In silico study of dynamic expression profile of mRNA Export Factors promoting progression of Diffuse large B-cell lymphoma

Sahni Deepak, Pandey Alok, Sharma Chandresh* and Bharati Akhilendra Pratap*
Department of Biotechnology, School of Life Sciences and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur-208024, INDIA
*chandreshsharma@csjmu.ac.in; akhilendrapratap@csjmu.ac.in

Abstract

Diffuse large B-cell lymphoma (DLBC) is the most common blood cancer in India, comprising of 30-40% of all lymphomas. It is a subtype of non-Hodgkin lymphoma (NHL), with approximately 75% recurrence cases. Current treatments involve chemotherapeutics such as Rituximab, Cyclophosphamide and Doxorubicin. However, frequent relapse highlights the need for alternative therapeutic strategies. mRNA export factors (MEFs) play critical roles in mRNA transport from the nucleus to the cytoplasm, facilitating translation in eukaryotes. While MEFs have been studied individually in cancers, the broader link between MEFs and neoplasia remains underexplored. Our results indicate that the MEF genes (THOCs, DDX19B, RAE1) exhibited significant expression differences in DLBC compared to normal tissues. We also found the positive correlation of most of the MEFs with that of mentioned markers associated with DLBC.

In addition, hyper-methylation of the THOC5 promoter (beta value >2.5) was observed in p53-mutant cases while THOC2 promoter hypermethylation (beta >0.3) was more frequent in females. Age-related methylation changes were noted for THOC7. Finally, we observed that the higher expression of MEFs (THOC2 and DDX19B) was linked to poor survival in Asian populations as compared to the White population. These findings highlight the role of MEFs as potential prognostic markers and could be used for assessment of DLBC progression, offering promising avenues for improving DLBC management.

Keywords: MEFs, DLBC, NHL, GTEx, TCGA, GEPIA2, UALCAN.

Introduction

Diffuse large B-Cell Lymphoma (DLBC) is most prevalent constituting almost one-third of all cases globally[34]. It is the most common blood cancer in India. The prevalence rate of DLBC is approximately 30-40% among all types of lymphomas[22] and it belongs to non-Hodgkin lymphoma (NHL), a cancer that originates in the lymphatic system[35] with about 75% recurrence cases. It is the most common form of aggressive lymphoma distinguished by its quick growth and metastasis. Most prevalent in older people, with a typical diagnostic age of roughly 60-65 years, it is and more prevalent in men than in women while it can also strike younger people[24]. DLBC risk factors include immunosuppression (e.g. HIV/AIDS, organ transplant recipients), autoimmune disorders and chronic infections (e.g. Helicobacter pylori in stomach DLBC, Epstein-Barr virus)[41].

The signs and symptoms may include fevers, sweats at night, unintentional weight loss, or exhaustion, or they may be due to swollen lymph nodes, breathlessness or pain in the chest (the chest or mediastinal nodes are affected by lymphoma). Splenomegaly or swollen lymph nodes are less frequently unintentional observations made while evaluating for other medical conditions[15]. Due to its heterogeneity, DLBC is divided into subgroups according to morphology and gene expression profiles. By using gene expression profiling, two molecular subtypes of DLBC have been identified: germinal-center B cell-like (GCB) generally associated with a better prognosis and activated B cell-like (ABC) associated with poorer prognosis compared to GCB[16].

These subtypes are lymphomas that originate from various lymphoid differentiation stages (cell-of-origin) and are caused by various oncogenic processes[4]. MYC and BCL2 and/or BCL6 gene rearrangements are characteristics of double-hit and triple-hit lymphomas, which are linked to an aggressive clinical course[29]. DLBC accounts for 25-30% of all lymphomas including Hodgkin and non-Hodgkin types. Approximately 30,000 to 35,000 new cases of NHL are recorded in India each year, with DLBC accounting for between 9,000 to 14,000 of these instances. This represents a significant health burden[24,28]. DLBC accounts for about 30-40% of NHL cases in Western countries, while its proportion varies greatly by region in Asia[29].

There are many known markers of DLBC, some of them are CD274, ALK, CD40, FOXM1 and some associative markers of neoplasia like p53, MYC and KRAS[23]. R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone) are chemotherapy drugs that have been utilized so far to stop the growth and spread of DLBC. DLBC relapses are rather common despite all of these indicators and chemotherapy regimens. Up to 45% to 50% of patients will relapse despite the safety and effectiveness of this approach. Continuing research into the transcriptome and genetic landscape of DLBC has shown subgroups of individuals who have a poor prognosis for chemo-immunotherapy[35].

Relapses are common emphasizing the necessity for different approaches for treatment. In eukaryotes, mRNA export factors (MEFs) are essential for the movement of mRNA from the nucleus to the cytoplasm[31]. A crucial stage in gene expression, this transport makes sure that mRNAs are accessible in the cytoplasm for protein translation. Numerous mRNA export factors that interact with the nuclear pore complex (NPC), a large protein structure that extends across the nuclear envelope, mediate this process[6]. The wider connection between MEFs and neoplasia is still poorly understood, despite the fact that MEFs have been examined separately in malignancies.

MEFs also play a very important role in many diseases like osteogenesis imperfecta type I, motor neuron diseases, viral infections and in cancer, where a particular single factor is targeted but there is a research gap relating all MEFs and neoplasia. A group of specialized proteins that attach to mature mRNA molecules and help them pass through the NPC, are mostly responsible for the nuclear export of mRNA (Figure 1). The main export pathway is Mex67/MTR2-NXF1/NXT which identifies and attaches to messenger ribonucleoprotein particles (mRNPs) that contain processed mRNAs[13,33].

CRM1 (for specialized mRNAs) and NXF1 (for bulk mRNA export) are the important proteins to ensure that only processed mRNAs are exported, they engage with complexes like TREX Complex (Transcription-Export Complex) which is a multi-protein assembly coupling transcription, splicing and export consisting of THO Subcomplex (THOC1-7) that coordinates with UAP56/DDX39B to bind mRNA during transcription and adapters like ALYREF (ALY) to build mRNA-protein complexes (mRNPs)[10]. Other factors like GANP (germinal center-associated nuclear protein) and ALYREF are also important because they promote the selective export of certain mRNA subsets, ensuring that only processed mRNAs are exported, thereby preserving cellular homeostasis[11,43].

The importance of mRNA export dysregulation in cancer biology is demonstrated by the fact that it has been linked to several carcinomas. The function of several mRNA export factors is implicated in different kinds of carcinomas[3,45,47]. Overexpression of NXF1 may promote the formation of lymphomas by increasing the export of oncogenic mRNAs such as MYC and BCL[66]. The export of transcripts linked to cell division, immune evasion and apoptosis can be changed by dysregulation of THOC components[6,11]. Dysregulated TREX components disrupt apoptosis and immune evasion gene expression[19]. TREX disruption also leads to Intellectual Disability Syndrome[42]. THOC1, THOC2, or THOC5 mutations can disrupt the coupling of transcription and export, leading to genomic instability or aberrant gene expression[18]. In many carcinomas, the expression and function of mRNA export factors are altered, leading to aberrant mRNA export.

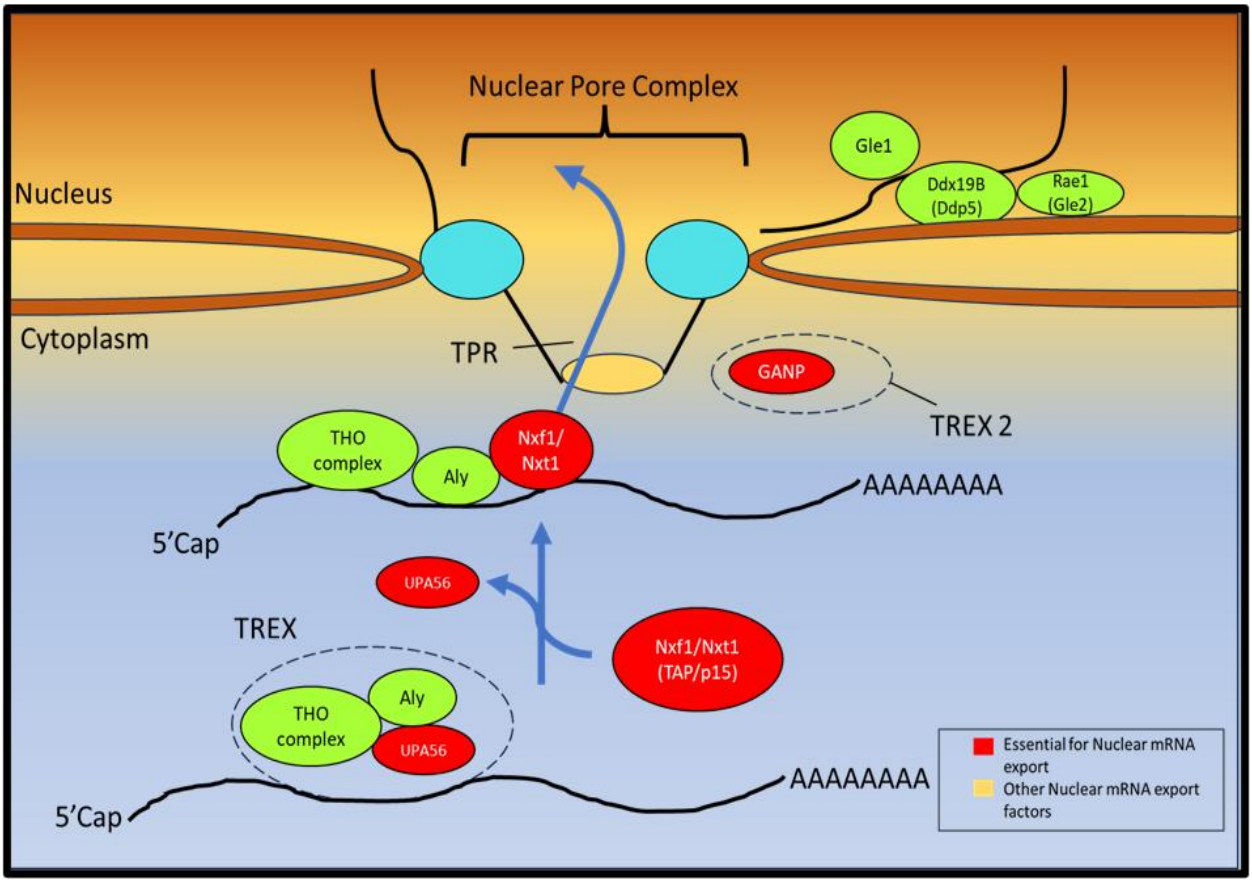

Figure 1: Mechanism of mRNA export.

For instance, ZC3H11A has been identified as an mRNA export factor whose expression is significantly elevated in various cancers including gliomas. High levels of ZC3H11A correlate with poor prognosis, suggesting its role as a potential biomarker for tumor aggressiveness[1]. Similarly, XPO1 (Exportin-1) has been shown to be overexpressed in lymphoma cells, enabling adaptive regulation of mRNA export under genotoxic stress conditions. This adaptation facilitates DNA repair processes critical for cell survival in the face of genomic instability[25]. ZC3H11A is essential for regulating mRNA export pathways in gliomas. Poor clinical outcomes and higher tumor cell proliferation have been associated with elevated levels of this protein. Because ZC3H11A is dysregulated, therapeutic approaches for treating gliomas may involve focusing on this factor[21].

XPO1 is essential for preserving mRNA export in lymphoma cells when they are under stress. By exporting mRNAs encoding DNA repair proteins, its overexpression helps these cells effectively handle genotoxic stress. It has been demonstrated that inhibiting XPO1 increases susceptibility to drugs that damage DNA, suggesting that it may be a promising therapeutic target[25]. The study has reported the expression profile of MEFs in DLBCL as well as explored the pan cancer analysis of MEFs. Furthermore, we have done correlation of the factors with that of the markers associated with the DLBC. We have further analysed the high and low expression of the factors associated with the survival rate of the patients using overall survival and disease-free survival plots. We have also analysed the gene ontology using DAVID databases and interaction analysis of MEFs with other markers and interacting partners, understanding the role that mRNA export factors play in the development of cancer brings up new treatment options. By focusing on certain export factors, such as ZC3H11A and XPO1, current treatments may be more effective, or new therapeutic approaches that restore normal RNA processing and export pathways may be developed (Figure 2).

Material and Methods

Assessment of expression profile of different MEFs and associated survival in DLBC: We obtained quantification expression data of 58 patients suffering from DLBC from The Cancer Genome Atlas (TCGA) project (https://www.cancer.gov/tcga)[39]. and same number of normal samples as a control from the Genotype-Tissue Expression (GTEx) project (https://www.gtexportal.org/home/). In order to thoroughly investigate the expression pattern of our gene of interests that belong to mRNA export factors (MEFs) family in case of DLBC, evaluate the expression profile of MEFs in DLBC.

MEFs are as follows: THOC1, THOC2, THOC3, THOC5, THOC6, THOC7, ALREF, DDX19B, ZC3H14, GLE1, NUP214, NUP98, RAE1 and NXF1. Both a log scale and raw read counts expressed in transcripts per million (TPM) were used to quantify the expression data. To compress the data and to make it more representable, fold change for each MEF is calculated by dividing the expression count in DLBC with the specific count in normal case, which provides us the actual increase in expression profile.

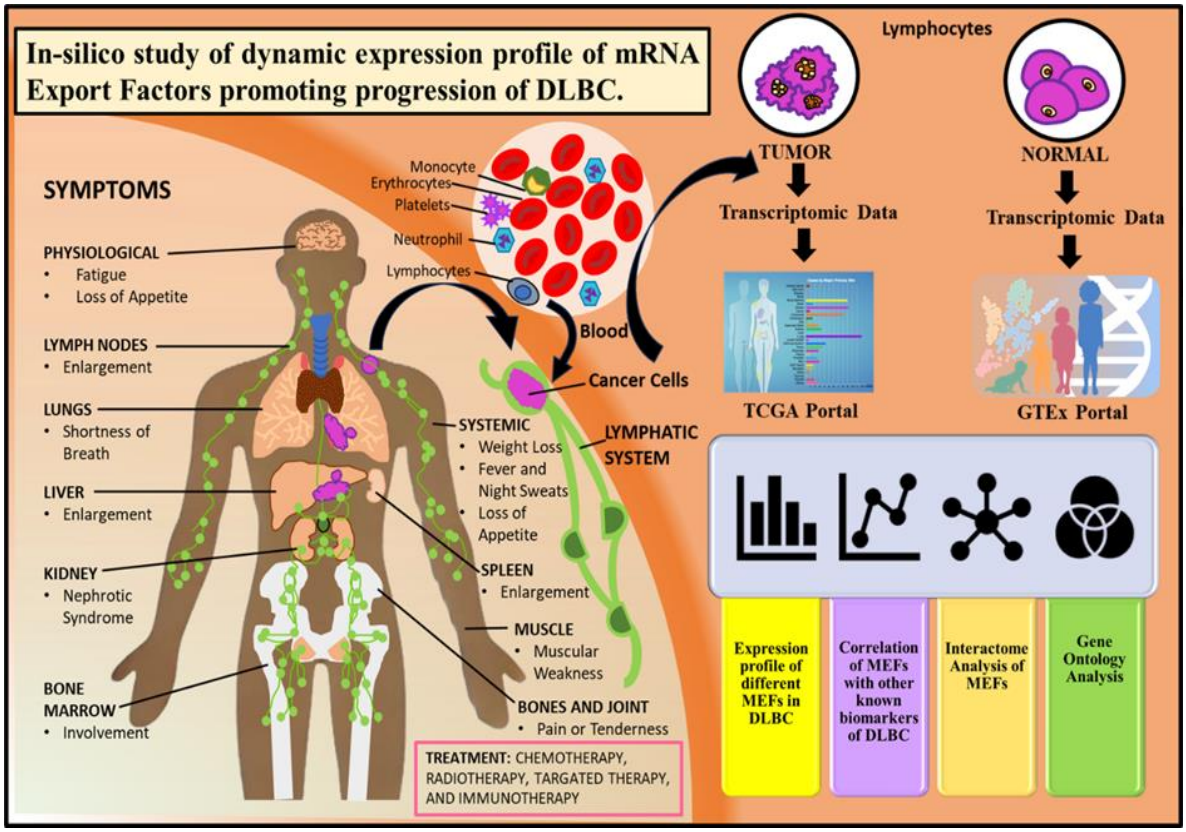

Figure 2: The graphical abstract of the manuscript

To find out the associated survival of DLBC patients with respect to change in expression profile of MEFs, we have generated survival maps using GEPIA2 tool (gepia.cancerpku.cn)[37,38], which extracts datasets from the TCGA database. Using this tool, data was plotted as overall survival and disease-free survival maps.

Assessment of Correlation of different MEFs with other known biomarkers: Spearman Correlation method was used to assess the relationship between different MEFs expression level and other DLBC-specific known biomarkers (CD274, ALK, CD40 and FOXM1) and other genes involved in tumor progression (KRAS, c-MYC and p53) that validate more on relatedness of MEFs with DLBC. For analysis, differentially expressed genes (MEFs in this case) in DLBC were gathered from GEPIA2 (gepia.cancerpku.cn)[37,38].

Methylation status of promoter of different MEFs based on different variables: To find out the methylation status of the promoter of different MEFs (THOC2, THOC5, THOC6 and THOC7), we used the UALCAN tool[8,9]. The beta value in this case, revealed the actual methylation ranging from 0 (unmethylated) to 1(fully-methylated). The specific cut-off for hypomethylation is 0.25 to 0.3 whereas for hyper-methylation it is 0.5 to 0.7. In this, beta value or level of methylation is one variable and other variables are different parameters (p53 mutation, obesity, sex and age group of the DLBC patients) that we are considered for specific MEF.

Race-based overall survival analysis and MEF-specific survival analysis: Using the GDC tool, the race dependent survival plot was generated. Many MEFs are also affecting the survival of DLBC patients, which is race dependent. In this way, we have conducted 2-variate analysis using the UALCAN tool[8,9], considering specific MEF as one variable and race as the second one. In this way we concluded how the race and expression profile of different variables affect the survival of DLBC patients.

Mutation profile of different MEFs in DLBC: By splitting the DLBC patient population into two different cohorts (Asian and White); using the GDC portal, we have segregated the mutation profile of patients. There were a total of 18 cases in the Asian cohort and 29 cases of White. We have separately analysed both the cohorts and found specific mutations.

This analysis could be a stepwise 2-variate analysis, first we divided the total population based on race and, secondly, looked for mutations in separate cohorts and finally considered it under the same umbrella of DLBC. Allele frequency of the specific mutations was extracted from cBioportal[7,12,17] which also uses TCGA[39] database.

Pathway enrichment analysis: Using TCGA[39], GEPIA2[37,38] and DAVID (Database for Annotation, Visualization and Integrated Discovery)[20,30], gene ontology enrichment analysis was performed with upregulated genes in TGCT.

Assessment of Direct and Indirect Interaction partners of MEFs: The Search Tool for the Retrieval of Interacting Genes/Proteins database (STRING, version 11.0) was utilized to find functional relationships between certain genes, their co-expression, gene enrichment and functional analysis. Based on anticipated product interactions, the STRING library connects genes in both direct (physical) and indirect (functional) ways[36].

Results and Discussion

Expression profile of different MEFs and associated survival in DLBC: The expression profile of mRNA export factors was calculated using datasets from TGCA. Considering the expression profiles of normal samples and patients with DLBC, we have calculated fold change for proper comparative analysis to analyse the actual change. We have calculated the fold change by dividing the expression in transcripts per million with that of normal count. We have plotted the bar diagram depicting the whole dynamic expression profile as shown in figure 3A. The data represent that THOC3 was having the highest fold change of 15.7968254, whereas NUP214 was the only one to be downregulated with fold change of -1.440256616. The pan cancer expression profiles of different MEFs were also performed using datasets from TGCA for different carcinomas samples and GTex for normal tissue expression profile.

Furthermore, the overall survival (OS) of the patients with high expression levels of different MEFs in DLBC was calculated. Heatmap of overall survival, as shown in figure 3B is depicting the hazard ratio (log10HR) which is directly proportional to the survival of patient at higher expression of MEFs (THOC1, THOC2, THOC3, THOC5, THOC6, THOC7, ALREF, DDX19B, ZC3H14, GLE1, NUP214, NUP98, RAE1 and NXF1) in DLBC. In this, high expression of DDX19B, THOC2, GLE1, NUP98 and ZC3H14 is associated with poor survival, while rest of the mentioned MEFs do not show any relevant effect on patient survival. Particularly, ALYREF and THOC5 are associated with better survival, which may be triggering some anti-tumor pathways[3,46,48].

Moreover, the ALYREF-MYC feedback pathway regulates proliferation of cells in Glioblastoma[40]. Disease-Free Survival (DFS) analysis of the same MEFs depicting the survival of patients without any sign and symptom of tumor was also performed (Figure 3C). The result indicates the highest hazard ratio in case of ZC3H14, indicating high levels of ZC3H13 is associated with poor survival. If we compare the data of OS and DFS, the data indicates that the HR ratio is even higher in case of DFS as compared to that of OS. This clearly deciphers the most probable reason for relapse of DLBC due to altered expression of MEFs associated with poor survival and early death.

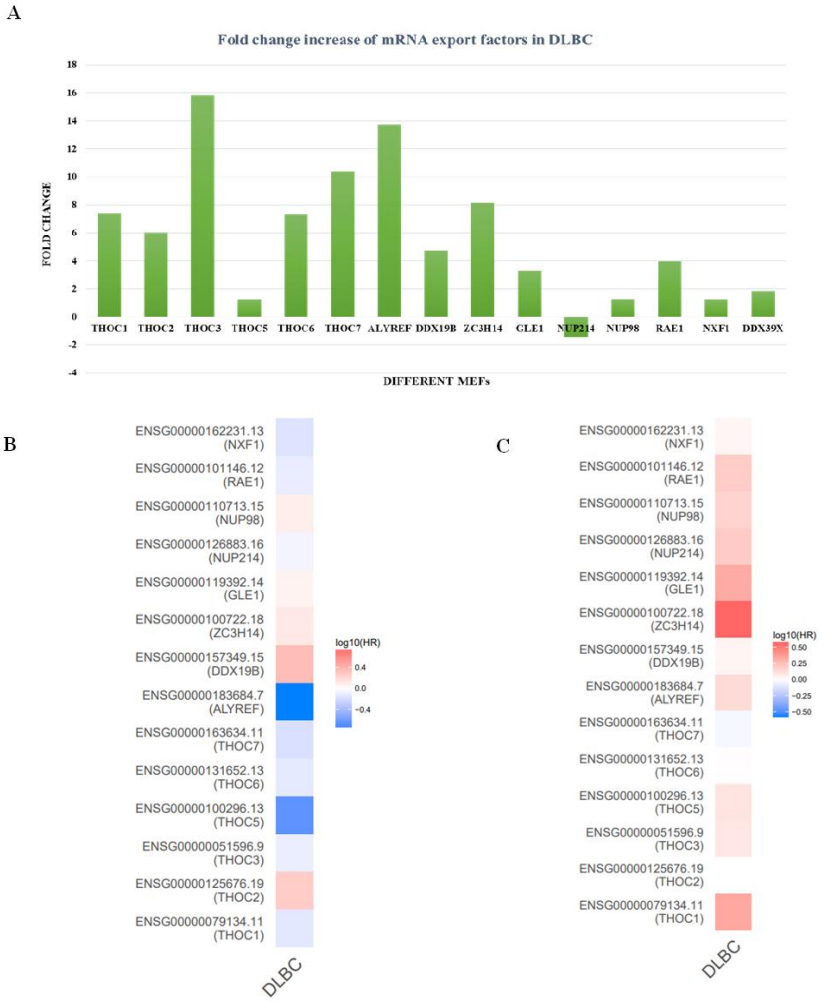

Figure 3: Expression Profile of MEFs and associated survival. (A) Fold Change Expression analysis of different MEFs in DLBC (Fold change calculated by dividing with normal expression level), (B) Overall Survival Maps, (C) Disease Free Survival Map. Heat map shows hazard ratios (HR) in log10 for respective genes. The red and blue blocks denote higher and lower risks respectively.

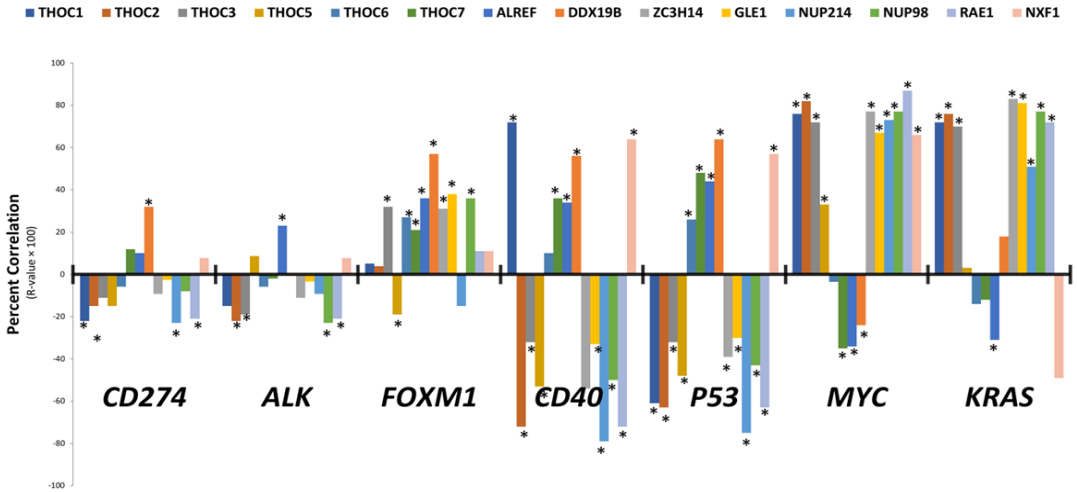

Figure 4: Correlation Analysis: Spearman Correlation method was used to assess the relationship between 14 different MEFs (THOC1, THOC2, THOC3, THOC5, THOC6, THOC7, ALREF, DDX19B, ZC3H14, GLE1, NUP214, NUP98, RAE1 and NXF1) expression level and other DLBC-specific known biomarkers (CD274, ALK, CD40 and FOXM1) and other genes involved in tumor progression (KRAS, c-MYC and p53). The clustered bar graph is showing combined results of 98 correlations. All asterisks' marks represent significant correlation having p-value less than 0.05.

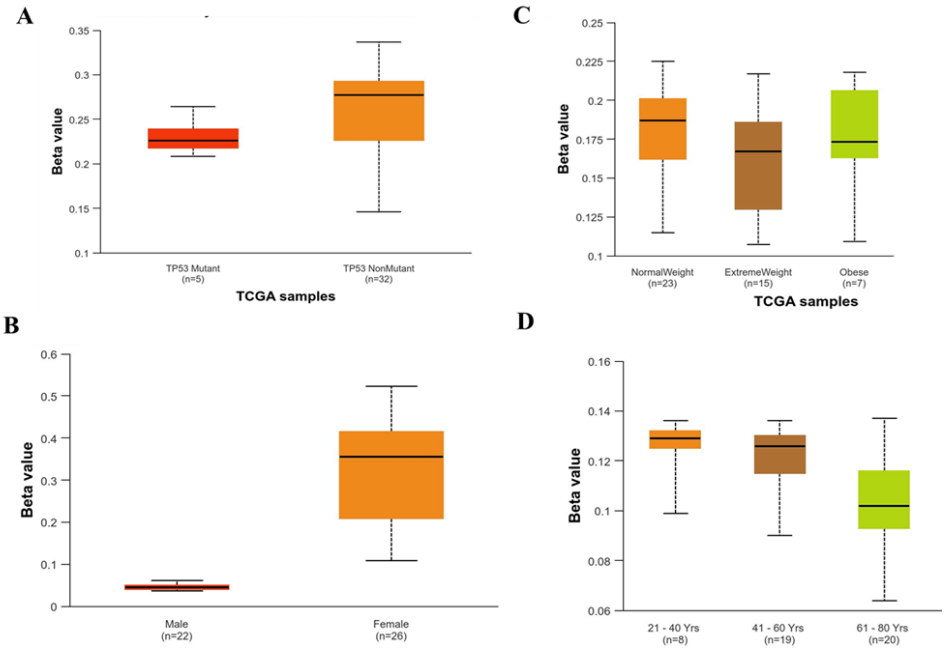

Figure 5: Methylation Status of different MEFs promoter based on different parameters. (A) Promoter methylation level of THOC5 in DLBC: Based on p53 mutation. (B) Promoter methylation level of THOC6 in DLBC: Based on sex of the DLBC patients. (C) Promoter methylation level of THOC2 in DLBC: Based on body weight of the DLBC patients. (D) Promoter methylation level of THOC7 in DLBC: Based on different age groups.

Correlation of different MEFs with other known biomarkers related to DLBC: The correlation of the candidate genes (MEFs) with already known biomarkers (CD274, ALK, CD40 and FOXM1) and other genes related to tumor proliferation (p53, KRAS and c-MYC) was studied. We performed a Spearman Correlation test to assess the relationship between the expression levels. The value of coefficient of correlation (R) lies between +1 and -1 and 0 (in case of no association). As shown in figure 4, it is crystal clear that the majority of MEFs are positively correlated with the other markers (CD274, ALK, CD40, FOXM1, p53, KRAS and c-MYC).

It is quite evident from the cluster bar graph (Figure 4) that the majority of MEFs are positively associated with the tumor promoting genes (KRAS and c-MYC) and with guardians of the genome (p53). Positive correlation may be indicative of stimulation and/or co-expression whereas negative correlation may be part of any feedback mechanism or inhibition. Correlation of MYC and ALYREF drives positive feedback which promotes cell proliferation[40].

Methylation status of promoter of different MEFs: We found significant variation in promoter methylation of different MEFs (THOC2, THOC5, THOC6 and THOC7) based on variables or filters opted in each case. The data was considered significant based on the beta value, which is indicative of level of methylation.

Generally, hypermethylation of promoters resulted in altered expression patterns (generally reduced expression), which could be modulator of any disease[2,26]. In our study, we found a specific methylation variation pattern between the cohort groups (based on variables like p53 mutation, sex, age and obesity) (Figure 5).

The first cohort/group was represented by the DLBC population with and without p53 mutation. We observed the group with p53 mutation is having less methylation as indicated by lower beta value, with significance of 2.873600E-02 (Figure 5A). Less methylated promoter of THOC5 would favour high expression of THOC5 which may lead to progression of DLBC. Moreover, mutated p53 would also add to the disease progression, resulting in a cumulative effect of both the factors. The second cohort, based on sex of the patient, observed that there was significant difference in the promoter methylation pattern of THOC2.

In this case, females have hypomethylated promoters as compared to males as the median of beta value is 0.356 (Figure 5B). This indicates that males are more prone to DLBC, though it is also reported in the epidemiological studies[14,44]. Similarly, when the DLBC population was divided based on body weight, there was significant change observed in promoter methylation of THOC6. On comparing the statistical significance between the normal weight vs extreme weight, we got the p-value of 4.930300E-02 (Figure 5C). Our analysis suggests that obesity is associated with less promoter methylation of THOC6 and it is also characterized by epidemiological study[14].

Dividing the DLBC population into different age groups and getting the promoter methylation profile of THOC7 was also supported by Epperla et al[14]. In this, we observed that 41-60 yrs of age group were more methylated as compared to that

of 61-80 yrs age group cohort (Figure 5D). On comparing the statistical significance between age (41-60Yrs) vs age (61-80 Yrs), we got a p-value of 3.486400E-02, which is quite significant.

Race-based overall survival analysis and MEF-specific survival analysis: From the genetic point of view, it is very important to conduct race-based analysis coupled with a specific MEF expression portfolio. The GDC portal was used to generate cohorts based on race, in accordance with the datasets available in the TCGA portal. There were 2 groups created to carry out overall survival of DLBC patients based on race (Figure 6 A). First cohort belongs to Asians with 18 cases (Males =8 and Females=10) and second was Whites with 29 cases (Males =14 and Female =15). Surprisingly, the Asian population was having poor survival as compared to that of White population. It is quite evident from figure 6B that the survival plot of White population was 4 times better from Asians, which makes Asians more prone to DLBC.

We further investigated the effect in DLBC patient survival rate in different races (Figure 6C). We observed that high expression of DDX19B in the Asian population is associated with poor survival while low/median expression is associated with comparatively better survival. On the contrary, Caucasians were having better survival at high as well as low/medium expression. The whole datasets were significant having p value of 0.048. Following the same trend, the survival plot of THOC2 with respect to different races was analysed (Figure 6D). The result indicates the poor survival of the Asian population as compared with the White as in case of DDX19B with a significant p-value score of 0.0042 (Figure 6D). Unlike DDX19B and THOC2, THOC5 and THOC7 low/medium expression profile is associated with poor survival in Asians and better survival in Caucasians (Figure 6E and F).

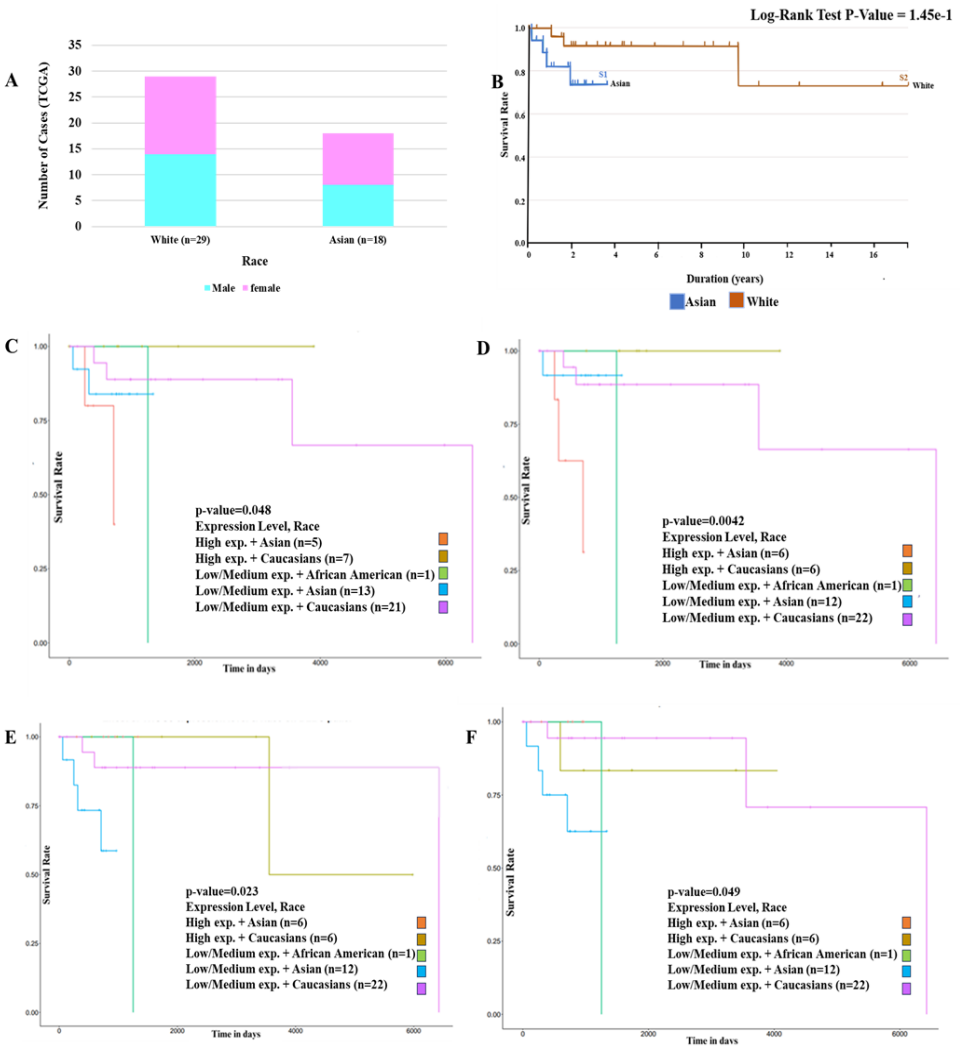

Figure 6: Race-based expression profile of different MEFs. (A) Case Distribution of DLBC based on Race and gender: Represents the no. of male and female patients across two cohorts; Asians (Male = 8 and Female = 10) and Whites (Male =14 and Female = 15). (B) Overall Survival based on Race (generated using GDC portal). (C) Effect of DDX19B expression level and Race on DLBC patient survival. (D) Effect of THOC2 expression level and Race on DLBC patient survival. (E) Effect of THOC5 expression level and Race on DLBC patient survival. (F) Effect of THOC7 expression level and Race on DLBC patient survival.

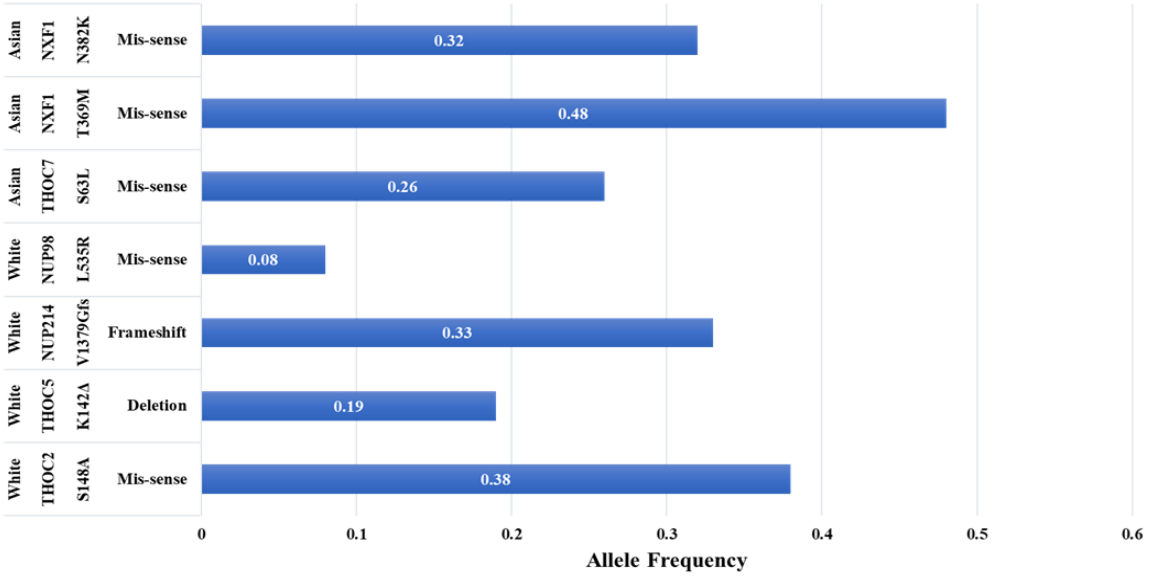

Figure 7: Race-based mutations in DLBC: Using GDC portal race-based cohorts were generated for the analysis and cBioportal was used to find out specific Allele frequency. In the bar graph the bar represents Allele frequency and the base provide details of specific cohort and mutations in protein sequence of specific MEF mentioned alongside.

Mutation profile of different MEFs in DLBC: Using the GDC portal, we have created similar race-based cohorts and analysed the mutations related to different MEFs. There were no common mutations between White and Asian subgroups of DLBC patient cases. In the case of White Cohort, we observed 5 different types of mutation in different MEFs (THOC2, THOC5, NUP214 and NUP98) as shown in figure 7. In Asians, we observed 4 different mutations as mentioned in table 1. In White population, the mutation was detected in THOC2, THOC5, NUP214 and NUP98 whereas the frequency of THOC2 mutation was higher (0.38) as compared to other mutations. Similarly, 3 mutations in two different factors (THOC7 and NXF1) were observed in the case of the Asian population. Two different mis-sense mutations were observed only in the NXF1 with the highest frequency (0.48 and 0.32).

Pathway enrichment analysis of MEFs: Using TCGA[39] and GEPIA2[37,38], it was observed that all the MEFs are overexpressed in DLBC, except NUP214 as observed in figure 3A. Furthermore, using DAVID (Database for Annotation, Visualization and Integrated Discovery)[20,30] gene ontology enrichment analysis was done. Based on functional gene ontology, the upregulated genes have been categorized into 3 main divisions: (a) Involvement in Biological Processes, (b) role in molecular function and (c) part of cellular component. According to these 3 divisions, cluster bar graph was plotted with -log10 of p-value at X-axis and sub-category on Y-axis.

The upregulated genes (MEFs) in DLBC are specifically involved in mRNA transport, mRNA splicing, mRNA processing, translocation, protein transport, host-virus interaction, nuclear pore complex formation and some of them function as RNA binding proteins (Figure 8B).

Interactome Analysis of different MEFs with known markers: It is very important to decipher the signalling pathway of the candidate biomarker, so that we can optimize the point of inhibition for our targeted therapy involving our candidate genes which could be potential markers of DLBC. Known interactions provide us clues, connecting which we can generate signalling pathway maps. Using STRING, a molecular interaction map was built up for different MEFs and other known biomarkers. From the bitmap (Figure 9), it is quite evident that all MEFs are interacting with each other by direct interaction and also by some indirect means, all these contribute to the export of mRNA out of the nucleus. NUPs (NUP214 and NUP98) show interaction with RUNX1, which plays a role in the development of blood cells and is involved in several types of cancer and heart disease[5,27,32].

Moreover, NUP98 also interacts with PBX1, which is Pre-B-Cell leukemia transcription factor 1 involved in re-B-cell acute lymphoblastic leukemia. It is noteworthy that other MEFs (except NUPS) interact with each other and may also interact with RUNX1 through NUPS (as no direct bonding can be seen). Similarly, with PBX1, only NUP98 is interacting and other MEFs may be interacting indirectly through NUP98. We already know that NUPs are involved in leukemia and here interactive analysis further validates the story. All these connections direct the hypothesis that NUPS (NUP98 and NUP 214) initiate the chain of mis-management which further builds up with the contribution of other MEFs in case of DLBC. Other markers like (ZNF513, SERTAD2, GSR, ANGPTL1 and NIR6A1) did not show any interaction with MEFs which dictates that there is a big lacuna in between which should be filled in order to complete the picture under study. In order to fill the gap, more work needs to be done in this direction.

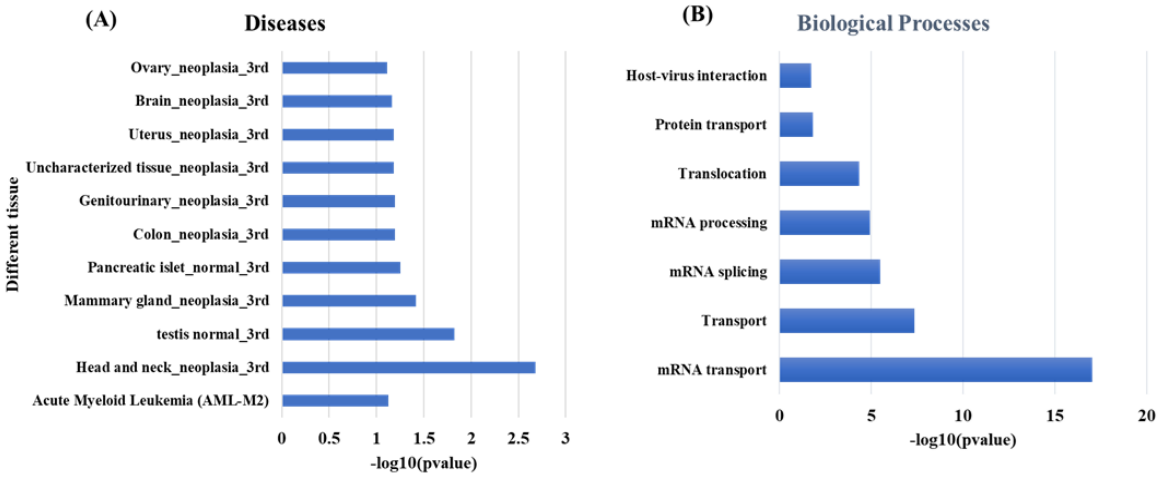

Figure 8: Gene Ontology Analysis. (A) Involvement of different MEFs in various diseases. (B) Involvement of various MEFs in different Biological Processes.

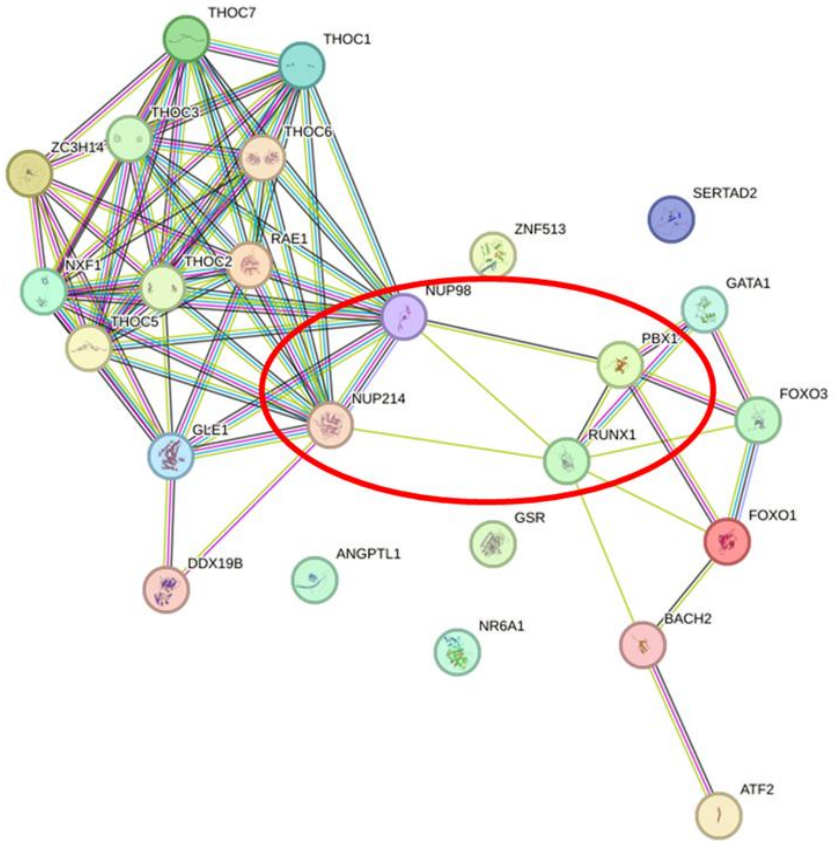

Figure 9: Interactome Analysis, depicting interaction of different MEFs with each other and encircled area shows the interaction of NUPS with transcription factors involved in leukemia (RUNX1 and PBX1).

Conclusion

DLBC is a type of cancer which is not explored much with respect to different MEFs. So far, DLBC is not studied based on race, mutation and promoter methylation level which adds novelty to the study. Moreover, targeting different MEFs in DLBC for diagnostic and therapeutic purposes would be of great impact and it would also open the new arena of personalized medicine by targeting the specific mutated MEF in specific patients. In this analysis of dynamic expression profiling, we found that THOC3 is having the highest fold change which is nearly about 16 times, such a rise in expression must be involved in some sort of signalling. In lung carcinoma, THOC3 is known to interact with YBX1 and promote the progression of tumor with the help of PFKBP4 mRNA modification[45]. Such findings are yet to be explored in DLBC, which will further decipher the reason for high expression of THOC3. For instance, we can consider THOC3 as a valid candidate to be considered as a

diagnostic marker in DLBC as fold change of 16 cannot be considered random, moreover the raw counts were in log10 of transcripts per million, here the digits are adding lots of significance to be considered. Comparing fold change profile and mutation data, NXF1 could also be the target molecule in the field of therapeutics.

It is noteworthy that NXF1 was the only candidate, having 2 separate mutations and highest allele frequency (that is 0.48) in the total DLBC population under study. It indicates that out of 100 Asian DLBC patients, 48 would have the specific point mutation (T369M) of NXF1. Correlations with factors promoting tumorigenesis (MYC and KRAS) also supports the candidature of selected MEFs (NXF1, THOC7, THOC3, NUP214, NUP98) to be considered as specific markers or targets of DLBC as discussed above. This analysis further deciphers the role of race, sex, age, body weight and promoter methylation profile of specific MEF in a 2-variate model to explore effect of two parameters in a single study.

Acknowledgement

Authors are grateful to DST-SERB (Grant No. EEQ/2022/ 000486) for the financial assistance. APB is grateful to EMEQ (Empowerment and Equity Opportunities for Excellence in Science) scheme of the DST-SERB (now ANRF) and DS is grateful to CSIR (Council of Scientific & Industrial Research) for the financial assistance.

References

1. Ali A., Contreras P., Darweesh M., Andersson L., Jin C., Essand M. and Yu D., Targeting ZC3H11A elicits immunogenic cancer cell death through augmentation of antigen presentation and interferon response, *Molecular Therapy Nucleic Acids*, **35(4) (2024)**

2. Arya A. K., Bhadada S. K., Singh P., Sachdeva N., Saikia U. N., Dahiya D. and Rao S. D., Promoter hypermethylation inactivates CDKN2A, CDKN2B and RASSF1A genes in sporadic parathyroid adenomas, *Scientific Reports*, **7(1)**, 3123 **(2017)**

3. Bai X., Ni J., Beretov J., Wang S., Dong X., Graham P. and Li Y., THOC2 and THOC5 regulate stemness and radioresistance in triple-negative breast cancer, *Advanced Science*, **8(24)**, 2102658 **(2021)**

4. Barraclough A., Hawkes E., Sehn L. H. and Smith S. M., Diffuse large B-cell lymphoma, *Hematological Oncology*, **42(6)**, e3202 **(2024)**

5. Bonifer C., Levantini E., Kouskoff V. and Lacaud G., Runx1 structure and function in blood cell development, *RUNX Proteins in Development and Cancer*, 65-81 **(2017)**

6. Borden K. L., The nuclear pore complex and mRNA export in cancer, *Cancers*, **13(1)**, 42 **(2020)**

7. Cerami E., Gao J., Dogrusoz U., Gross B. E., Sumer S. O., Aksoy B. A. and Schultz N., The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data, *Cancer Discovery*, **2**, 401-404 **(2012)**

8. Chandrashekar D. S., Bashel B., Balasubramanya S. A. H., Creighton C. J., Ponce-Rodriguez I., Chakravarthi B. V. and Varambally S., UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, *Neoplasia*, **19(8)**, 649-658 **(2017)**

9. Chandrashekar D. S., Karthikeyan S. K., Korla P. K., Patel H., Shovon A. R., Athar M. and Varambally S., UALCAN: An update to the integrated cancer data analysis platform, *Neoplasia*, **25**, 18-27 **(2022)**

10. Clarke B. P., Angelos A. E., Mei M., Hill P. S., Xie Y. and Ren Y., Cryo-EM structure of the CBC-ALYREF complex, *eLife*, **12**, RP91432 **(2024)**

11. Culjkovic-Kraljacic B. and Borden K. L., Aiding and abetting cancer: mRNA export and the nuclear pore, *Trends in Cell Biology*, **23(7)**, 328-335 **(2013)**

12. De Bruijn I., Kundra R., Mastrogiacomo B., Tran T. N., Sikina L., Mazor T. and Schultz N., Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal, *Cancer Research*, **83(23)**, 3861-3867 **(2023)**

13. De Magistris P., The great escape: mRNA export through the nuclear pore complex, *International Journal of Molecular Sciences*, **22(21)**, 11767 **(2021)**

14. Epperla N., Vaughn J. L., Othus M., Hallack A. and Costa L. J., Recent survival trends in diffuse large B-cell lymphoma—Have we made any progress beyond rituximab?, *Cancer Medicine*, **9(15)**, 5519-5525 **(2020)**

15. Friedberg J. W. and Fisher R. I., Diffuse large B-cell lymphoma, *Hematology/Oncology Clinics of North America*, **22(5)**, 941-952 **(2008)**

16. Frontzek F., Staiger A. M., Wullenkord R., Grau M., Zapukhlyak M., Kurz K. S. and Lenz G., Molecular profiling of EBV associated diffuse large B-cell lymphoma, *Leukemia*, **37(3)**, 670-679 **(2023)**

17. Gao J., Aksoy B. A., Dogrusoz U., Dresdner G., Gross B., Sumer S. O. and Schultz N., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, *Science Signaling*, **6(269)**, pl1-pl1 **(2013)**

18. Guo M. and Wang S. M., Genome instability-derived genes are novel prognostic biomarkers for triple-negative breast cancer, *Frontiers in Cell and Developmental Biology*, **9**, 701073 **(2021)**

19. Heath C. G., Viphakone N. and Wilson S. A., The role of TREX in gene expression and disease, *Biochemical Journal*, **473(19)**, 2911-2935 **(2016)**

20. Huang D. W., Sherman B. T. and Lempicki R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature Protocols*, **4(1)**, 44-57 **(2009)**

21. Li J., Song M., Liu Z., Nan F., Wang B., Qian D.,and Hu M., Analysis of the mRNA export protein ZC3H11A in HCMV infection and pan-cancer, *Frontiers in Microbiology*, **14**, 1296725 **(2023)**

22. Li S., Young K. H. and Medeiros L. J., Diffuse large B-cell lymphoma, *Pathology*, **50(1)**,74-87 **(2018)**

23. Lossos I. S. and Morgensztern D., Prognostic biomarkers in diffuse large B-cell lymphoma, *Journal of Clinical Oncology*, **24(6)**, 995-1007 **(2006)**

24. Martelli M., Ferreri A. J., Agostinelli C., Di Rocco A., Pfreundschuh M. and Pileri S. A., Diffuse large B-cell lymphoma, *Critical Reviews in Oncology/Hematology*, **87(2)**, 146-171 **(2013)**

25. Marullo R., Rutherford S. C., Revuelta M. V., Zamponi N., Culjkovic-Kraljacic B., Kotlov N. and Cerchietti L., XPO1 enables adaptive regulation of mRNA export required for genotoxic stress tolerance in cancer cells, *Cancer Research*, **84(1)**, 101-117 **(2024)**

26. Nesset K. A., Perri A. M. and Mueller C. R., Frequent promoter hypermethylation and expression reduction of the glucocorticoid receptor gene in breast tumors, *Epigenetics*, **9(6)**, 851-859 **(2014)**

27. Qi P., Zhai Q. and Zhang X., RUNX1 facilitates heart failure progression through regulating TGF-β-induced cardiac remodeling, *PeerJ*, **11**, e16202 **(2023)**

28. Sarma S. and Mehta J., Spectrum of lymphomas in India, *International Journal of Molecular and Immuno Oncology*, ***9*(1)**, 16-24 **(2024)**

29. Sehn L. H. and Salles G., Diffuse large B-cell lymphoma, *New England Journal of Medicine*, **384(9)**, 842-858 **(2021)**

30. Sherman B. T., Hao M., Qiu J., Jiao X., Baseler M. W., Lane H. C. and Chang W., DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update), *Nucleic Acids Research*, **50(W1)**, W216-W221 **(2022)**

31. Siddiqui N. and Borden K. L., mRNA export and cancer, *Wiley Interdisciplinary Reviews: RNA*, **3(1)**, 13-25 **(2012)**

32. Sriramoju Shamili, Avula Srinivas, Konda Sindhura and Siddoju Kavitha, The development, preparation and characterisation of novel pyran derivatives and their biological assessment, *Res. J. Chem. Environ.*, **28(3)**, 61-69 **(2024)**

33. Stewart M., Nuclear export of mRNA, *Trends in Biochemical Sciences*, **35(11)**, 609-617 **(2010)**

34. Sukswai N., Lyapichev K., Khoury J. D. and Medeiros L. J., Diffuse large B-cell lymphoma variants: an update, *Pathology*, **52(1)**, 53-67 **(2020)**

35. Susanibar-Adaniya S. and Barta S. K., 2021 update on diffuse large B cell lymphoma: a review of current data and potential applications on risk stratification and management, *American Journal of Hematology*, **96(5)**, 617-629 **(2021)**

36. Szklarczyk D., Gable A. L., Lyon D., Junge A., Wyder S., Huerta-Cepas J. and Mering C. V., STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research*, **47(D1)**, D607-D613 **(2019)**

37. Tang Z., Kang B., Li C., Chen T. and Zhang Z., GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis, *Nucleic Acids Research*, **47(W1)**, W556-W560 **(2019)**

38. Tang Z., Li C., Kang B., Gao G., Li C. and Zhang Z., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Research*, **45(W1)**, W98-W102 **(2017)**

39. Tomczak K., Czerwińska P. and Wiznerowicz M., Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, *Contemporary Oncology/Współczesna Onkologia*, **19(1A)**, 68-77 **(2015)**

40. Wang J., Li Y., Xu B., Dong J., Zhao H., Zhao D. and Wu Y., ALYREF drives cancer cell proliferation through an ALYREF-MYC positive feedback loop in glioblastoma, *OncoTargets and Therapy*, **14**, 145-155 **(2021)**

41. Wang S. S., Epidemiology and Etiology of Diffuse Large B-Cell Lymphoma (DLBCL), *Seminars in Hematology*, **60(5)**, 255–266 **(2023)**

42. Werren E. A., LaForce G. R., Srivastava A., Perillo D. R., Li S., Johnson K. and Schaffer A. E., TREX tetramer disruption alters RNA processing necessary for corticogenesis in THOC6 Intellectual Disability Syndrome, *Nature Communications*, **15(1)**, 1640 **(2024)**

43. Wickramasinghe V. O., Andrews R., Ellis P., Langford C., Gurdon J. B., Stewart M. and Laskey R. A., Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP, *Nucleic Acids Research*, **42(8)**, 5059-5071 **(2014)**

44. Yildirim M., Kaya V., Demirpençe Ö. and Paydas S., The role of gender in patients with diffuse large B cell lymphoma treated with rituximab-containing regimens: a meta-analysis, *Archives of Medical Science*, **11(4)**, 708 **(2015)**

45. Yu T., Zhang Q., Yu S. K., Nie F. Q., Zhang M. L., Wang Q. and Lu K. H., THOC3 interacts with YBX1 to promote lung squamous cell carcinoma progression through PFKFB4 mRNA modification, *Cell Death and Disease*, **14(7)**, 475 **(2023)**

46. Yuan Y., Fan Y., Tang W., Sun H., Sun J., Su H. and Fan H., Identification of ALYREF in pan cancer as a novel cancer prognostic biomarker and potential regulatory mechanism in gastric cancer, *Scientific Reports*, **14(1)**, 6270 **(2024)**

47. Zhang J., Zhao Q., Zhao J., Cui X. and Chen X., A pan-cancer analysis of the oncogenic and immunological roles of THOC3 in human cancer **(2024)**

48. Zhao Y., Xing C. and Peng H., ALYREF (Aly/REF export factor): a potential biomarker for predicting cancer occurrence and therapeutic efficacy, *Life Sciences*, **338**, 122372 **(2024)**

(Received 21st March 2025, accepted 18th May 2025)