWAR5, CIRBP, CITED2, and RBM3 were some of the down-regulated genes and IFIT2, SALL1, LGALS3BP, CPQ, GBP1, IFI44L, DTX3L, CYP2J2, PARP14, BST2, ISG15, BTN3A3, IFIT3, IRF9, STAT1, GBP3, FOXO4, IFIH1, MX1, and GPNMB were some of the up-regulated genes. UBE2T, PWAR5, and SST were the only downregulated and SALL1, CDK18 were the only upregulated genes among HIVE compared to HIV. BTN3A3, MX1, IFIT3, IFI44L, PLSCR1, STAT1, GBP1, IRF9, IFI6, GADD45A, GPNMB, BST2, ISG15 are some of the downregulated genes among HAD compared to HIV, and none of the genes were up regulated between those two subject groups. A comparison between within white matter showed 48 down-regulated and 26 up-regulated genes among HIVE compared to HAD subjects. OPALIN, CDR1, CIRBP, LOC105377335, GSTZ1, ALDH1A1, LINC01102, SEPT10, LRRCC1, ZNF808 were there within the set of down-regulated genes among HIVE subjects. GPNMB, BTN3A3, ISG15, LY6E, IFIT3, BTN3A2, CMPK2, STAT1, IFIT1, IFIH1, GBP2, MX1, HERC6 were there among the up-regulated genes within the same subjects. OPALIN, LRRCC1, SEPT10, RBP1, CDR1, ALDH1A1, CNN3, CIRBP were some of the down-regulated genes while GPNMB, ISG15, SEPHS2, BTN3A3, IFIT1, IFIH1, RCC2, GATB, BTN3A2 were some of the upregulated genes among HIVE compared to HIV within the white matter. HAPLN2, LOC101930085, PDE1A, LOC100128079, CHADL, ENPP6, GATB, RSRP1 are some of the upregulated genes among HAD compared to HIV while none of the genes were downregulated between them. Details of the full list of genes that were up and downregulated among different subject groups across three brain sectors can be found in Table S2.

Results of RF Method 1

Three best RF models created using RF method 1 for basal ganglia, frontal neocortex, and the white matter had OOB error rates 47.36%, 42.34%, and 47.36%, respectively. Due to the limited subjects that we had in our sample, it was not possible to reduce the OOB error rates more than these. Genes within each brain sector were ranked according to the mean decrease in accuracy and given in Table S3 (supplementary materials). The higher the mean decrease associated with a gene, the more important it is incorrectly predicting that subject group.

Figure 4 depicts the top 25 genes identified within basal ganglia using RF method 1. Table S3 contains the ranked genes as per the mean decrease in accuracy along with corresponding fold change and q values related to SAM analyses. This helps to assess the relative importance of each gene based on the RF method-1 and SAM between every two subject groups. Although our discussion here is focused on identifying DE genes among the subject groups HIVE vs HAD and HAD vs HIV, readers can refer Table S3 to find out any other important relationships among different subject groups.

As mentioned in Sec 3.2, there were no down-regulated genes within basal ganglia among HIVE compared to HAD. Among the top 25 genes identified by RF method 1, the genes IFIT2, IFIH1, LGALS3BP, HERC5, PSMB8, STAT1, BTN3A3, BTN3A2, MX1, and IRF9 were up-regulated at least by a 3-fold and had q values < 0.03 within HIVE compared to HAD subjects. Among those, BTN3A2, IFIH1, and STAT1 had at least a 6-fold change.

IFIT3, ISG15, IFI6, and ZNF80 were the only four genes that were downregulated HAD compared to HIV and none of them were ranked within the first 25 of RF method 1. Among these four, IFI6 and ZNF808 were significantly downregulated with a q value < 0.00. IFI6, a protein-coding gene that is related to cytokine signaling in the immune system pathway and it is associated with many diseases including Hepatitis C (Chen, Li and Chen, 2016, Qi and others, 2017) and Dengue virus-2 (Qi and others, 2015). This gene was not DE within HIVE and HIV subjects. Nevertheless, it was significantly up-regulated (fold change: 9.12 with a q value: 0.00) among HIVE compared to HAD and hence needed to be further investigated due to its potential to be used a biomarker to detect HAD patients in their early stages (This gene was ranked #69 according to the gene ranking in RF method 1). The top 25 genes identified by RF method 1 within the frontal cortex are given in Fig. 5. SST, which was ranked #7 by RF method 1 was significantly down-regulated with a q-value 0.00 among HIVE compared to HAD subjects. The genes, MX1, MAP4K4, CPQ, SALL1, LGALS3BP, PARP14, and IFIT2, were significantly up-regulated with q- values < 0.04 among them. In fact, MX1 and GADD45A were significantly down-regulated with q-values < 0.00 among HAD compared to HIV. Since GADD45A was down-regulated only between HAD and HIV, this need to be further investigated to be used as a potential biomarker.

The top 25 genes identified within white matter using RF method 1 is given in Fig. 6. BTN3A3, ISG15, IFIH1, MX1, and IFIT1 which were among the top 25 ranked genes, were significantly upregulated with at least 4 folds and q-values < 0.03 among HIVE compared to HAD subjects. Although not ranked within the first 25, RBP1, LRRCC1, OPALIN, CPAMD8, LOC100506114, CIRBP and RBM3 were there among the 48 down-regulated genes within HIVE compared to HAD. None of the other top 25 genes were significantly up or down-regulated (with a q-value < 0.05) within white matter among HIVE and HIV or HAD and HIV subjects. A set of genes within the first 100 rankings, LOC101930085, LOC100128079, HAPLN2, and ENPP6, had at least 2-fold change, but with q values < 0.09 (these are not < 0.05) among HAD compared to HIV subjects. Even though they were not significant at the 5% level, it might worth further examine them as they were not significantly up or downregulated within any other subject groups. For example, HAPLN2 has been identified to play a pivotal role in the formation of the hyaluronan-associated matrix in the central nervous system, which facilitates neuronal conduction and general structural stabilization within the white matter of mutant mice (Bekku and others, 2010). However, we did not find any detailed research on the other three genes.

Results of RF Method 2

Here, we present the results of our second random forest model based on the most informative clusters that we found out using the GXNA tool. The corresponding rankings of genes within each cluster are also given in Table S3. Note that, we present the results of two clusters for each brain sector.

Figure 7 depicts the genes expression pattern of basal ganglia given by RF method 2. The two graphs given here are related to the top 2 ranking gene subnetworks or the clusters that we have identified through GXNA. One cluster had eight genes, and the other one had just four genes. STAT1 and IRF9 within the first cluster and the LGALS3BP from the second cluster were significantly upregulated within the HIVE compared to HAD with q-values < 0.03. None of the genes within two clusters were DE both among HAD and HIVE compared to HIV under q-value < 0.05.

Figure 8 depicts the genes expression pattern of the frontal cortex given by RF method 2. FOX04 within the first cluster was significantly up-regulated with a q-value 0.04, and COL5A2 within the

second cluster was significantly down-regulated among HIVE compared to HAD within the frontal cortex. GADD45A within the first cluster was the only gene that is significantly down-regulated among HAD compared to HIV with a q-value 0.00. None of the genes within two clusters were DE among HIVE compared to HIV.

Figure 9 depicts the genes expression pattern of white matter given by the RF method 2. RBP1 was identified to be significantly down-regulated while, and MX1 was significantly up-regulated within the white matter with a q-values < 0.04 among HIVE compared to HAD. RBP1 also significantly downregulated among HIVE compared to HIV subjects. Surprisingly, the bottom cluster did not show any DE regulated genes among any of the subject group comparisons.

Model Validation

We evaluated all the random forest model that we developed using a testing data set based on model accuracy, macro precision, and macro recall rates. The testing data set consists of one control, one HIV, two HAD, and one HIVE subject making a total of five. Table 1 summarizes the model evaluations in terms of the OOB error rates of the training data set and accuracy, macro precision, and macro recall rates for the testing data set. The resulting RF models have reasonable OOB error rates.

Table 1
Random Forest Model Evaluations

Brain Region	Model	OOB Error Rate(%)	Macro Precision	Macro Recall	Accuracy
Basal Ganglia	RF Method 1	47.38%	0.63	0.63	0.60
	RF Method 2 (Cluster 1)	57.89%	0.52	0.50	0.40
	RF Method 2 (Cluster 2)	47.38%	0.52	0.50	0.40
Frontal Cortex	RF Method 1	57.89%	0.56	0.50	0.60
	RF Method 2 (Cluster 1)	47.38%	0.75	0.50	0.60
	RF Method 2 (Cluster 2)	52.63%	0.75	0.50	0.60
White Matter	RF Method 1	47.38%	0.67	0.63	0.60
	RF Method 2 (Cluster 1)	57.89%	0.89	0.75	0.80
	RF Method 2 (Cluster 2)	57.89%	0.25	0.25	0.40

Supporting Documents:

Although the other model evaluations are not that convincing with low accuracies, micro-precision, and micro recall rates within the testing data set. It is important to note that one misclassification would contribute to a large reduction in accuracy, precision, and recall rates due to this small sample size. In fact, one misclassification within controls, HIV and HIVE subject can result in zero precision or

recall rates for that class. Hence, it is important to test these RF models on a large sample. Moreover, we have noted that is RF methods 1 and 2 likely to have resulted in incomparable accuracy, precision, and recall rates.

Discussion:

Patients with HIV-1 infection have displayed different levels of neurocognitive impairment; some show mild impairment, while others appear to be affected more severely impaired (Dufour and others, 2018, Levine and others, 2013). If the impairment is severe enough, it might trigger an inflammatory response in the brain that leads to cell death with this comes difficulties with performing fine motor skills such as the finger-tapping task. The results have provided us with a better insight into a potential biomarker for HIV patients with different severity levels of HAND.

Although we have used the gene expression data from the same brain specimen as Gelman et al. (Gelman et al., 2012) we have applied a non-specific gene filter and hence ended up with a different number of DE genes within each brain sector. In contrast to our analysis, they have seen a clustering pattern among the four subject groups after performing principal component analysis. This can be due to the gene filtration effect.

Based on our analysis, the two brain regions that show the highest gene expression activities are frontal neocortex and basal ganglia. Frontal neocortex involves higher functions such as sensory perception, generation of motor commands, spatial reasoning, and conscious thoughts. Basal ganglia is responsible for motor function and other functions such as motor learning skills, executive functions, and emotions.

As per linear model analyses, CIRBP was significantly down-regulated while BTN3A3, IFIT3 MX1, HERC5, PARP14, ISG15 significantly up-regulated among HIVE compared to HAD both within basal ganglia and frontal cortex. RBM3 was found to be down-regulated among the same subjects within the frontal cortex.

SAM analyses reveals that genes, CIRBPRBM3, ZNF808, CDR1, THAP9-AS1, LOC1001928307 were down-regulated among HIVE subjects compared to HAD within frontal cortex and white matter whereas IFI44L, BST2, ISG15, BTN3A3, IFIT3, STAT1, IFIH1, MX1, and GPNMB were up-regulated

among HIVE subjects compared to HAD across all three brain regions. There were no any common down-regulated genes across all three bran regions among HIVE and HIV while GPNMB, ISG15, BTN3A3, LOC10041958, IFIT1, IFIH1, RCC2, and BTN3A2 were up-regulated within basal ganglia and white matter. IFI6, IFIT3, and ISG15 were found to be down-regulated among HAD and HIV across basal ganglia and frontal cortex. As CIRBP, RBM3, GPNMB, ISG15, IFIT6 genes were DE among the different subject groups with HAND, we find potentials to be used as biomarkers to detect HAND under further investigations.

Both CIRBP and RBM3 appear to play a role in the inflammatory response that leads to cell death in these patients and are associated with Alzheimer's disease they are very similar RNA-binding proteins that are up regulated in response to hypothermia (low temperatures) both of which appear to play similar roles in the regulation of numerous cellular events (Lanciego, Luquin and Obeso, 2012). CIRBP gene appears to activate the Akt and Erk pathways in neurons that block mitochondrial apoptosis, preventing cell death during hypothermic conditions (Li and others, 2012, Zhang and others, 2015). The Akt pathway is activated through the use of a kinase, phosphoinositide-3-kinase (PI3K), once activated it enhances cell proliferation and survival (Hemmings and Restuccia, 2015); the Erk pathway is activated by the binding of receptor tyrosine kinase (RTK) to a ligand, this induces cellular proliferation and activation of transcription factors that aid in cell survival of under certain conditions such as stress (McCain, 2013). Both pathways play a role in the protection of cells t. Thus, we can conclude that CIRBP expression in the brain prevents neuron cell death during both hypothermic and oxidative stress conditions (Liu and others, 2015).

Similarly, RBM3 plays a neuroprotective role under mild hypothermic conditions; studies have shown that RBM3 expression is inversely related to neuronal apoptosis (Chip and others, 2011, Peretti and others, 2015). ISG15 activity is tightly regulated by specific signaling pathways that have a role in innate immunity. ISG15 has identified as an interferon-stimulated gene since its expression is induced in response to type I interferons or lipopolysaccharide treatment (Malakhova and others, 2002). PNMB has been reported to be expressed in various cell types, including melanocytes, osteoclasts, osteoblasts, dendritic cells, and it is overexpressed in various cancer types (Zhou and others, 2012).

IFIT3 is related with IFN-induced antiviral protein which acts as an inhibitor of cellular as well as viral processes, cell migration, proliferation, signaling, and viral replication.

The gene, GADD45A, was consistently ranked among the top genes by both RF methods 1 and 2 within the frontal cortex brain region. Also, this was found to be significantly downregulated among HAD compared to HIV within the frontal cortex. This is a protein-coding gene whose transcript levels are increased with stressful growth arrest conditions (Li and others, 2018). Hence, researchers should give priority to investigating its potential to be used as an efficient biomarker.

Through our analyses, we were able to identify potential biomarkers in patients with HAND that could help detect the potential development of neurocognitive impairment before it occurs. Hyperthermia has shown to enhance HIV replication within the brain (Roesch and others, 2012); this is seen in patients with fever ranging from 38–40 ℃ and with this comes the inhibition of RBM3 and CIRBP genes with both appear to play a neuroprotective role. The gene expression of RBM3, CIRBP, GADD45A, and HIVE patients differed significantly from that of patients with HAND, which is why these genes can serve as potential biomarkers to diagnose neurocognitive impairments associated with HIV beforehand.

Abbreviations

HAND

HIV associated neurological disorders; cART:Combined antiretroviral treatment; DE:Differentially expressed; PCA:Principal component analyses; SAM:Significance analysis of microarrays; RF:Random forests; HIV:Human immunodeficiency virus; HAD:HIV-associated dementia; HIVE:HIV encephalitis; IFRG:Interferon response genes; GXNA:gene expression network analysis; GEO:Gene expression omnibus

Declarations

Competing interests

The authors declare that they have no competing interests.

Funding:
The authors acknowledge the research support from the College of Health Professions at UTRGV.
Supported by grants from the NIH/NINDS, R15NS108815-01(UR).

Authors’ contributions:

HR designed and did the data retrieval of the study. BM and HR led the implementation of the method and performed the data analysis. BM, RDLG, UR, and HR helped with the interpretation, description of

the results, and drafted the manuscript.

Acknowledgments
Not applicable.

Availability of data and materials

The data generated throughout the present study has been deposited to NCBI’s Gene Expression

Omnibus (GEO) under the GEO accession number of GSE35864.

Ethics approval and consent to participate

Not applicable.

References

Aouf M, Liyanage L, Hansen S (Year). Critical Review of Data Mining Techniques for Gene Expression Analysis. Proceedings of the 2008 4th International Conference on Information and Automation for Sustainability, 367–371.
Aschard H, Lutz S, Maus B, Duell EJ, Fingerlin TE, Chatterjee N, Kraft P, Van Steen K. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. Hum Genet. 2012;131:1591–613.
Bekku Y, Vargová L, Goto Y, Vorísek I, Dmytrenko L, Narasaki M, Ohtsuka A, Fässler R, Ninomiya Y, Syková, E. and Oohashi T. (2010). Bral1: Its Role in Diffusion Barrier Formation and Conduction Velocity in the CNS. The Journal of Neuroscience 30, 3113.
Boehm U, Klamp T, Groot M, Howard JC. Cellular responses to interferon-gamma. Annu Rev Immunol. 1997;15:749–95.
Breiman L. Bagging Predictors. Mach Learn. 1996;24:123–40.
Breiman L. Random Forests. Mach Learn. 2001;45:5–32.
Campbell EM, Hope TJ. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nat Rev Microbiol. 2015;13:471–83.
Chen S, Li S, Chen L. Interferon-inducible Protein 6–16 (IFI-6-16, ISG16) promotes Hepatitis C virus replication in vitro. J Med Virol. 2016;88:109–14.
Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012;99:323–9.
Chip S, Zelmer A, Ogunshola OO, Felderhoff-Mueser U, Nitsch C, Buhrer C, Wellmann S. The RNA-binding protein RBM3 is involved in hypothermia induced neuroprotection. Neurobiol Dis. 2011;43:388–96.
Chu G, Narasimham B, Tibshirani R, Tusher VG. (2005). SAM "Significance Analysis of Microarrays" Users guide and technical document. Policy.
Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016;1418:93–110.
Danforth K, Granich R, Wiedeman D, Baxi S, Padian N. (2017). Global Mortality and Morbidity of HIV/AIDS. In rd, K. K. Holmes, S. Bertozzi, B. R. Bloom and P. Jha, editors, Major Infectious Diseases. Washington (DC).
Dufour CA, Marquine MJ, Fazeli PL, Umlauf A, Henry BL, Zlatar Z, Montoya JL, Ellis RJ, Grant I, Moore DJ. A Longitudinal Analysis of the Impact of Physical Activity on Neurocognitive Functioning Among HIV-Infected Adults. AIDS Behav. 2018;22:1562–72.
Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
Everall I, Salaria S, Roberts E, Corbeil J, Sasik R, Fox H, Grant I, Masliah E. Methamphetamine stimulates interferon inducible genes in HIV infected brain. J Neuroimmunol. 2005;170:158–71.
Fischer T, Wyatt CM, D'Agati, Croul VD, McCourt S, Morgello L, S. and Rappaport J. Mononuclear

phagocyte accumulation in visceral tissue in HIV encephalitis: evidence for increased monocyte/macrophage trafficking and altered differentiation. Curr HIV Res. 2014;12:201–12.
Garcia-Mesa Y, Jay TR, Checkley MA, Luttge B, Dobrowolski C, Valadkhan S, Landreth GE, Karn J and Alvarez-Carbonell, D. (2017). Immortalization of primary microglia: a new platform to study HIV regulation in the central nervous system. J Neurovirol 23, 47–66.
Gelman, B.B., Chen, T., Lisinicchia, J.G. Soukup, V.M., Carmical J.R., Starkey, J., Masliah, E, Commins, D.L., Brandt, D., Grant, I., Singer, E.J., Levine, A. J., Miller, J., Winker, J.M., Fox, H.S., Luxon, B.A., Morgello, S., (2012). The national neuroAIDS tissue consortium brain gene array: two types of HIV-associated neurocognitive impairment. PLos One 7, e46178.
Ghafouri M, Amini S, Khalili K, Sawaya BE. HIV-1 associated dementia: symptoms and causes. Retrovirology. 2006;3:28.
Hemmings BA, Restuccia DF. (2015). The PI3K-PKB/Akt pathway. Cold Spring Harb Perspect Biol 7.
Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002;18(Suppl 1):233–40.
Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23:187–200.
Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15:379–93.
Karpievitch YV, Hill EG, Leclerc AP, Dabney AR, Almeida JS. An Introspective Comparison of Random Forest-Based Classifiers for the Analysis of Cluster-Correlated Data by Way of RF++. PLoS One. 2009;4:e7087.
Lanciego JL, Luquin, N, Obeso JA. Functional neuroanatomy of the basal ganglia. Cold Spring Harb Perspect Med. 2012;2:a009621.
Levine AJ, Miller JA, Shapshak P, Gelman B, Singer EJ, Hinkin CH, Commins D, Morgello S, Grant I, Horvath S. Systems analysis of human brain gene expression: mechanisms for HIV-associated neurocognitive impairment and common pathways with Alzheimer's disease. BMC Med Genomics. 2013;6:4.
Li FH, Han N, Wang Y, Xu Q. Gadd45a knockdown alleviates oxidative stress through suppressing the p38 MAPK signaling pathway in the pathogenesis of preeclampsia. Placenta. 2018;65:20–8.
Li S, Zhang Z, Xue J, Liu A, Zhang H. Cold-inducible RNA binding protein inhibits H(2)O(2)-induced apoptosis in rat cortical neurons. Brain Res. 2012;1441:47–52.
Liu J, Xue J, Zhang H, Li S, Liu Y, Xu D, Zou M, Zhang Z, Diao J. Cloning, expression, and purification of cold inducible RNA-binding protein and its neuroprotective mechanism of action. Brain Res. 2015;1597:189–95.
Malakhova O, Malakhov M, Hetherington C, Zhang DE. Lipopolysaccharide activates the expression of ISG15-specific protease UBP43 via interferon regulatory factor 3. J Biol Chem. 2002;277:14703–11.
Masliah E, Roberts ES, Langford D, Everall I, Crews L, Adame A, Rockenstein E, Fox HS. Patterns of gene dysregulation in the frontal cortex of patients with HIV encephalitis. J Neuroimmunol. 2004;157:163–75.
McCain J. The MAPK (ERK) Pathway: Investigational Combinations for the Treatment Of BRAF-Mutated Metastatic Melanoma. P T. 2013;38:96–108.
Nacu S, Critchley-Thorne R, Lee P, Holmes S. Gene expression network analysis and applications to immunology. Bioinformatics. 2007;23:850–8.
Peretti D, Bastide A, Radford H, Verity N, Molloy C, Martin MG, Moreno JA, Steinert JR, Smith T, Dinsdale D, Willis AE and Mallucci, G. R. (2015). RBM3 mediates structural plasticity and protective effects of cooling in neurodegeneration. Nature 518, 236–239.
Qi H, Chu V, Wu NC, Chen Z, Truong S, Brar G, Su SY, Du Y, Arumugaswami V, Olson CA, Chen SH, Lin CY, Wu TT, Sun R. Systematic identification of anti-interferon function on hepatitis C virus genome

reveals p7 as an immune evasion protein. Proc Natl Acad Sci U S A. 2017;114:2018–23.
Qi Y. Random Forest for Bioinformatics. In: Zhang C, Ma Y, editors. Ensemble Machine Learning: Methods and Applications. Boston: Springer US; 2012. pp. 307–23.
Qi Y, Li Y, Zhang Y, Zhang L, Wang Z, Zhang X, Gui L, Huang J. IFI6 Inhibits Apoptosis via Mitochondrial-Dependent Pathway in Dengue Virus 2 Infected Vascular Endothelial Cells. PLoS One. 2015;10:e0132743.
Roberts ES, Masliah E, Fox HS. CD163 identifies a unique population of ramified microglia in HIV encephalitis (HIVE). J Neuropathol Exp Neurol. 2004;63:1255–64.
Roberts ES, Zandonatti MA, Watry DD, Madden LJ, Henriksen SJ, Taffe MA, Fox HS. Induction of pathogenic sets of genes in macrophages and neurons in NeuroAIDS. Am J Pathol. 2003;162:2041–57.
Roesch F, Meziane O, Kula A, Nisole S, Porrot F, Anderson I, Mammano F, Fassati A, Marcello A, Benkirane M, Schwartz O. Hyperthermia stimulates HIV-1 replication. PLoS Pathog. 2012;8:e1002792.
Shityakov S, Dandekar T, Forster C. Gene expression profiles and protein-protein interaction network analysis in AIDS patients with HIV-associated encephalitis and dementia. HIV AIDS (Auckl). 2015;7:265–76.
Slonim DK. From patterns to pathways: gene expression data analysis comes of age. Nat Genet. 2002;32 Suppl:502–8.
Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3.
Spudich SS. Immune activation in the central nervous system throughout the course of HIV infection. Curr Opin HIV AIDS. 2016;11:226–33.
Tin Kam H. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell. 1998;20:832–44.
Tin Kam H, Hull JJ, Srihari SN. Decision combination in multiple classifier systems. IEEE Trans Pattern Anal Mach Intell. 1994;16:66–75.
Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98:5116–21.
Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. J Am Stat Assoc. 2004;99:909–17.
Z.J, W. and R., I. (2010). Description of gcrma package.
Zhang HT, Xue JH, Zhang ZW, Kong HB, Liu AJ, Li SC, Xu DG. Cold-inducible RNA-binding protein inhibits neuron apoptosis through the suppression of mitochondrial apoptosis. Brain Res. 2015;1622:474–83.
Zhou LT, Liu FY, Li Y, Peng YM, Liu YH, Li J. Gpnmb/osteoactivin, an attractive target in cancer immunotherapy. Neoplasma. 2012;59:1–5.

Figures

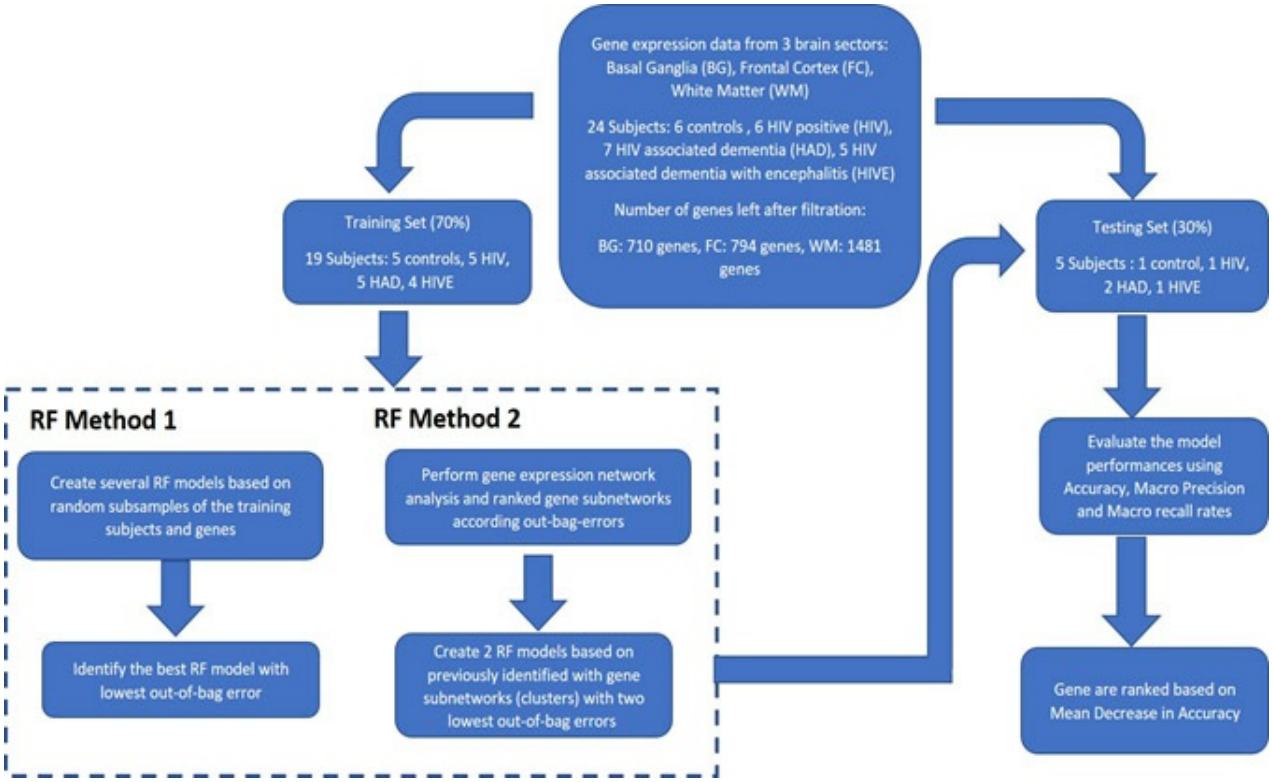


Figure 1

Details of the RF method 1 and 2. RF models within the first approach are created by taking random subsamples of filtered genes from each brain regions and the best RF models per each brain region are selected based on the lowest OOB error rate. There were two steps in the second RF approach. First, correlated gene subnetworks were identified using the gene expression network analysis (GXNA) tool. Following that, the high-scoring (The top two clusters based on the T∑ scoring introduced in Nacu et al, 2007) gene subnetworks are used to develop the RF models.

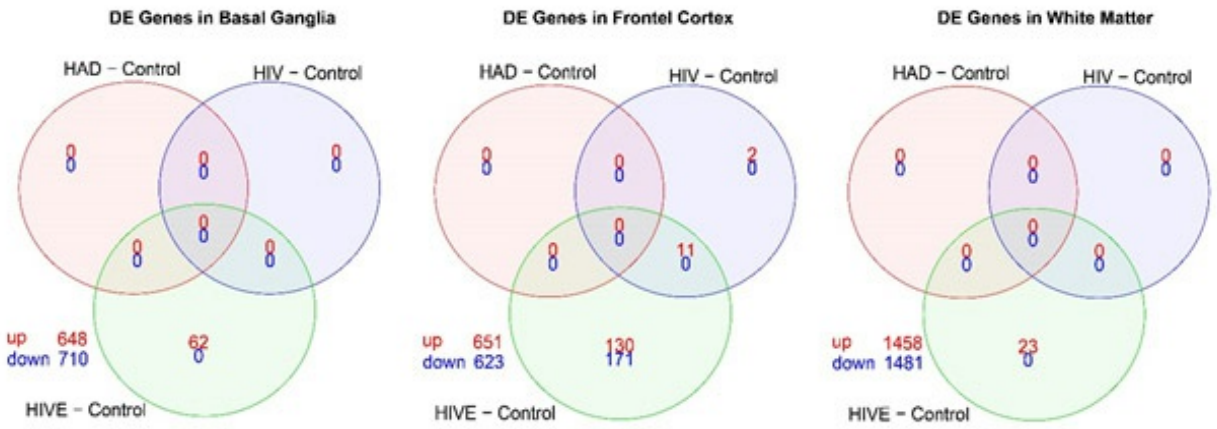


Figure 2

Genes with an adjusted p-value less than 0.01, using Benjamini-Hochberg correction for multiple testing comparisons, were identified as differentially expressed. Differentially Expressed Genes (a) Basal Ganglia (b) Frontal Cortex (c) White Matter. The subjects with HIVE had the most regulated genes in all three brain sectors. Many of the genes associated with the HIVE group within the frontal cortex were downregulated, while many of them are upregulated in the white matter and basal ganglia. Subjects with HIV, but without either HAD or HIVE, had 11 upregulated genes common with HIVE subjects, namely, IFI44L*, DTX3L, GBP1*, IFIT3*, AQP1, MX1*, PLSCR1, RARRES3, SPARC, TIMP1, and SERPING1. As per above mentioned DE gene selection criterion, subjects with HAD did not exhibit any DE genes over control subjects across all three brain sectors. compared to HAD. Genes marked with * are interferon response genes that are related to the cell functions in the immune system and autoimmunity.

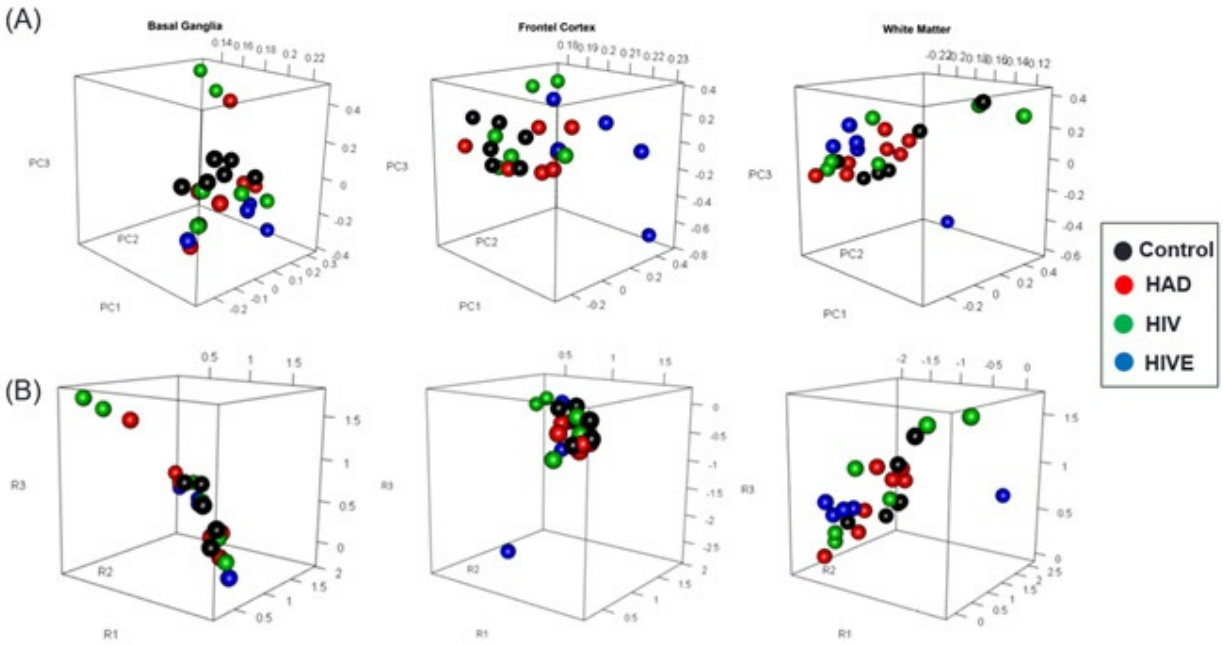


Figure 3

(a): Principal Component Analysis of the controls and other 3 HIV-1 subject groups within the three brain regions, basal ganglia, frontal cortex, and white matter. No clear groupings were observed among the different subject groups (b) Rotated Loadings with Varimax factor rotation for basal ganglia, frontal cortex, and white matter. Even after the Varimax rotation of PCA, no significant clusters were observed.

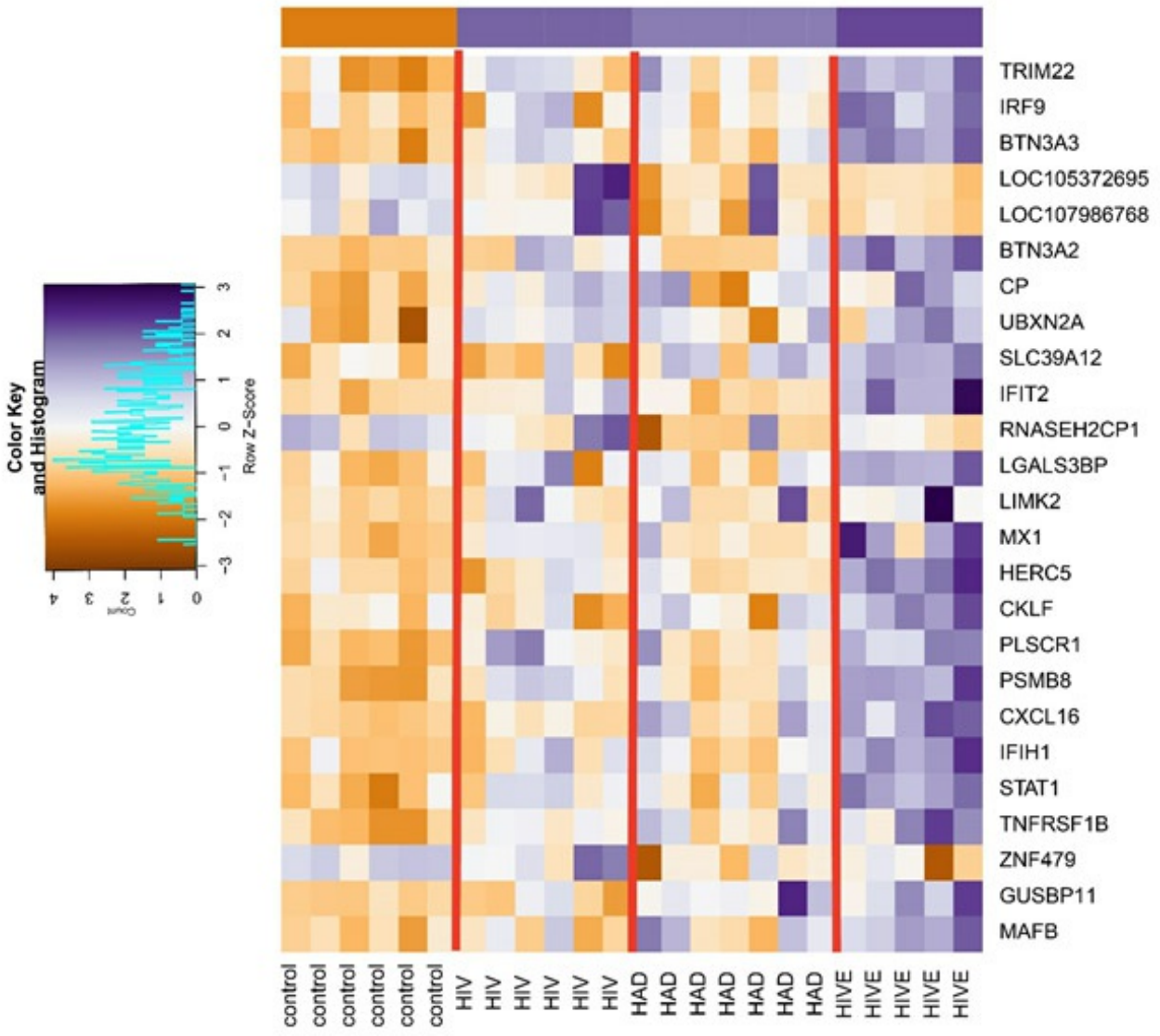


Figure 4

Genes expression levels of basal ganglia identified by the RF Method 1 (Refer Table S4) to get specific gene rankings with respect to the mean decrease in accuracy along with their fold change and q values). Among the top 25 genes identified by RF method 1, the genes IFIT2, IFIH1, LGALS3BP, CKLF, HERC5, PSMB8, PLSCR1, STAT1, BTN3A3, CXCL16, BTN3A2, TRIM22, MX1, and IRF9 were up-regulated at least by 3 fold and had q values < 0.07 within HIVE compared to HAD subjects. Among those, BTN3A2, IFIH1 and STAT1 had at least a 6 fold change. There were no up or down-regulated genes identified by SAM among the top 25 genes of RF method 1 between HAD and HIV subjects.

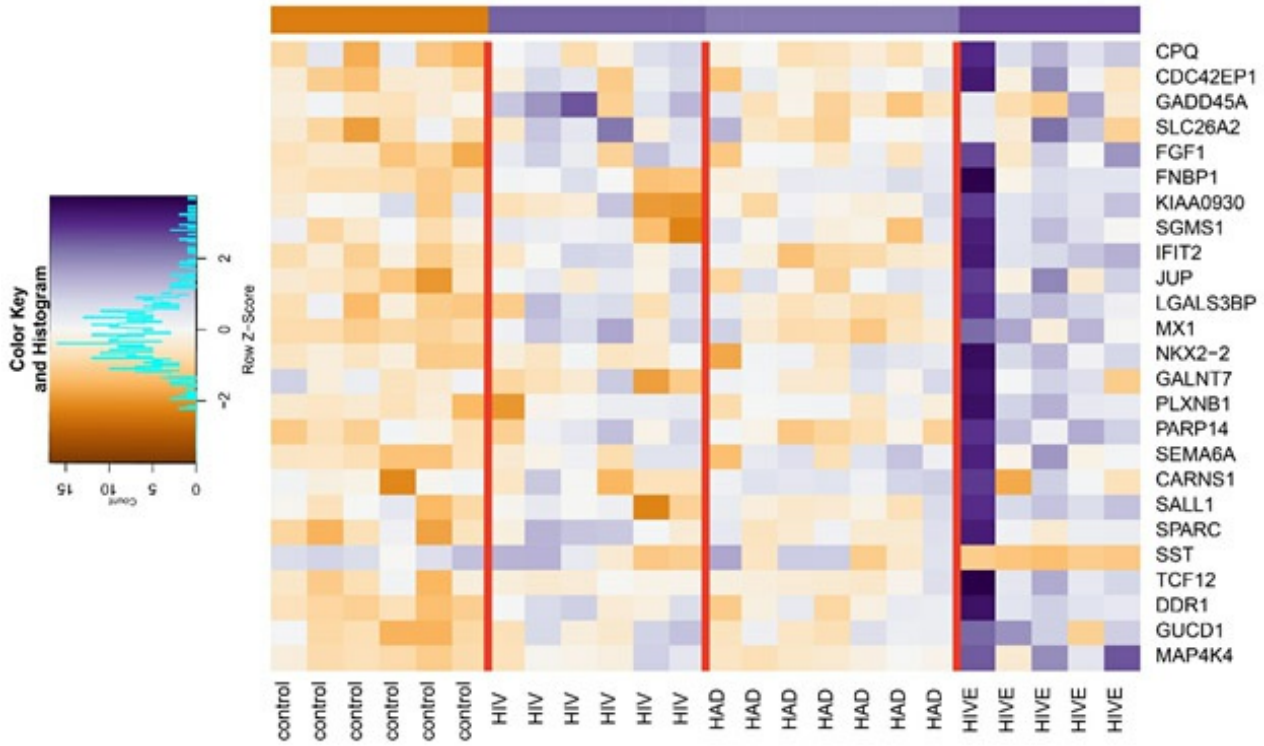


Figure 5

Gene expression levels of the frontal cortex by RF Method 1. SST which was ranked #7 by RF method 1 was significantly down-regulated with a q-value 0.00 among HIVE compared to HAD subjects. The genes, MX1, MAP4K4, CPQ, SALL1, LGALS3BP, PARP14, and IFIT2, were significantly up-regulated with q- values< 0.04 among them. MX1 and GADD45A were significantly down-regulated with q-values<0.00 among HAD compared to HIV. GADD45A is a protein-coding gene whose transcript levels are increased with stressful growth arrest conditions. Since, GADD45A was down-regulated only between HAD and HIV, this need to be further investigated to be used as a potential biomarker.

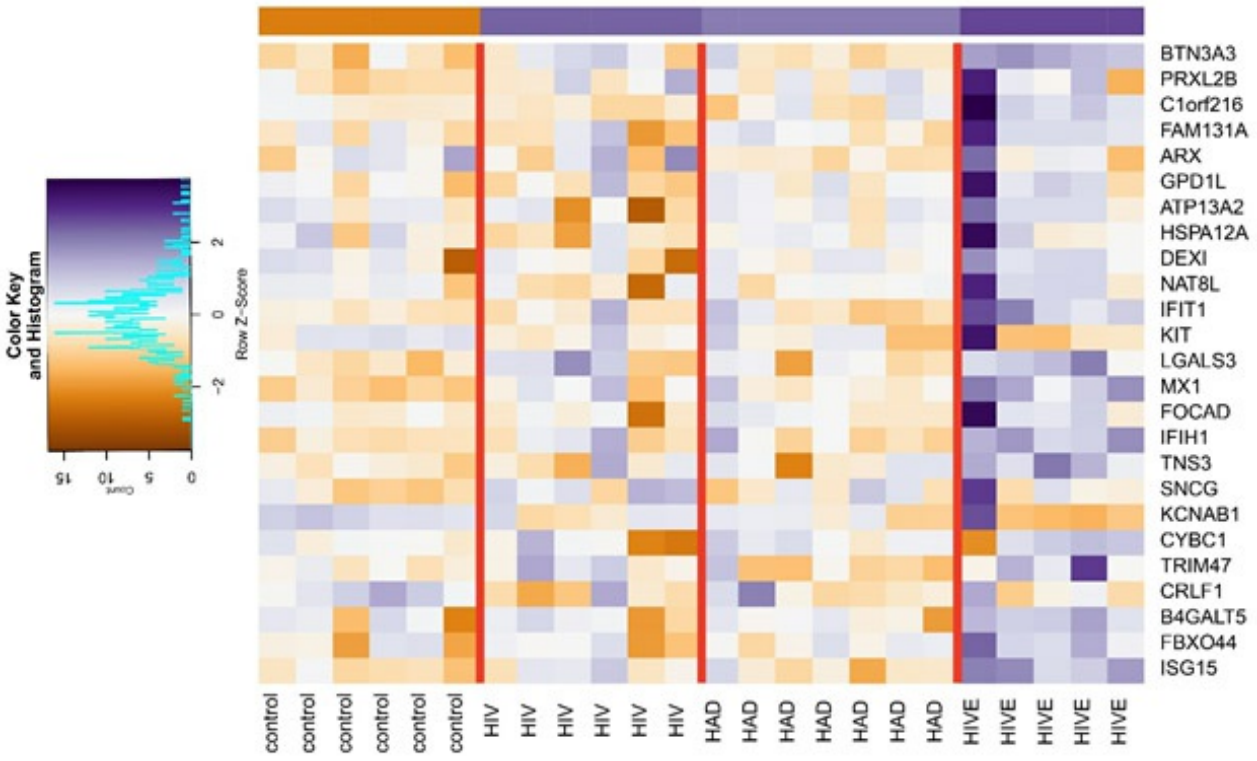
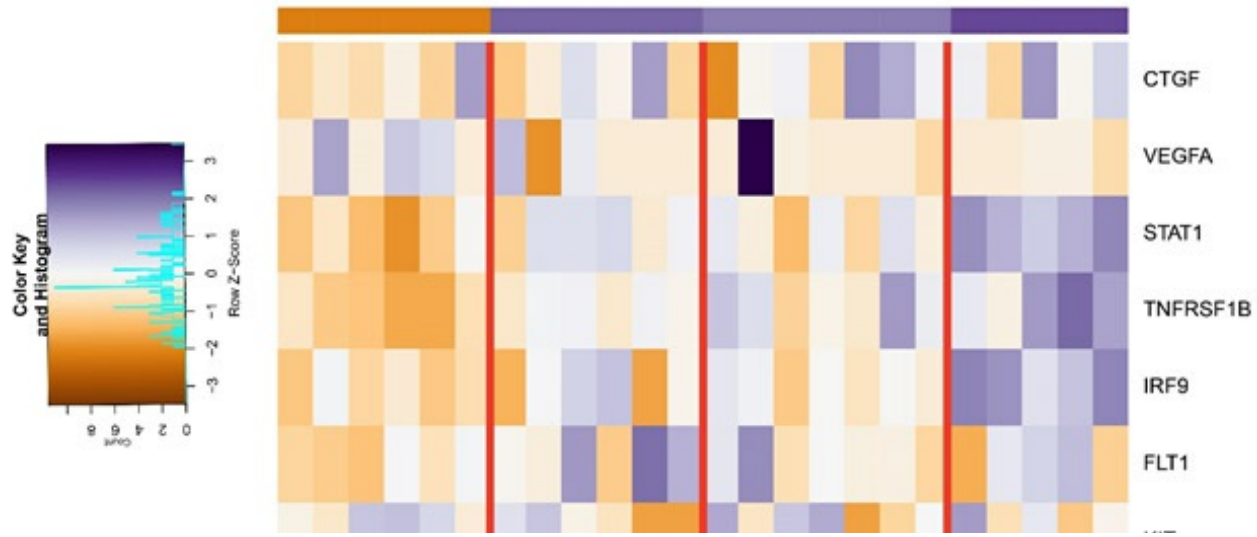


Figure 6

Gene expression levels of white matter by RF Method 1. BTN3A3, ISG15, IFIH1, MX1, and IFIT1 which were among the top 25 ranked genes were significantly upregulated with at least 4 folds and q-values < 0.03 among HIVE compared to HAD subjects. Although not ranked within the first 25, RBP1, LRRCC1, OPALIN, CPAMD8, LOC100506114, CIRBP, and RBM3 were there among the 48 down-regulated genes within the HIVE compared to HAD. None of the other top 25 genes were significantly up or down-regulated (with a q-value < 0.05) within white matter among HIVE and HIV or HAD and HIV subjects.



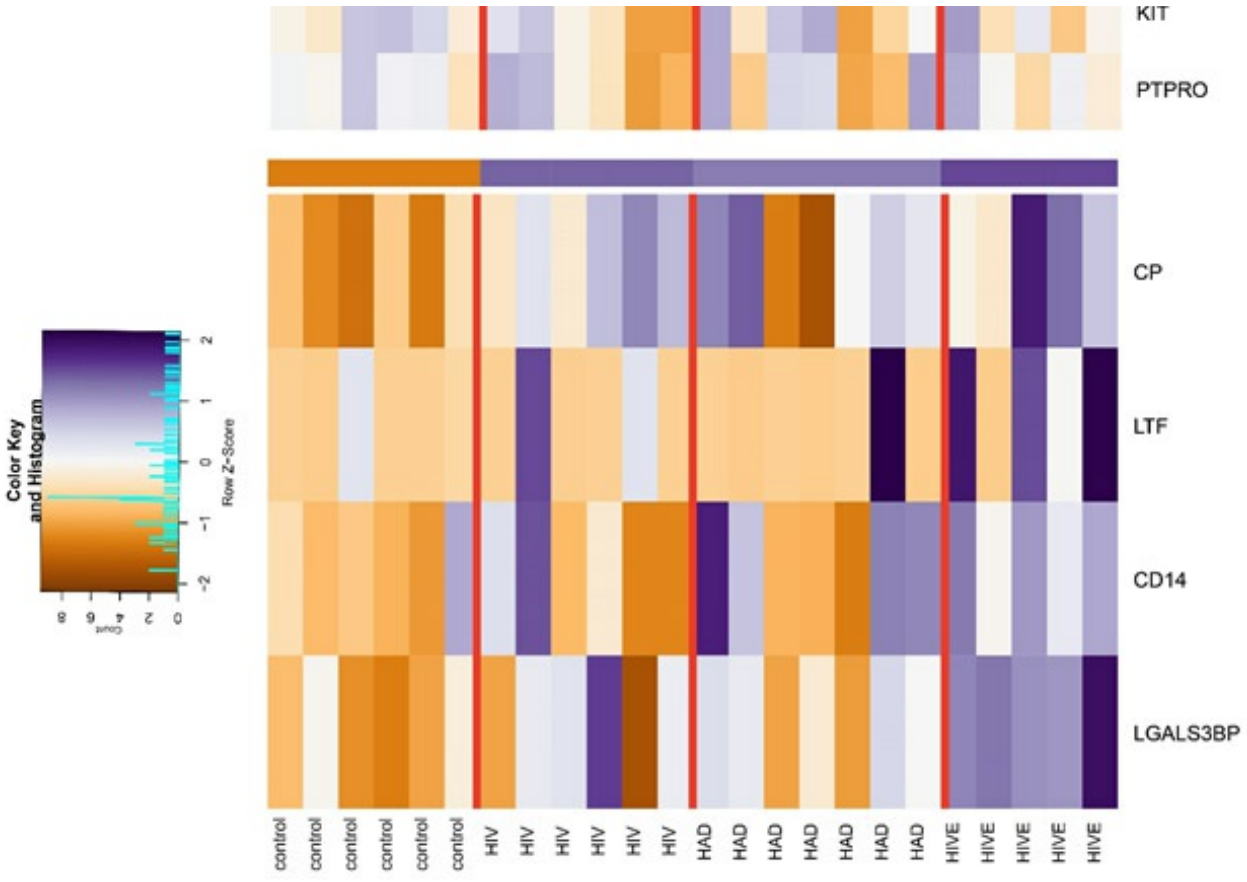
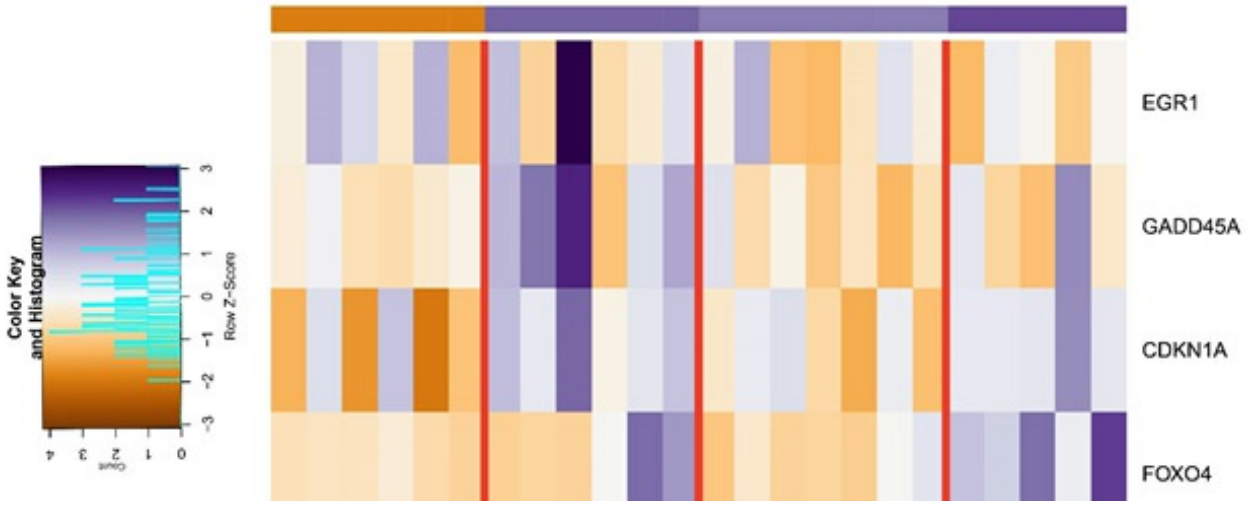


Figure 7

Genes expression levels of basal ganglia identified by RF Method 2. Top and bottom graphs are related to the two top-ranking gene networks identified by GXNA. One cluster had eight genes and the other one had just four genes. STAT1 and IRF9 within the first cluster and the LGALS3BP from the second cluster were significantly upregulated within the HIVE compared to HAD with q-values < 0.03. None of the genes within two clusters were DE both among HAD and HIVE compared to HIV under q-value < 0.05.



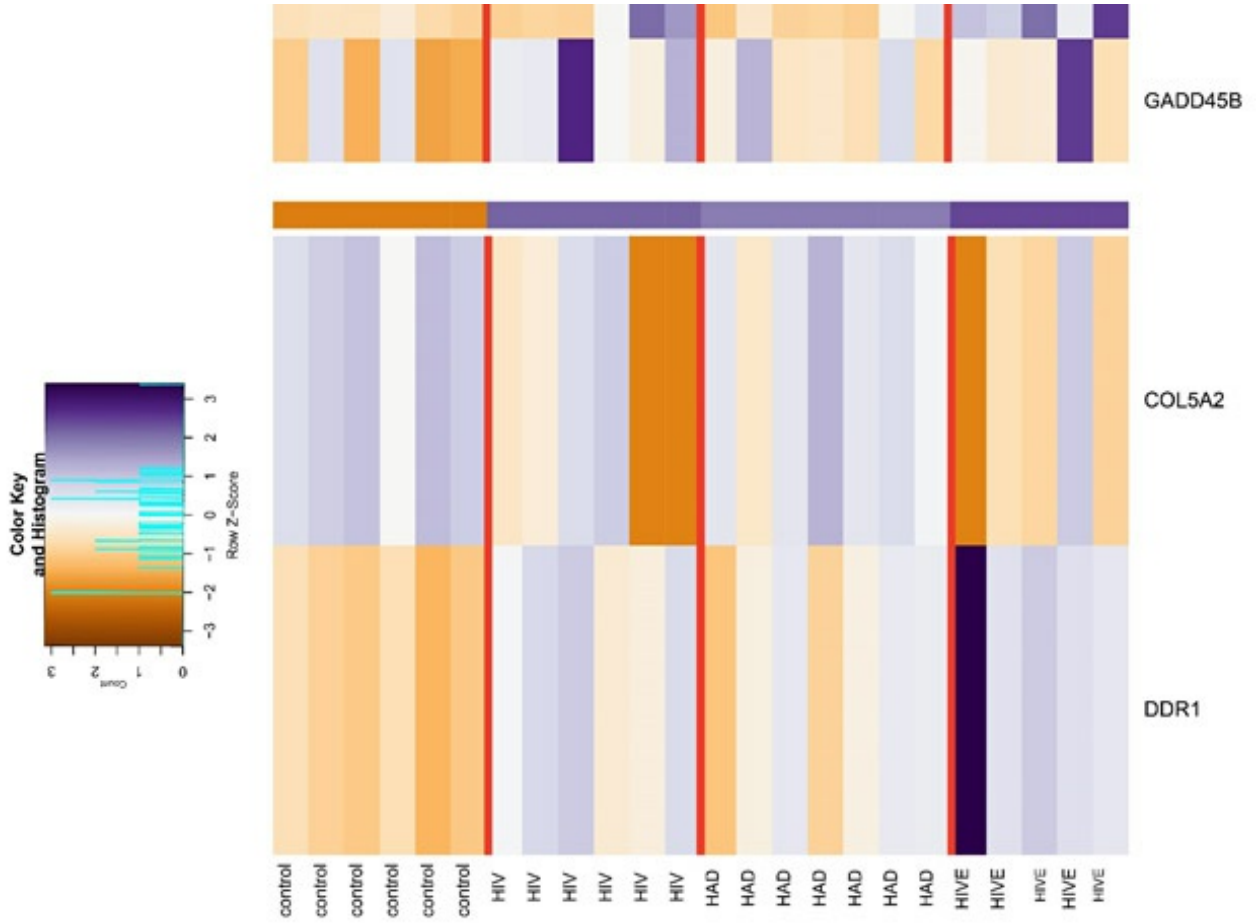
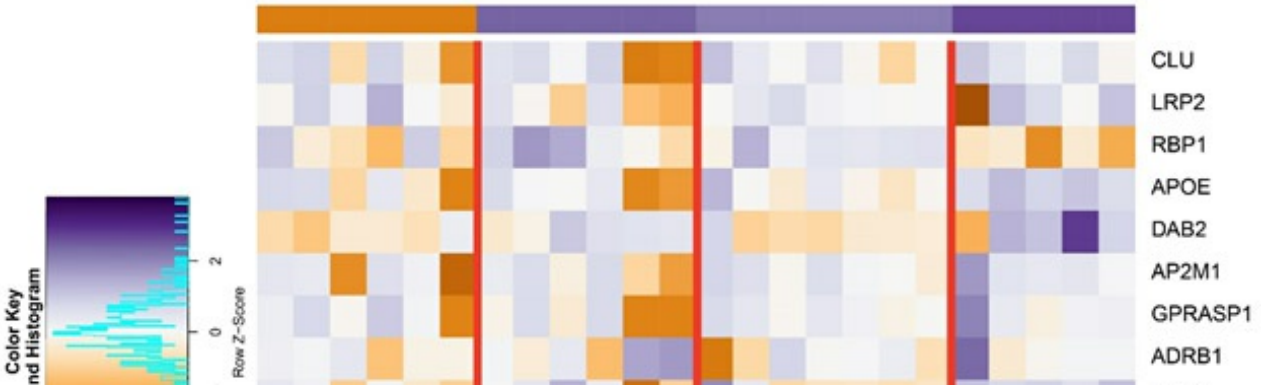


Figure 8

Genes expression levels of the frontal cortex identified by RF Method 2. FOX04 within the first cluster was significantly up-regulated with a q-value 0.04 and COL5A2 within the second cluster was significantly down-regulated among HIVE compared to HAD within the frontal cortex. GADD45A within the first cluster was the only gene that is significantly downregulated among HAD compared to HIV with a q-value 0.00. This gene was consistently ranked among the top genes by both RF methods 1 and 2. None of the genes within two clusters were DE among HIVE compared to HIV.



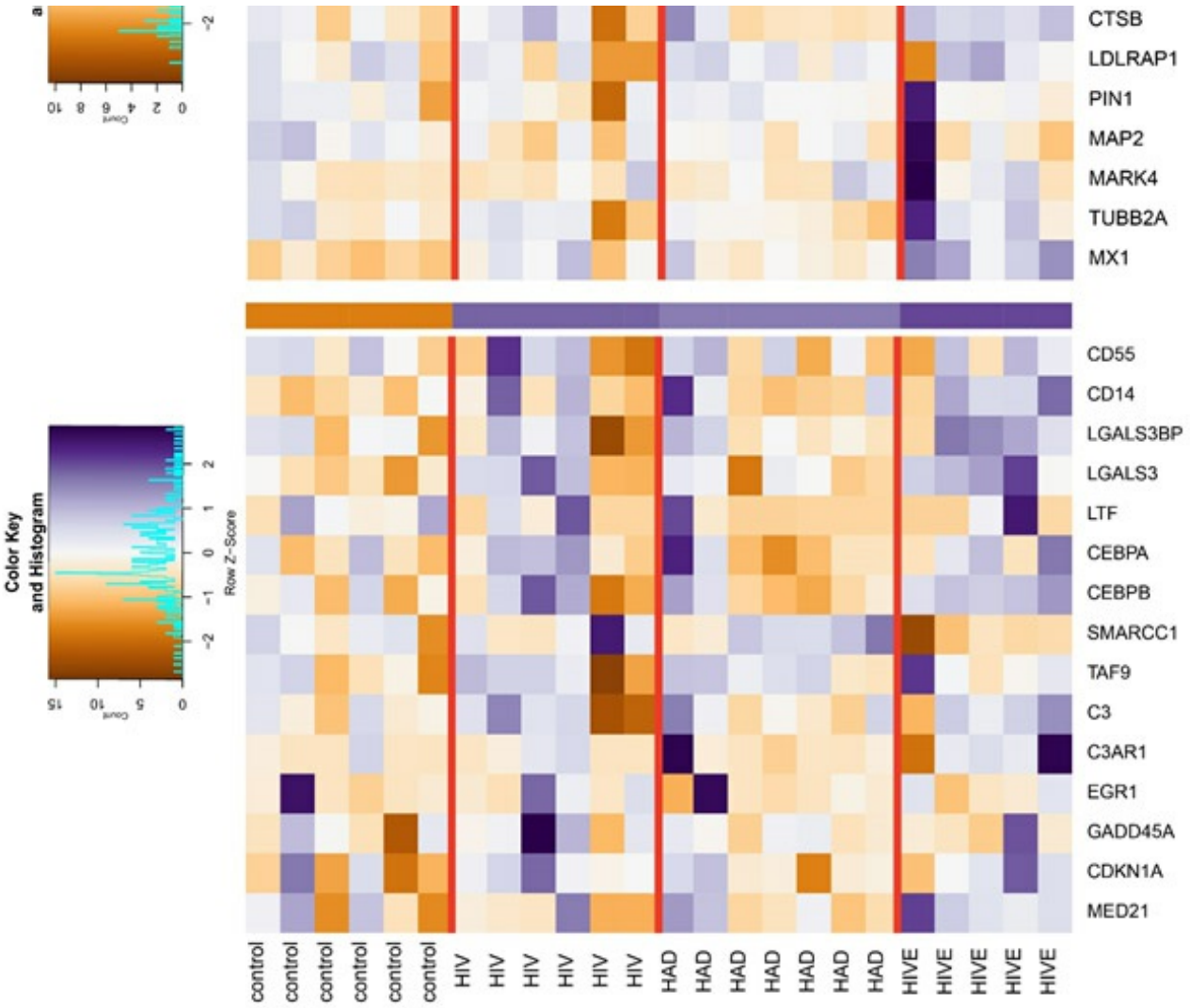


Figure 9

Genes expression levels of white matter identified by RF Method 2. RBP1 was identified to be significantly downregulated while and MX1 was significantly upregulated within the white matter with a q-values < 0.04 among HIVE compared to HAD. RBP1 also significantly downregulated among HIVE compared to HIV subjects. Surprisingly, the bottom cluster did not show any DE regulated genes among any of the subject group comparisons.