ORIGINAL INVESTIGATION

Gene biomarkers and classifiers for various subtypes of HTLV-1-caused ATLL cancer identified by a combination of differential gene co-expression and support vector machine algorithms

Mohadeseh Zarei Ghobadi[1] · Elaheh Afsaneh[2] · Rahman Emamzadeh[1]

Abstract

Adult T-cell leukemia/lymphoma (ATLL) is pathogen-caused cancer that is progressed after the infection by human T-cell leukemia virus type 1. Four significant subtypes comprising acute, lymphoma, chronic, and smoldering have been identified for this cancer. However, there are no trustworthy prognostic biomarkers for these subtypes. We utilized a combination of two powerful network-based and machine-learning algorithms including differential co-expressed genes (DiffCoEx) and support vector machine-recursive feature elimination with cross-validation (SVM-RFECV) methods to categorize disparate ATLL subtypes from asymptomatic carriers (ACs). The results disclosed the significant involvement of CBX6, CNKSR1, and MAX in chronic, MYH10 and P2RY1 in acute, C22orf46 and HNRNPA0 in smoldering subtypes. These genes also can classify each ATLL subtype from AC carriers. The integration of the results of two powerful algorithms led to the identification of reliable gene classifiers and biomarkers for diverse ATLL subtypes.

Keywords HTLV-1 · ATLL · Cancer · SVM · DiffCoEx · ATLL subtypes

Introduction

Human T-cell lymphotropic virus type 1 (HTLV-1) is a deltaretrovirus that infects CD4 + T and can develop cancer named adult T-cell leukemia/lymphoma (ATLL). HTLV-1 may induce cell transformation and proliferation, as well as immune deficiency in the infected individuals [1, 2]. ATLL is developed in almost 5% of the HTLV-1 infected subjects. ATLL is primarily classified into four subtypes based on Shimoyama classification: lymphoma, acute, chronic, and smoldering [3, 4]. The lymphoma and acute types are specified by very poor prognosis and aggressive manner. The smoldering and chronic subtypes are nominated by an indolent clinical course and disparate clinicopathologic traits. The cancer cells are mainly in the blood in acute ATLL in spite of lymphomatous ATLL, in which cancer cells are mostly placed in the lymphatic organs and lymph nodes [5]. Chronic ATLL is specified as a slow-growing or low-grade cancer. Smoldering subtype is also a low-grade disease and can make no problem without any signs for a long time [6].

Exploration of the most effective functional players in the development of a disease help find potential biomarkers and also possible therapeutic targets. One of the proper methods is the identification of co-expressed genes that probably have similar functions and also regulate the same biological pathways. Algorithms such as DiffCoEx that find the differential co-expressed genes between two conditions result in finding the most particular co-expressed genes in a determined disease state. The remarkable benefit of this approach is the identification of differential co-expressed gene groups in the presence of the within-group correlation between two conditions [7].

Support vector machine (SVM) method is a type of machine-learning approach that contains a series of supervised learning approaches that could be applied for regression, outlier detection, and classification. It can be used to

---

Edited by Matthias J. Reddehase.

✉ Mohadeseh Zarei Ghobadi
mohadesehzaree@gmail.com

✉ Rahman Emamzadeh
r.emamzadeh@sci.ui.ac.ir

1 Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran

2 Independent Researcher, Tehran, Iran

classify biological data and find the possible subtypes of a specified disease. SVM basically categorizes the samples into two classes. However, it can be utilized for multiclass classification by crushing the problem into several binary classification ones. It is carried out by considering a high dimensional space and mapping data points to it to acquire a reciprocal linear differentiation between every two classes known as the one-to-one method [8].

The poor prognosis of ATLL and its subtypes is a concerning problem in the endemic regions [9]. Therefore, in this study, we used a combination of Diffcoex and SVM algorithms for categorizing three ATLL subtypes. It resulted in the identification of the gene classifiers between asymptomatic carriers (ACs), who carry HTLV-1 without any sign, and patients with one of the three ATLL subtypes. The determined genes may have critical roles in the progression of each cancer subtype and would be considered potential biomarkers.

Results

Identification of differential co-expressed modules

The overall steps of analysis in this study are summarized in Fig. 1. The differential co-expressed gene modules (DiffCoexGMs) between ACs and three ATLL subtypes were found utilizing the DiffCoEx algorithm [10]. This method actually determines the groups of genes that are specifically co-expressed in an ATLL subtype. According to the steps of executing the algorithm explained in Materials and methods, adjacency differences matrices between two conditions

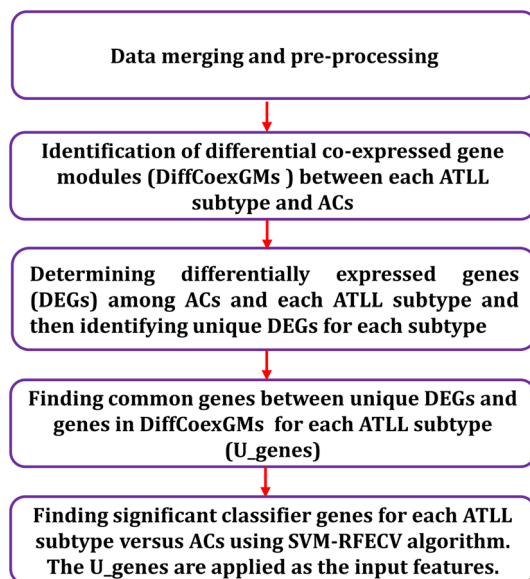**Fig. 1** Flowchart of the suggested procedure in this study

were initially determined. Afterward, dissTOMs were calculated with β power of 9, 6, and 5 for ACs-ATLL_acute, ACs-ATLL_chronic, and ACs-ATLL_smoldering, respectively. Next, the hierarchical tree was constructed using the flashClust function. The differential co-expressed modules were then identified by cutting the branches of the tree. Figure 2a–c shows a dendrogram of identified modules based on a topological overlap dissimilarity matrix named diss-TOM (1-TOM), in which each color denotes a module. Two genes have high topological overlap if they are connected to a similar group of genes in the network. The dissTOM shows the common genes with shared neighbors in a graph, which finally leads to identifying modules with a high correlation with a condition. Furthermore, the heatmaps depicting the discrepancies between the correlation patterns in DiffCoexGMs are indicated in Fig. 3. The upper diagonal of the heatmap reveals a correlation between pairs of genes in each module among ATLL subtypes (the red color denotes positive correlations and the blue color denotes negative correlations). The lower diagonal of the heatmap represents a correlation between similar gene pairs in ACs. The high intensity of color shows a higher correlation between the gene expression profiles in a module in a condition. A total of six DiffCoexGMs were found between ACs and ATLL_acute, in which the green module was highly correlated with ATLL_acte (Fig. 3a). In comparison between ACs and ATLL_chronic, five DiffCoexGMs were found. Among them, the turquoise module was correlated with ATLL_chronic (Fig. 3b). Two of the five DiffCoexGMs between ACs and ATLL_smoldering including blue and brown were correlated with ATLL_smoldering (Fig. 3c). The identified modules comprise co-expressed genes that are possibly more involved in the progression of ACs toward each ATLL subtype (Supplementary data file 1). The pathway enrichment analysis of the identified modules was also performed. The results revealed the involvement of the genes in mostly cancer-related pathways. The related enriched pathways are mentioned in Supplementary data file 2. In the next step, we found differentially expressed genes (DEGs) among ACs and each subtype and then the unique ones for each subtype (Supplementary data file 3). Afterward, we found the common genes between unique DEGs and the genes in each specific DiffCoexGMs for each subtype (Supplementary data file 4). These unique genes (U_genes) could be further used as features in the machine-learning classification algorithm.

Finding classifiers using support vector machine-recursive feature elimination with cross-validation (SVM-RFECV) analysis

To find the best unique classifiers between ACs and ATLL subtypes, the SVM-RFECV algorithm was used. The U_genes obtained from previous analyses were applied as the
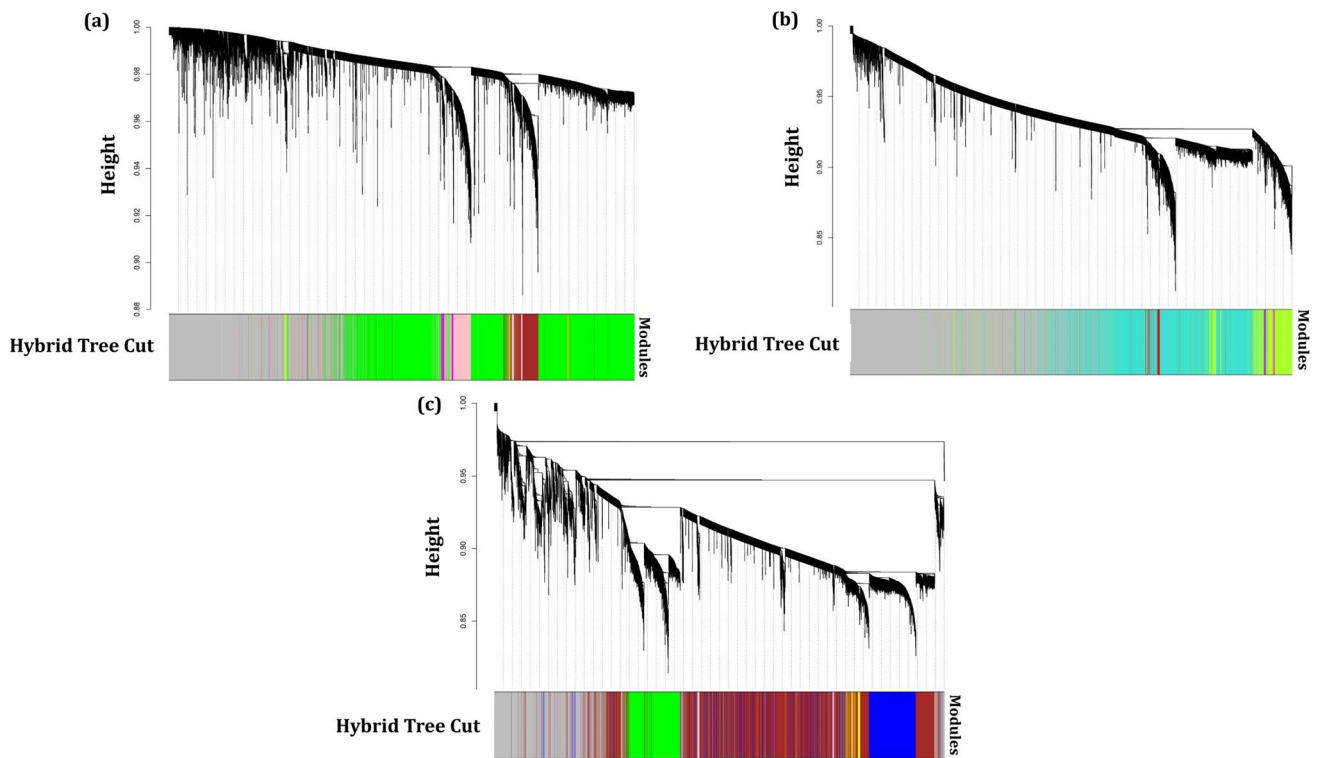
Fig. 2 Dendrogram of clustered genes based on (1-TOM) with defined module colors of **a** ACs-ATLL_acute, **b** ACs-ATLL_chronic, and **c** ACs-ATLL_smoldering. The colors represent the modules (groups of genes)
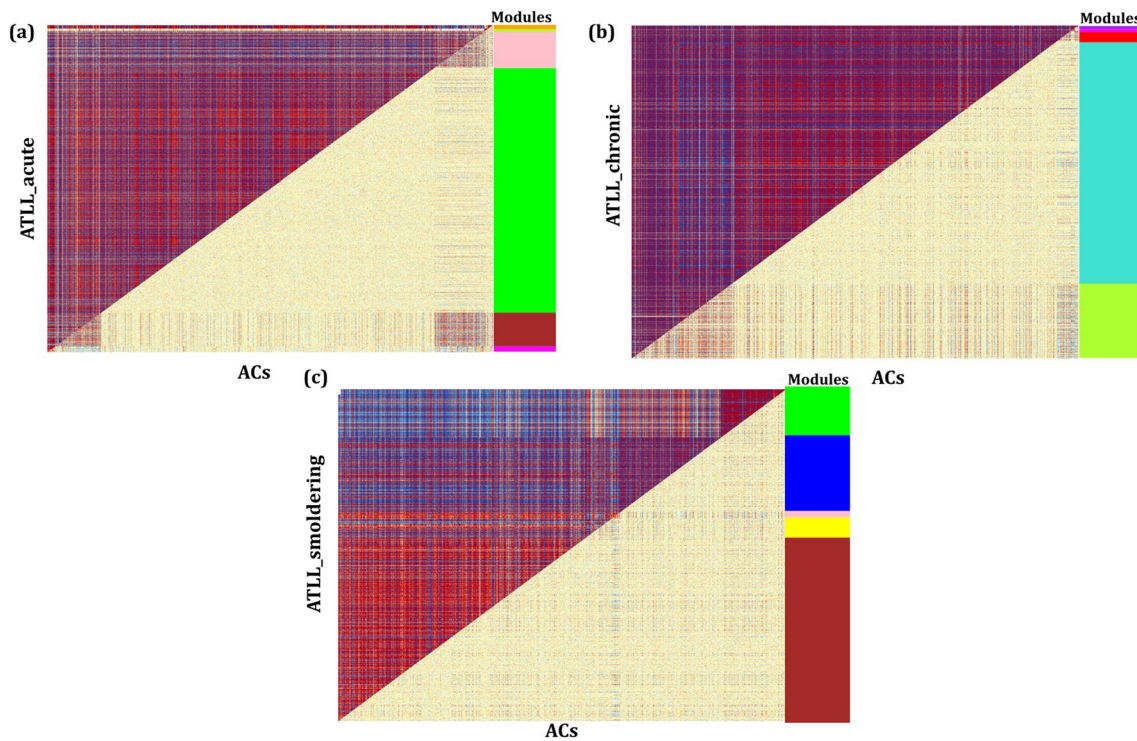

Fig. 3 Heatmap of comparative correlation comprising differentially co-expressed modules between **a** ACs and ATLL_acute, **b** ACs and ATLL_chroninc, and **c** ACs and ATLL_smoldering. The upward and downward diagonals denote the correlations between gene pairs in ATLL and ACs, respectively

features for each analysis. Therefore, a total of 56, 14, and 103 U_genes were employed as the input features for ATLL_acute, ATLL_chronic, and ATLL_smoldering, respectively (Supplementary data file 4). The data were randomly divided to train and test sets (65/35). The obtained accuracies for train datasets were found as 0.98, 1.00, and 1.00 for ATLL_acute, ATLL_chronic, and ATLL_smoldering, respectively. Afterward, the test datasets were used to assess the constructed models. Figure 4a–c indicates the confusion matrices and classification reports (precision, recall, and F1-score) for three subtypes of ATLL. Precision is the ratio of correctly observed positive outcomes to all observed positive outcomes. Recall is the ratio of correctly observed positive outcomes to the total observations in a desired class. F1-score is a substantial performance metric. It is defined as the harmonic average of precision and recall. The F1-score ranges between 0 and 1. A value larger than 0.5 shows a relatively good classification and a value lower than 0.5 reveals a failed classification. As indicated in Fig. 4, the constructed models for all three subtypes showed successful classification results. The accuracies of test datasets were obtained as 1.00 for ATLL_acute, 0.95 for ATLL_chronic, and 0.94 for ATLL_smoldering. The model identified MYH10 and P2RY1 as the best classifier for classifying ATLL_acute, CBX6, CNKSR1, and MAX for ATLL_chronic, C22orf46 and HNRNPA0 for ATLL_smoldering from ACs. These

genes are the potential biomarkers and possible therapeutic targets for each ATLL subtype.

Discussion

In this study, we carried out the consecutive stages to find the best diagnostic gene classifiers between asymptomatic carriers and the subjects with one of the three ATLL subtypes. For this purpose, the differential co-expressed modules were initially used to identify the specific modules related to each subtype. Next, the shared genes between the unique DEGs and genes in the selected modules were determined. Afterward, the unique genes were used as the input variables in the SVM-RFE machine-learning algorithm.

The analysis revealed *MYH10* and *P2RY1* as the classifier between ACs and ATLL_acute subjects. MYH10 is a cellular protein myosin that has possibly several functions in cell shape, cytokinesis, secretion, and capping [11]. It also has a significant role in the cytoskeleton reorganization and lamellipodial extension during cell spreading [12]. Non-muscle myosin IIB (NM IIB) is encoded by *MYH10* which is implicated in tumor invasion, cell migration, and the generation of extracellular matrix (ECM) [13].

P2RY1 encodes a protein belonging to the G-protein coupled receptors family, which contains a receptor for extracellular adenine nucleotides such as ADP [14, 15].
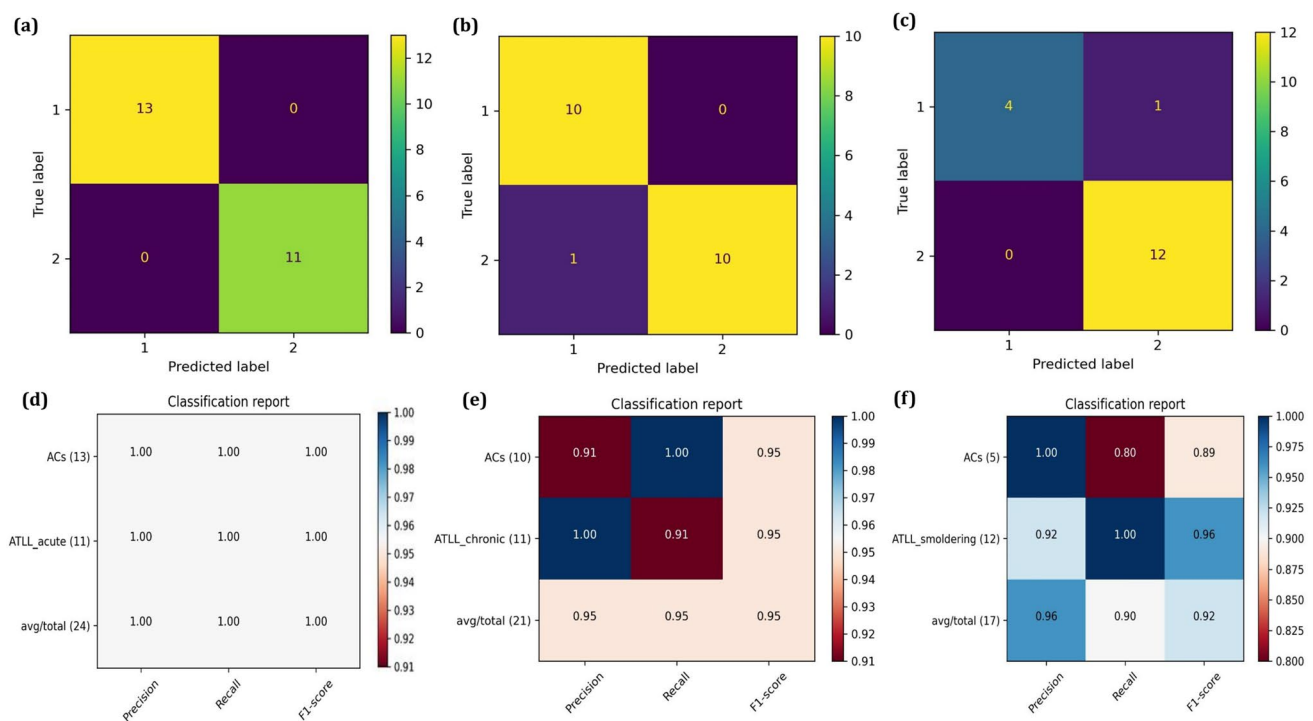

Fig. 4 The confusion matrix (**a–c**) and classification reports (**d–f**) for ATLL_acute, ATLL_chronic, and ATLL_smoldering subtypes, respectively

P2RY1 liberates Ca^{2+} from the intracellular reservoir by coupling to phospholipase C, resulting in reversible platelet aggregation [16, 17]. It is also involved in the control of several physiological functions [18]. It has been demonstrated that the P2RY1 receptor mediates apoptosis in prostate cancer and astrocytoma cells in correlation with the ERK1/2 activation. It is also known as a cell proliferation inhibitor. Therefore, P2RY1 receptor is a potential anticancer target [19, 20].

CBX6, *MAX*, and *CNKSR1* were identified as the best gene classifier for ATLL_chronic. CBX6 is a subunit of polycomb repressive complex 1 that intercedes the suppression of the epigenetic gene and operates as a tumor suppressor or an oncogene [21]. It can promote G1/S phase transition, proliferation, invasion, and metastasis capability of the cancer cells through adjusting transcription factors snail/zeb1-mediated EMT mechanism [22].

MAX is a transcription factor, which is capable of a complex constitution with Myc. Myc is an oncoprotein involved in cell differentiation, proliferation, and apoptosis [23]. MYC-MAX boosts tumorigenesis and has a significant function in lymphomagenesis. MYC-MAX heterodimers steer a feed-forward circuit that amplifies high MYC expression in tumors. MYC regulates the kinase Plk1 to retain its stability in an aggressive type of lymphoma. Disturbance in this process, Max null B cells leads to a decrease of MYC protein expression and full repeal of lymphomagenesis [24, 25]. MAX deletion results in metabolic rewiring and context-specific tumor repression [26].

CNKSR1 encodes a kinase inhibitor of ras gene 1 which is a linker enhancer for an enzyme. It is a necessary component in the receptor tyrosine kinase pathway and can be utilized to target tyrosine phosphorylation. CNKSR1 acts as an oncogene and accelerates RAF/MEK/ERK signaling through the positive feedback effects via its plasmalemmal localization [27]. Moreover, the PH domain of CNKSR1 incorporates mut-KRAS to interdict the development of mut-KRA cells, so can treat some cancers [28].

Two genes including *C22orf46* and *HNRNPA0* were found to be potentially important genes involved in the pathogenesis of ATLL_smoldering. *C22orf46* is a pseudogene that is predicted to be located in the extracellular region. The function of C22orf46 in cancer has not been found yet. There is only one report that proposes the oncogenic role of C22orf46 in adrenocortical carcinoma cells (ACC) [29]. It has been shown that circ-CCAC1 facilitates ACC progression through miR-514a-5p/C22orf46 signaling. Therefore, the function of C22orf46 should be further surveyed. *HNRNPA0* is a member of the A/B family of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs). It encodes a protein that has two copies of quasi-RRM domains that attach RNAs, followed by a glycine-rich C-terminus. It has been indicated that the attendance of a variant in the regulatory region of *HNRNPA0* is related to increased cancer occurrence [30]. Furthermore, it is a powerful regulator for the growth of cancer cells [31]. However, it was downregulated in ATLL_smoldering, so it should be investigated in subsequent studies.

Conclusion

In this study, we found the genes that likely have significant roles in the development of three ATLL subtypes. The outcomes indicate the implication of MYH10 and P2RY1 in ATLL_acute, CBX6, CNKSR1, and MAX in ATLL_chronic, C22orf46 and HNRNPA0 in ATLL_smoldering. However, further experimental investigations in a large sample number are required to validate these potential gene regulators.

Materials and methods

Datasets and preprocessing

We downloaded the microarray datasets GSE33615 [32] and GSE55851 [33] containing the gene expression values of the ATLL samples as well as GSE29312 [34] and GSE29332 [34] comprising the expression amounts of genes in the AC subjects from the database Gene Expression Omnibus (GEO). The samples had taken from whole blood or peripheral blood mononuclear cells (PBMCs). The expression data of each condition were merged, individually. On the whole, 23, 29, and 10 samples from the ATLL individuals with chronic, acute, and smoldering subtypes, respectively, and also 37 ACs samples comprising the expression of 14,887 genes were used for further analysis. The feasible batch effects were excluded by employing the function of removeBatchEffect in the Limma package [35]. The data were also log2-transformed and quantile normalized.

Identification of differential co-expressed modules

Gene co-expression analysis finds groups of genes that are expressed in a concordant way and may regulate similar biological processes. On the other hand, identifying the different gene groups between two conditions could be utilized to compare them. Therefore, differential co-expression analysis aims to find the correlated gene groups in a specific condition that are not correlated in the other condition [7]. Differential co-expression may reveal the rewiring of transcriptional networks in reply to various subtypes of a disease. To identify the differential co-expressed gene clusters (modules) between two states, the DiffCoEx algorithm was applied [13]. DiffCoEx is based on the weighted gene

co-expression analysis (WGCNA) [36] and executed as follows: (i) construction of an adjacency matrix containing Pearson correlations between each gene pair; (ii) calculation of the squared correlation coefficients and then the construction of an adjacency difference (d_{ij}) matrix, in which the greater d_{ij} amounts denote the remarkable co-expression alterations between two genes. Each element of d_{ij} reaches the power β which is defined so that the network pursues a scale-free topology [15]. Then d_{ij} values reach the power of β; (iii) computing the topological overlap dissimilarity matrix (dissTOM) to find the common genes with joint neighbors; (iv) the hierarchical clustering is then carried out with the "flashClust" function [16]. Then, "dynamic-TreeCut" function is applied to find modules from the constructed dendrogram, and "mergeCloseModules" function is used to merge near modules; (v) the statistical significance of differential co-expression is assessed utilizing the dispersion statistic to express the correlation change between two states. In this study, we determined the differential co-expressed modules between ACs and ATLL-acute, ACs and ATLL_chronic, ACs and ATLL_smoldering.

Determining differentially expressed genes and pathway enrichment analysis

To identify differentially expressed genes between ACs and ATLL subtypes, the Limma package (version 3.54) was applied in R environment. Benjamini–Hochberg FDR adjusted p values < 0.05 were considered to determine DEGs [37]. For pathway enrichment analysis of the specific modules for each subtype, the ToppGene database was employed [38]. The terms with p value < 0.05 were determined as statistically remarkable.

Support vector machine-recursive feature elimination with cross-validation

SVM is a supervised classification machine-learning algorithm that has been widely used to classify various subtypes of a disease. SVM detects a hyperplane to detach the inputs into individual groups. Support vectors are used to identify the optimum hyperplane. The points near the hyperplane are considered support vectors. The output hyperplane is one that has the maximum distance from support vectors. Generally, the output of the linear function is captured in SVM [39]. In this study, we employed SVM with the tenfold cross-validation (CV) to find the most significant classifier for various ATLL subtypes [6, 40]. The features were the significant co-expressed genes that had been determined by DiffCoEx algorithm and DEGs. In addition, the recursive feature elimination that is a wrapper feature selection algorithm was used to determine the feature's classifiers. This algorithm in association with SVM (SVM-RFE) is

executed so that the top-ranked variables are selected as the most appropriate conditional variables on the given ranked subset in the model [41]. The top-ranked variables in the ultimate iteration of SVM-RFE are the essential informative features and the bottom-ranked variables are the unimportant variables that could be excluded. The SVM-RFECV is executed as follows: (i) training dataset by the SVM; (ii) arranging the features utilizing the weights of the attained classifier; (iii) removing the features with the lowest weight; (iv) updating the training dataset based on the selected features; (v) recurring the stages with the training set confined with the remaining features. We employed SVM- RFECV in Python 3.9.

Declarations

References

1. Ashrafi F, Ghezeldasht SA, Ghobadi MZ (2021) Identification of joint gene players implicated in the pathogenesis of HTLV-1 and BLV through a comprehensive system biology analysis. Microb Pathog 160:105153

2. Hajj HE, Nasr R, Kfoury Y, Dassouki Z, Nasser R, Kchour G et al (2012) Animal models on HTLV-1 and related viruses: what did we learn? Front Microbiol 3:333

3. Oshiro A, Tagawa H, Ohshima K, Karube K, Uike N, Tashiro Y et al (2006) Identification of subtype-specific genomic alterations in aggressive adult T-cell leukemia/lymphoma. Blood 107(11):4500–4507

4. Shimoyama M (1985) Adult T-cell leukemia/lymphoma and its clinical subtypes from the viewpoints of viral etiology. Human T-Cell Leukemia Virus. Springer, pp 113–125

5. Ghobadi MZ, Afsaneh E, Emamzadeh R, Soroush M (2023) Potential miRNA-gene interactions determining progression of various ATLL cancer subtypes after infection by HTLV-1 onco-virus. BMC Med Genomics 16(1):62. https://doi.org/10.1186/s12920-023-01492-0

6. Ghobadi MZ, Emamzadeh R, Afsaneh E (2022) Exploration of mRNAs and miRNA classifiers for various ATLL cancer subtypes using machine learning. BMC Cancer 22(1):1–8

7. Tesson BM, Breitling R, Jansen RC (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics 11(1):1–9

8. Chamasemani FF, Singh YP. Multi-class support vector machine (SVM) classifiers—an application in hypothyroid detection and classification. 2011 sixth international conference on bio-inspired computing: theories and applications: IEEE; 2011. p. 351–6.

9. Ghobadi MZ, Emamzadeh R, Mozhgani S-H (2021) Deciphering microRNA-mRNA regulatory network in adult T-cell leukemia/lymphoma; the battle between oncogenes and anti-oncogenes. PLoS One 16(2):e0247713

10. Zarei Ghobadi M, Emamzadeh R, Teymoori-Rad M, Mozhgani S-H (2021) Decoding pathogenesis factors involved in the progression of ATLL or HAM/TSP after infection by HTLV-1 through a systems virology study. Virol J 18(1):175. https://doi.org/10.1186/s12985-021-01643-8

11. Cai L, Fritz D, Stefanovic L, Stefanovic B (2010) Nonmuscle myosin-dependent synthesis of type I collagen. J Mol Biol 401(4):564–578

12. Betapudi V (2010) Myosin II motor proteins with different functions determine the fate of lamellipodia extension during cell spreading. PLoS One 5(1):e8560

13. Wang Y, Yang Q, Cheng Y, Gao M, Kuang L, Wang C (2018) Myosin heavy chain 10 (MYH10) gene silencing reduces cell migration and invasion in the glioma cell lines U251, T98G, and SHG44 by inhibiting the Wnt/β-catenin pathway. Med Sci Monit 24:9110

14. Jin J, Daniel JL, Kunapuli SP (1998) Molecular basis for ADP-induced platelet activation. II. The P2Y1 receptor mediates ADP-induced intracellular calcium mobilization and shape change in platelets. J Biol Chem 273(4):2030–2034. https://doi.org/10.1074/jbc.273.4.2030

15. Léon C, Hechler B, Vial C, Leray C, Cazenave JP, Gachet C (1997) The P2Y1 receptor is an ADP receptor antagonized by ATP and expressed in platelets and megakaryoblastic cells. FEBS Lett 403(1):26–30. https://doi.org/10.1016/s0014-5793(97)00022-7

16. Gremmel T, Yanachkov IB, Yanachkova MI, Wright GE, Wider J, Undyala VV et al (2016) Synergistic inhibition of both P2Y1 and P2Y12 adenosine diphosphate receptors as novel approach to rapidly attenuate platelet-mediated thrombosis. Arterioscler Thromb Vasc Biol 36(3):501–509

17. Yanachkov IB, Chang H, Yanachkova MI, Dix EJ, Berny-Lang MA, Gremmel T et al (2016) New highly active antiplatelet agents with dual specificity for platelet P2Y1 and P2Y12 adenosine diphosphate receptors. Eur J Med Chem 107:204–218

18. Tan Y, Zhang T, Zhou L, Liu S, Liang C (2019) MiR-34b-3p represses the multidrug-chemoresistance of bladder cancer cells by regulating the CCND2 and P2RY1 genes. Med Sci Monit 25:1323

19. Hua Y, Liu Y, Li L, Liu G (2023) Activation of hypermethylated P2RY1 mitigates gastric cancer by promoting apoptosis and inhibiting proliferation. Open Life Sci 18(1):20220078

20. Sellers LA, Simon J, Lundahl TS, Cousens DJ, Humphrey PP, Barnard EA (2001) Adenosine nucleotides acting at the human P2Y1 receptor stimulate mitogen-activated protein kinases and induce apoptosis. J Biol Chem 276(19):16379–16390

21. Deng H, Guan X, Gong L, Zeng J, Zhang H, Chen MY et al (2019) CBX6 is negatively regulated by EZH2 and plays a potential tumor suppressor role in breast cancer. Sci Rep 9(1):1–13

22. Wang J, He H, Jiang Q, Wang Y, Jia S (2020) CBX6 promotes HCC metastasis via transcription factors snail/zeb1-mediated EMT mechanism. OncoTargets therapy 13:12489

23. Comino-Méndez I, Leandro-García LJ, Montoya G, Inglada-Pérez L, de Cubas AA, Currás-Freixes M et al (2015) Functional and in silico assessment of MAX variants of unknown significance. J Mol Med 93(11):1247–1255

24. Ren Y, Bi C, Zhao X, Lwin T, Wang C, Yuan J et al (2018) PLK1 stabilizes a MYC-dependent kinase network in aggressive B cell lymphomas. J Clin Investig 128(12):5517–5530

25. Mathsyaraja H, Freie B, Cheng P-F, Babaeva E, Catchpole JT, Janssens D et al (2019) Max deletion destabilizes MYC protein and abrogates Eμ-Myc lymphomagenesis. Genes Development 33(17–18):1252–1264

26. Augert A, Mathsyaraja H, Ibrahim AH, Freie B, Geuenich MJ, Cheng P-F et al (2020) MAX functions as a tumor suppressor and rewires metabolism in small cell lung cancer. Cancer Cell 38(1):97–114 (**e7**)

27. Nishiyama K, Maekawa M, Nakagita T, Nakayama J, Kiyoi T, Chosei M et al (2021) CNKSR1 serves as a scaffold to activate an EGFR phosphatase via exclusive interaction with RhoB-GTP. Life Sci Alliance 4(9):e202101095

28. Wang L, Liu X, Yue M, Liu Z, Zhang Y, Ma Y et al (2021) Identification of hub genes in bladder cancer based on weighted gene co-expression network analysis from TCGA database. Cancer Reports 5(9):e1557

29. Li W, Liu R, Wei D, Zhang W, Zhang H, Huang W et al (2020) Circular RNA circ-CCAC1 facilitates adrenocortical carcinoma cell proliferation, migration, and invasion through regulating the miR-514a-5p/C22orf46 axis. Biomed Res Int 2020:3501451. https://doi.org/10.1155/2020/3501451

30. Wei C, Peng B, Han Y, Chen WV, Rother J, Tomlinson GE et al (2015) Mutations of HNRNPA0 and WIF1 predispose members of a large family to multiple cancers. Fam Cancer 14(2):297–306

31. Konishi H, Fujiya M, Kashima S, Sakatani A, Dokoshi T, Ando K et al (2020) A tumor-specific modulation of heterogeneous ribonucleoprotein A0 promotes excessive mitosis and growth in colorectal cancer cells. Cell Death Dis 11(4):1–16

32. Yamagishi M, Nakano K, Miyake A, Yamochi T, Kagami Y, Tsutsumi A et al (2012) Polycomb-mediated loss of miR-31 activates NIK-dependent NF-κB pathway in adult T cell leukemia and other cancers. Cancer Cell Int 21(1):121–135

33. Fujikawa D, Nakagawa S, Hori M, Kurokawa N, Soejima A, Nakano K et al (2016) Polycomb-dependent epigenetic landscape in adult T-cell leukemia. Blood J Am Soc Hematol 127(14):1790–1802

34. Tattermusch S, Skinner JA, Chaussabel D, Banchereau J, Berry MP, McNab FW et al (2012) Systems biology approaches reveal a specific interferon-inducible signature in HTLV-1 associated myelopathy. PLoS Pathog 8(1):e1002480

35. Zarei Ghobadi M, Emamzadeh R (2022) Integration of gene co-expression analysis and multi-class SVM specifies the functional players involved in determining the fate of HTLV-1 infection toward the development of cancer (ATLL) or neurological disorder (HAM/TSP). PLoS One 17(1):e0262739

36. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9(1):559

37. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 57(1):289–300

38. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids Res 37(suppl_2):W305–W311

39. Afsaneh E, Sharifdini A, Ghazzaghi H, Ghobadi MZ (2022) Recent applications of machine learning and deep learning models

in the prediction, diagnosis, and management of diabetes: a comprehensive review. Diabetol Metab Syndrome 14(1):1–39

40. Sanz H, Valim C, Vegas E, Oller JM, Reverter F (2018) SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinform 19(1):1–18

41. Zarei Ghobadi M, Emamzadeh R, Teymoori-Rad M, Afsaneh E (2022) Exploration of blood-derived coding and non-coding RNA diagnostic immunological panels for COVID-19 through a co-expressed-based machine learning procedure. Front Immunol 13:1001070