BMC Veterinary Research

RESEARCH

Open Access

# A preliminary study of gene expression changes in Koalas Infected with Koala Retrovirus (KoRV) and identification of potential biomarkers for KoRV pathogenesis

Lipi Akter[1,2†], Md Abul Hashem[1,2,3†], Mohammad Enamul Hoque Kayesh[1,2,4†], Md Arju Hossain[5], Fumie Maetani[6,7], Rupaly Akhter[1,2], Kazi Anowar Hossain[1,2], Md Haroon Or Rashid[1,2], Hiroko Sakurai[6], Takayuki Asai[6], M. Nazmul Hoque[8*] and Kyoko Tsukiyama-Kohara[1,2*]

## Abstract

**Background** Koala retrovirus (KoRV), a major pathogen of koalas, exists in both endogenous (KoRV-A) and exogenous forms (KoRV-A to I and K to M) and causes multiple disease phenotypes, including carcinomas and immunosuppression. However, the direct association between the different KoRV subtypes and carcinogenesis remains unknown. Differentially expressed gene (DEG) analysis of peripheral blood mononuclear cells (PBMCs) of koalas carrying both endogenous (KoRV-A) and exogenous (KoRV-A, B, and C) subtypes was performed using a high-throughput RNA-seq approach. PBMCs were obtained from three healthy koalas: one infected with endogenous (KoRV-A; Group I) and two infected with exogenous (KoRV-B and/or KoRV-C; Group II) subtypes. Additionally, spleen samples ($n=6$) from six KoRV-infected deceased koalas (K1- K6) and blood samples ($n=1$) from a live koala (K7) were collected and examined to validate the findings.

**Results** All koalas were positive for the endogenous KoRV-A subtype, and eight koalas were positive for KoRV-B and/or KoRV-C. Transcription of KoRV *gag*, *pol*, and *env* genes was detected in all koalas. Upregulation of cytokine and immunosuppressive genes was observed in koalas infected with KoRV-B or KoRV-B and -C subtypes, compared to koalas infected with only KoRV-A. We found 550 DEG signatures with significant (absolute $p<0.05$, and absolute $\log_2$ Fold Change (FC) > 1.5) dysregulation, out of which 77.6% and 22.4% DEGs were upregulated ($\log_2$FC > 1.5) and downregulated ($\log_2$FC < −1.5), and downregulated ($\log_2$ FC < −1), respectively. We identified 17 unique hub genes (82.3% upregulated and 17.7% down-regulated), with *KIF23*, *CCNB2*, *POLR3F*, and *RSL24D1* detected as the potential hub genes modified with KoRV infection. Real-time RT-qPCR was performed on seven koalas to ascertain the expression levels of four potential hub genes, which were subsequently normalized to actin copies. Notably,

†Lipi Akter, Md Abul Hashem and Mohammad Enamul Hoque Kayesh these authors contributed equally to this work.

*Correspondence:
M. Nazmul Hoque
nazmul90@bsmrau.edu.bd
Kyoko Tsukiyama-Kohara
kkohara@vet.kagoshima-u.ac.jp
Full list of author information is available at the end of the article

all seven koalas exhibited distinct expression signatures for the hub genes, especially, *KIF23* and *CCNB2* show the highest expression in healthy koala PBMC, and *POLR3F* shows the highest expression in koala with lymphoma (K1).

**Conclusion** Thus, it can be concluded that multiple KoRV subtypes affect disease progression in koalas and that the predicted hub genes could be promising prognostic biomarkers for pathogenesis.

**Keywords** Koala retrovirus, RNA-seq, PBMCs, Differentially expressed gene, Biomarkers

## Background

The population of koalas (*Phascolarctos cinereus*), an iconic Australian marsupial species, is decreasing owing to habitat loss, anthropogenic challenges, and microbial infections. The Australian government declared the species endangered by extinction under the Environment Protection and Biodiversity Conservation (EPBC) Act on February 12, 2022 (https://rb.gy/4zygx). Neoplastic diseases in koalas, driven by the koala retrovirus (KoRV), pose a substantial threat to both koala health and conservation efforts [1–3]. Unlike most retroviruses, KoRV exists as both an endogenous and exogenous virus [4]. In 2000, researchers successfully sequenced the complete genome of KoRV, revealing the presence of essential genes (*gag*, *pol*, and *env*) flanked by long terminal repeats (LTRs) at both the 5′ and 3′ ends [5, 6]. Based on phylogenetic analysis and receptor-binding differences in *env*, KoRVs are categorized into 12 subtypes: KoRV-A [5], KoRV-B [1], KoRV-C [7, 8], KoRV-D [8], KoRV-E, KoRV-F [9], KoRV-G, KoRV-H, KoRV-I [10], KoRV-K, KoRV-L, and KoRV-M [4]. Recently, defective KoRV subtypes of KoRV-D and -E with significant deletions in the *gag* and *pol* genes have been reported [11, 12]. KoRV-A, which exists in both endogenous and exogenous forms, is the most extensively characterized subtype, whereas pathogenic KoRV-B is the most widely documented exogenous subtype [13]. KoRV-B stands out as the most pathogenic subtype that likely plays a significant role in the occurrence of neoplasia/lymphoma and exacerbation of chlamydial disease in koala populations [1, 14, 15]. Furthermore, the roles of different KoRV subtypes and their co-infection in disease occurrence have not yet been confirmed.

Virus-borne neoplastic diseases pose significant health threats to various species, including koalas. To unravel the pivotal genetic elements in the progression of the host gene transformation by oncogenic viruses, next-generation sequencing (NGS)-based RNA-seq has been used extensively in life science research for comprehensive transcriptome analysis and to provide valuable genetic information [16–21]. Furthermore, RNA-seq facilitates the identification of unknown mRNA transcripts and spliced isoforms by enabling quantitative determination of expression levels in whole transcriptomes [22, 23]. Transcriptomic analyses of virus-infected cells are beneficial for identifying the host immune response dynamics and gene regulatory networks (GRNs). GRNs can provide valuable insights into significant gene ontology modules and can pinpoint key genes, often referred to as "hub genes," which play a crucial role in the development of a specific tumor [24–26]. Moreover, comparative transcriptome analysis can effectively compare gene expression patterns among different groups and provide insights into biological processes [10, 27].

In addition to habitat loss and vehicular injuries, KoRV is considered a major threat to koala health and conservation [13, 28]. Moreover, the cause of disease in KoRV subtype infections is yet to be established [29]. Therefore, extensive investigations are necessary to enhance our understanding of the mechanisms driving KoRV pathogenesis, particularly in koalas infected with multiple KoRV subtypes concurrently. In the study, RNA-seq was used to analyze the gene expression profiles of peripheral blood mononuclear cells (PBMCs) from healthy koalas. Three specimens in total were examined, including one from a koala infected with the endogenous KoRV-A subtype (Group I) and two from koalas infected with exogenous subtypes (KoRV-A, -B, and -C) (Group II). The objective of the present study was to obtain a comprehensive perspective on KoRV and its subtypes within the host, which could advance our understanding of host-virus interactions and KoRV pathogenesis using various bioinformatic tools.

## Results

### KoRV subtypes and health status of koalas

Specimens from ten koalas, including three healthy adults (RNA-seq study), six dead koalas, and one healthy adult (K7) were used. All ten koalas tested positive for endogenous KoRV-A subtype. Among these, eight koalas were positive for KoRV-B and/or KoRV-C, along with KoRV-A. Additionally, koalas KHB-02 and KHL-03 tested positive for KoRV-B and -C, respectively, presented as Group II (Table S1), and compared with KHY-04 (only KoRV-A positive, presented as Group I). The hematological analysis results of the koalas used in the present study are listed in Table S2.

**Table 1** Mapping of the transcriptome data against the koala retrovirus (KoRV) reference genome

| Sample ID | Total reads | Total mapped | Multiple mapped | Uniquely mapped | Read1 | Read2 | Reads map to '+' | Reads map to '-' | Non_splice reads | Splice reads | Reads mapped in proper pairs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| KHY-04 | 52,977,420 | 56,351 (0.106368%) | 6044 (0.0114086%) | 50,307 (0.0949593%) | 24,955 | 25,352 | 25,116 | 25,191 | 49,496 | 811 | 45,414 |
| KHB-02 | 54,709,258 | 48,503 (0.0886559%) | 4663 (0.00852324%) | 43,840 (0.0801327%) | 21,786 | 22,054 | 22,141 | 21,699 | 43,646 | 194 | 36,654 |
| KHL-03 | 60,103,334 | 52,183 (0.0868221%) | 7592 (0.0126316%) | 44,591 (0.0741906%) | 22,987 | 21,604 | 21,660 | 22,931 | 44,319 | 272 | 39,942 |

None of the three healthy koalas exhibited any visible abnormalities (Table S1).

#### Transcriptome sequencing, assembly, gene annotation and expression

In total, 242,542,642 raw reads were generated from the three koala PBMCs using RNA-Seq. A total of 227,382,728 clean reads (i.e., 6.25% filtered out) with a total data of 31,497,518,861 bp (approximately 31.5 GB) were retained after removing adaptors, cleaning up contaminants, and filtering out low-quality reads. The average Q30 and GC percentages were 93.96% and 60.34%, respectively (Table S3). Based on the high-quality data, 14,425,510 contigs with 708,983,496 bp of data were assembled. Finally, 417,284 unigenes were identified in the contigs using paired-end relationships. A summary of the unigene numbers and length distributions is provided in Table S4. The expected numbers of fragments per kilobase of transcript per million mapped (FPKM) reads were determined using a density saturation curve (Fig. S1) and sequence homogeneity (Fig. S2) at the transcriptome level. Subsequently, high-quality reads were aligned against the reference KoRV genome (NC_039228.1), which showed compatibility between the reference KoRV genome and transcriptome data obtained in the present study (Table 1). The distribution of reads in the reference genome was determined based on genomic features, including exons, introns, and intergenic regions (Fig. S3). Most of the sequences were in the exon regions, indicating that mature RNA was produced by each koala. We also calculated the read density on each chromosome as the $\log_2$ value of the read count for 1 kb (Fig. S4).

To annotate the assembled unigenes, their sequences were aligned against publicly available nucleotide/protein databases. Out of the 76,549 annotated unigenes, 74,413, 58,662, and 37,303 were annotated in the (NR, Swiss-Prot, and COG databases, respectively (Fig. 1A). After eliminating redundancies from the different databases, 17,029 unigenes, including 16,109 in NR, 570 in Swiss-Prot, and 350 in COG, were annotated at least once (Fig. 1A). After functional annotation, the numbers of sequences from different species that matched koala unigenes were calculated from the annotation results. The top 10 enriched species based on NR functional annotation indicated high unigene variation between koalas of this study and reference genomes (Fig. 1B). Out of the 37,303 unigenes mapped to 26 COG categories, clusters were frequently mapped to signal transduction mechanisms, general function prediction, transcription, post-translational modifications, protein turnover, and chaperones. Using RSEM software, the read counts for each gene were obtained and FPKM analysis was conducted accordingly. The number of genes with different expression levels is summarized in Table S5. The general distribution and dispersal of the expression of the genes are indicated in Fig. 2. The general distribution and FPKM density distribution of gene expression in different samples were similar for all koalas (Fig. S5). Gene diversity analysis was performed using HT-seq (version 0.6.1) Expression of the *gag*, *pol*, and *env* sequences of KoRV was detected in each koala PBMC sample (Table 2).

#### Analysis of differentially expressed genes

To elucidate whether DEGs contribute to KoRV infections in koalas, three RNA-seq datasets [including one dataset ($n=1$; KHY-04) from endogenous (KoRV-A) subtype (Group I) and two datasets ($n=2$; KHB-02, and KHL-03) from exogenous (KoRV-B and -C) along with koRV-A subtypes (Group II)] of PBMCs from the three healthy koalas were analyzed. To perform RNA-seq analysis, reference genome data were retrieved from NCBI belonging to a previously published BioProject under accession number PRJNA359763 (https://www.ncbi.nlm.nih.gov/bioproject/384067). We particularly focused on the dysregulation (up- or down-regulation) of the identified DEGs during koala retroviral pathophysiology and their overlap with the healthy state of koalas. After excluding duplicate genes, 550 DEGs were identified (absolute $p<0.05$, and absolute $\log_2$ FC > 1.5) (Fig. 2). Among the detected DEGs, 77.6% (427/550) DEGs were found to be upregulated ($\log_2$ FC > 1.5), whereas 22.4% (123/550) DEGs were downregulated ($\log_2$FC < −1.5). The upregulated and downregulated DEGs in Group I and Group II koalas are illustrated using volcano plots (Fig. 2).

#### Cluster analysis and selection of significant hub proteins from protein–protein interactions

Identified DEGs ($n=550$) were initially submitted to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database to create a protein-protein interaction (PPI) network. In the PPI network (Fig. 3), proteins or genes were designated as nodes, and the linkages among the nodes were designated as edges. We identified 96 nodes and 158 edges in the PPI network. The density and heterogeneity of the networks were 0.071 and 0.524, respectively (Fig. 3). PPI analysis further showed that the average number of neighbors, network diameter, network radius, characteristic path length, clustering coefficient, and network centralization were 3.045, 11.0, 6.0, 5.063, 0.395, and 0.072, respectively. To elucidate the functional, active, and protein-related functions of the detected DEGs, cluster analysis was performed. Using the MCODE technique and the ClusterViz plugin in Cytoscape v3.10.1, four major clusters were identified within the PPI network. Cluster-1 contained

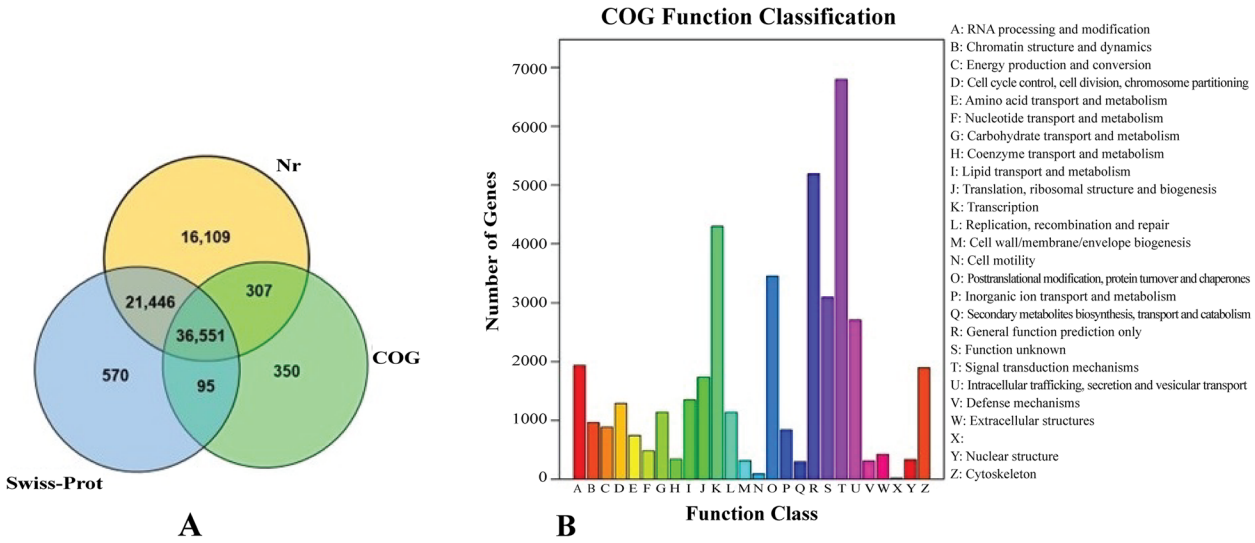


**Fig. 1** Unigene annotations. **A** Venn diagram of various unigene annotation results based on different databases (Nr [Non-redundant protein], Swiss-Prot, and COG [Clusters of Orthologous Groups of proteins]). **B** Gene Ontology (GO) functional annotation. The x-axis indicates the number of genes, and the y-axis indicates the GO classification; color code has been used to distinguish the first-level GO classification. X; unidentified

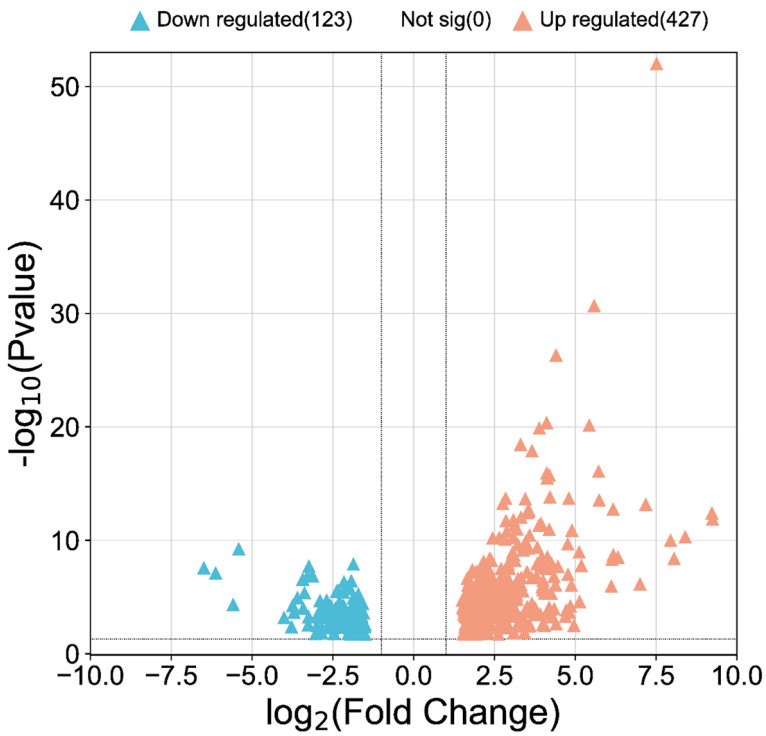


**Fig. 2** Volcano plots showing the differentially expressed genes (DEGs) associated with KoRV subtypes in three koala samples (Group I represents one sample KoRV-A as control and Group II represents two samples KoRV-B and C as cases). We identified 550 DEGs comprising 427 upregulated and 123 down-regulated DEGs based on p-value and log2 FC value. Peach and cyan color dots indicate the upregulated and down-regulated genes, respectively. The x-axis represents the $\log_2$ fold change (logFC), and the y-axis represents $-\log_{10}$ (*p*-value) of each DEG

**Table 2** Gene diversity analysis of koala retrovirus (KoRV) in study koalas

| Gene_ID | Exonic gene size | KHY-04 | KHY-04_FPKM | KHB-02 | KHB-02_FPKM | KHL-03 | KHL-03_FPKM | Gene name |
|---|---|---|---|---|---|---|---|---|
| Gene-D1Y44_gp1 | 1566 | 3984 | 142,540.41 | 5386 | 218,662.08 | 4761 | 203,115.3 | *gag* |
| Gene-D1Y44_gp2 | 3384 | 4978 | 82,420.45 | 6888 | 129,408.16 | 5057 | 99,838.68 | *pol* |
| Gene-D1Y44_gp3 | 1980 | 8886 | 251,449.95 | 3455 | 110,938.36 | 5150 | 173,771.39 | *env* |

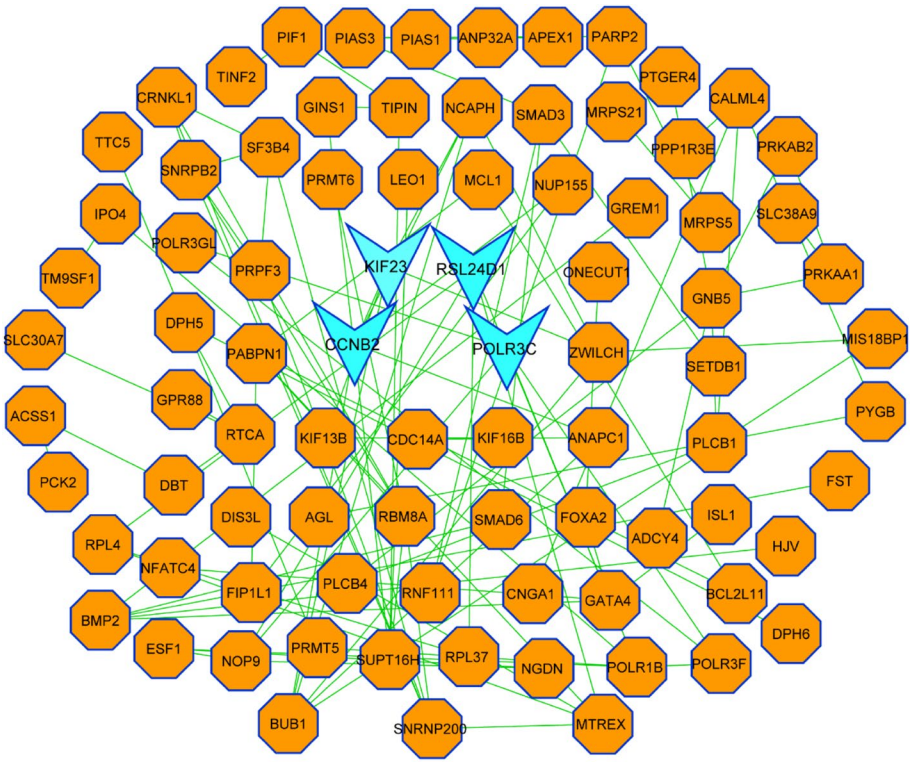


**Fig. 3** Protein–protein interaction (PPI) network of common DEGs in koalas infected with both endogenous and exogenous subtypes. The nodes represent the proteins, and the edges represent the interactions across the proteins. V-shaped and cyan color nodes indicate the most connected significant nodes in the PPI network

14 nodes with 56 edges, whereas Cluster-2 contained six nodes with 25 edges (Fig. 4). Proteins with several connecting edges were identified as hubs. Five different algorithms (MCC: maximal clique centrality, MNC: maximum neighborhood component, EPC: edge-percolated component, Closeness, and Degree) were applied (Fig. 5). As anticipated by the five methods, 17 unique hub nodes were recognized as potential hub proteins (KIF23, BUB1, CCNB2, POL3RF, NGDN, POLR3C, IPO4, ESF1, POLR1B, RSL24D1, MTREX, SUPT16H, BMP2, RBM8A, GNB5, TRMT6, and NUP155) (Figs. 5A–E). Only four hub proteins (KIF23, CCNB2, POLR3F, and RSL24D1) were shared among the five topological matrices (Fig. 5F). Among the shared hub proteins, KIF23 showed the highest score in the MNC, Degree, Closeness, and EPC methods, indicating the highest protein connectivity between PPI networks. However, RSL24D1 showed the highest scores in terms of EPC and Degree algorithms (Figs. 5A-E). Among the detected hub genes ($n=17$), 14 were upregulated, and three (BUB1, POLR3C, and IPO4) were downregulated (Fig. 6).

### Functional and pathway enrichment analysis of unique hub genes

To gain a broader perspective on the DEGs, functional enrichment analysis of different gene ontology (GO) terms, such as biological process (BP), cellular component (CC), and molecular function (MF), was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) online server (https://

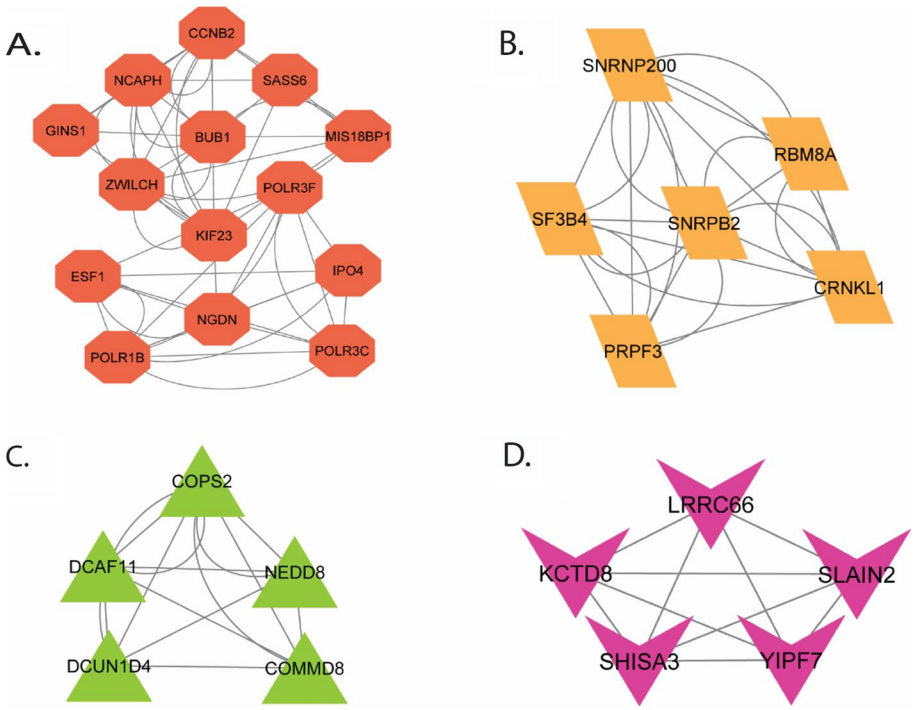


**Fig. 4** Clustering analysis from protein–protein interaction (PPI) network using the ClusterViz plugin of Cytoscape. **A** Cluster 1 with 14 nodes and 56 edges, **B** Cluster 2 with 6 nodes and 25 edges, **C** Cluster 3 with 5 nodes and 14 edges, and **D** Cluster 4 with 5 nodes and 10 edges

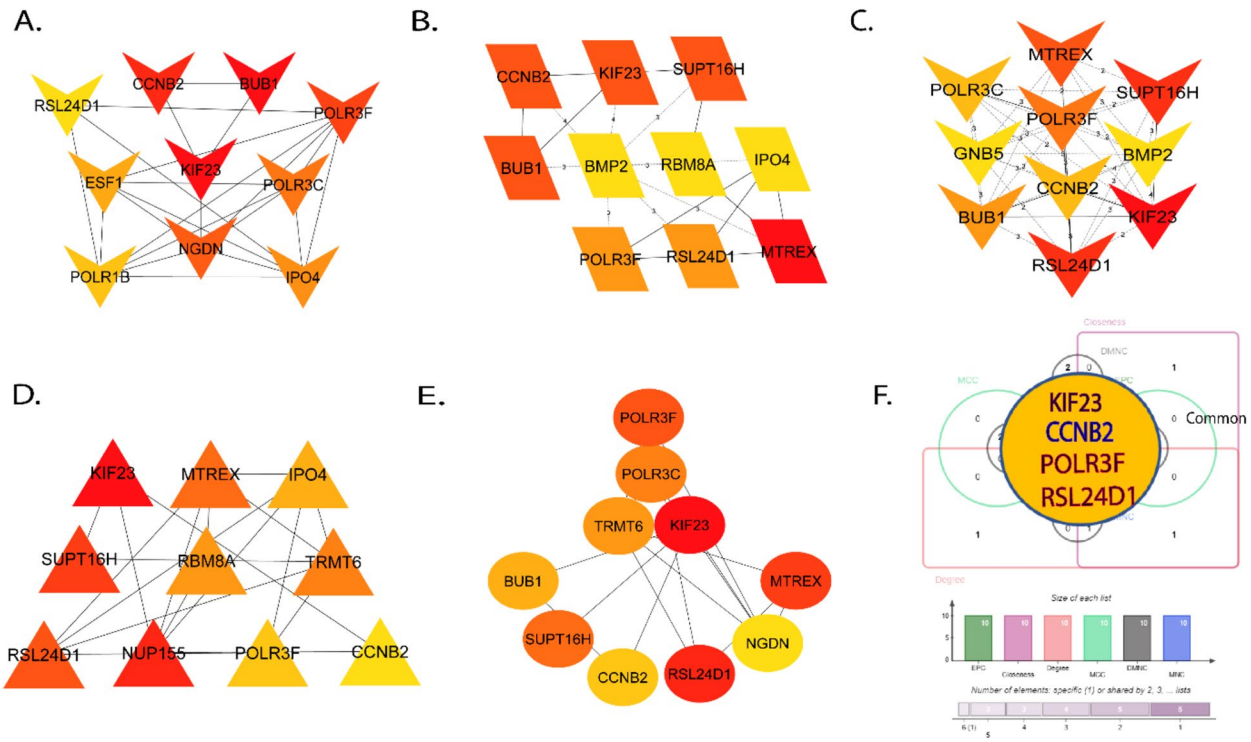


**Fig. 5** Determination of hub genes from the protein–protein interaction (PPI) networks using the CytoHubba plugin in Cytoscape. Five algorithms of the Cytohubba plugin were applied to obtain the hub genes. Here (**A**) MCC (Maximal Clique Centrality), (**B**) MNC (Maximum Neighborhood Component), (**C**) Degree, (**D**) Closeness, and (**E**) EPC (Edge Percolated Component) algorithms. Red to yellow color gradients indicate the higher ranking of hub genes. **F** Venn diagram showing the common four hub proteins based on five metrics

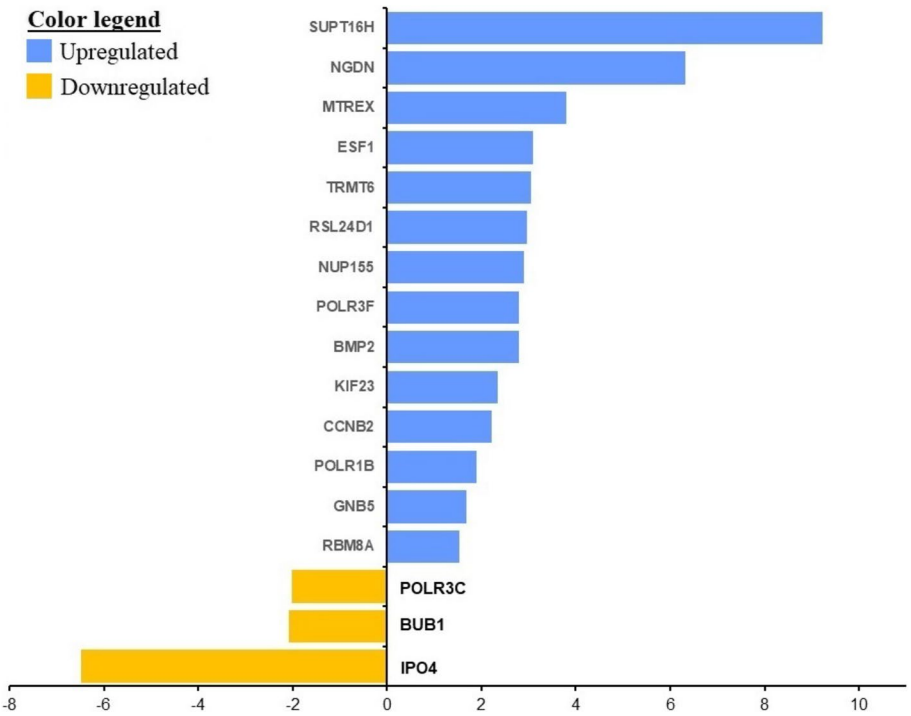


**Fig. 6** Bar plot depicting the up- and down-regulated hub genes in koalas infected with both endogenous and exogenous KoRV subtypes. Bar plot representing 17 unique hub genes that were significantly ($p < 0.05$) dysregulated (including 14 upregulated; logFC > 1.5 and 3 downregulated; logFC <—1.5) based on absolute logFC values

david.ncifcrf.gov/). Significantly enriched GO terms were identified if the enrichment yielded a high logarithmically adjusted *p*-value. The top enriched GOs terms for BP ($n=9$), CC ($n=7$), and MF ($n=11$) are illustrated in Fig. 7. We further performed enrichment analysis of the different molecular pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database incorporated within the DAVID database. The significance of the KEGG pathways was determined using p-values, and the highest count of DEGs was plotted against 30 pathways (Fig. 8).

### Expression patterns of koala hub mRNAs in tissues

To investigate the expression patterns of the hub genes, the mRNA expression of *CCNB2*, *POLR3F*, *KIF23*, and *RSL24D1* was assessed in both spleen tissues (K1-K6) and blood (K7) of koalas (Tables S2 and S6). These tissues tested positive for KoRV-A and/or KoRV-A, -B, and KoRV-C, as shown in Table S6. We observed diverse hub gene mRNA expression patterns across all koalas (Fig. 9). Specifically, for *CCNB2*, the highest expression was detected in koala K7, whereas the lowest expression was observed in koala K4. *POLR3F* mRNA exhibited the highest expression in koala K1 followed by K2, K5, and K6. It had the lowest expression in K3 followed by K4 and K7. In the case of *KIF23* mRNA, koala K7 displayed the highest expression levels, while all the other deceased koalas showed low expression levels. When examining *RSL24D1* mRNA expression, koala K5 exhibited the highest levels, followed by koala K4 and K2, whereas K3 exhibited the lowest levels, followed by K1, K6, and K7. Thus, *KIF23* and *CCNB2* showed the highest expression in healthy koala PBMC, while *POLR3F* showed the highest expression in koala with lymphoma (koala K1).

## Discussion

Rapid advancements in bioinformatics and the emergence of high-throughput techniques such as microarray and NGS have enabled a comprehensive understanding of the processes underlying carcinogenesis and the progression of various cancer types. High-throughput platforms have found widespread applications in early diagnosis, histological characterization, molecular classification, prognosis prediction, and analysis of drug resistance in cancer [19, 30]. DEGs, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), and differentially methylated CpG sites have the potential to offer valuable insights into biomarkers associated with retroviral cancer [30, 31].

Here, we provide insights into KoRV subtype-mediated dysregulation of several DEGs in the PBMCs of three healthy koalas in Japan. Among the 12 KoRV subtypes

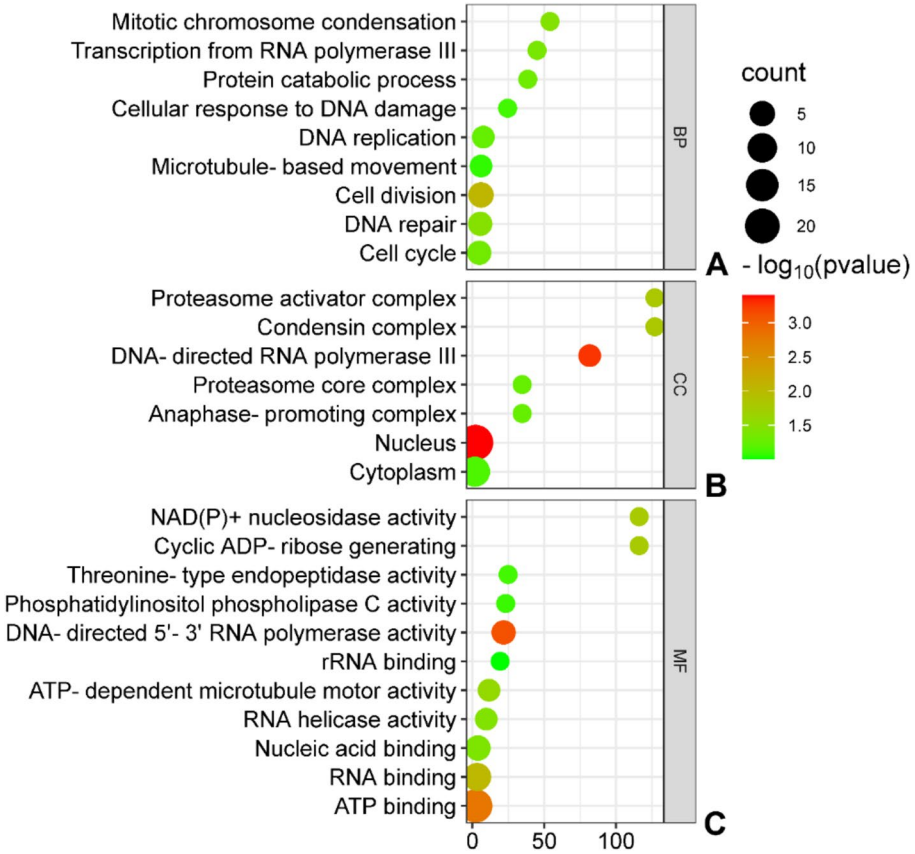


**Fig. 7** Gene Ontology (GO) pathway enrichment analysis based on the $-\log_{10}$(*p*-value) values. The pathways have been formed by combining the DEGs that are common in the (**A**) gene ontology (GO) biological process (BP), (**B**) GO cellular component (CC), and (**C**) molecular function (MF). The circular shapes indicate the top 27 functional pathway terms found in the peripheral blood of koalas with KoRV exogenous and endogenous subtypes, while red to green color gradients indicate significantly higher to lower pathway scoring. The size of the circle demonstrates the count of hub genes. The most important GO pathways included (**A**) cell division, DNA repair, mitotic chromosome condensation, transcription using RNA polymerase III, cell cycle, protein catabolic process, DNA replication, cellular response to DNA damage, and microtubule-based movement for the BP, (**B**) nucleus, DNA-directed RNA polymerase III, proteasome activator complex, condensin complex, proteasome core complex, anaphase-promoting complex and cytoplasm for CC, and (**C**) DNA-directed 5´-3´ RNA polymerase activity, ATP binding, RNA/rRNA binding, NAD(P) + nucleosidase activity, cyclic ADP-ribose generating, ATP-dependent microtubule motor activity, RNA helicase activity, nucleic acid binding, threonine-type endopeptidase activity, phosphatidylinositol phospholipase C activity for MF

identified to date [12], except for KoRV-A and KoRV-B, none have been studied comprehensively. KoRV-B has been suggested to be the most pathogenic based on increased infectivity measured *in-vitro* and evidence supporting its association with disease [1, 3, 15]. Therefore, to gain more insight into host responses based on KoRV subtype differences, transcriptome analysis was performed using RNA-seq of koala PBMCs obtained from three healthy koalas originally from Queensland, Australia, at a Japanese zoo.

The three koalas reared in a Japanese zoo had KoRV belonging to both endogenous (KoRV-A; Group I) and exogenous (KoRV-A, B, and C; Group II) subtypes. To the best of our knowledge, this is the first transcriptome analysis of koala PBMCs obtained from individuals with both endogenous and exogenous KoRV subtypes to investigate gene expression profiles and analyze their interactions in the host based on subtype differences. KoRV is a retrovirus that integrates its genetic material into the DNA of infected cells, leading to long-term infection [14, 32]. Similar to other retroviral counterparts, KoRV is involved in the transformation of regulatory genetic elements of cells and can weaken the immune system of koalas, making them more susceptible to cancer and infectious diseases. Koalas infected with KoRV may experience symptoms such as increased susceptibility to infections, reproductive abnormalities, and lymphoma (a type of cancer) [14]. However, not all koalas infected with KoRV develop symptoms, and the severity of the disease can vary among individuals. We observed

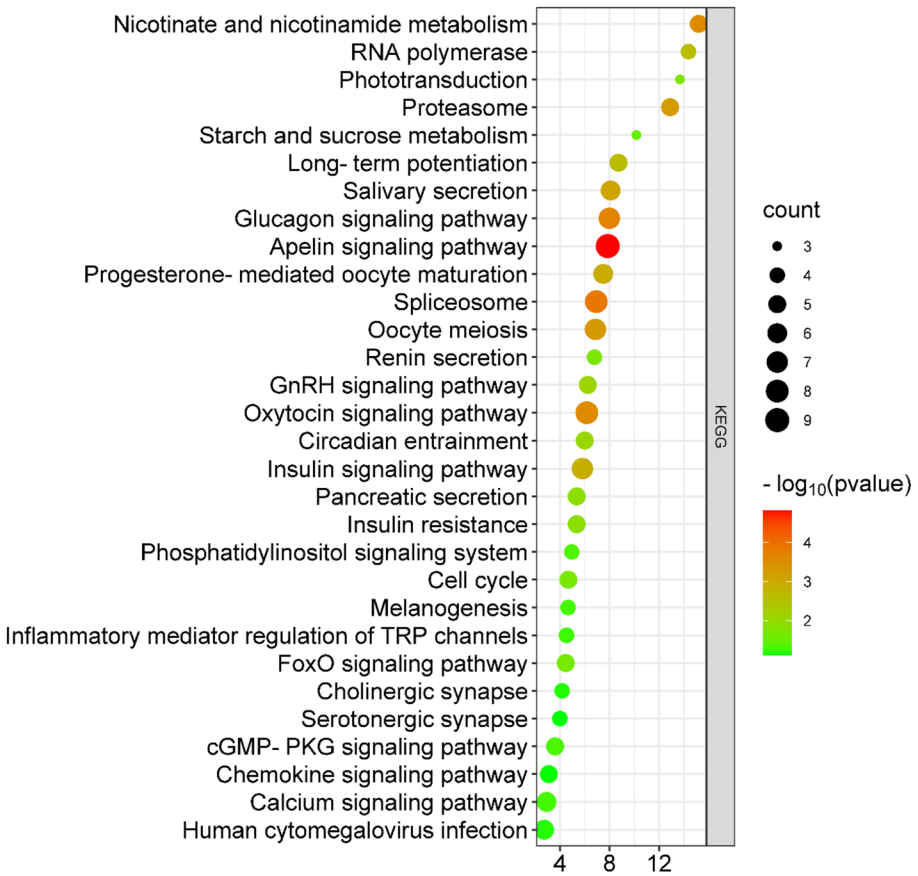


**Fig. 8** Molecular pathway enrichment analysis of top 30 terms based on the $-\log_{10}$(p-value) values. The circular shapes indicate the pathway terms, while red to green color gradients indicate significantly higher to lower pathway scoring. The size of the circle demonstrates the count of hub genes. The most significant KEGG pathways were the apelin signaling pathway; glucagon and calcium signaling pathway; starch, sucrose, nicotinate, and nicotinamide metabolism; salivary and pancreatic secretion; insulin signaling pathway; GnRH and Oxytocin signaling pathways; long-term potentiation; phototransduction; FoxO and cGMP-PKG signaling pathways; insulin resistance; renin secretion; inflammatory mediator regulation of TRP channels; and chemokine signaling pathway

that genes were differentially expressed in both groups (Group I vs. Group II), suggesting an association of KoRV subtype differences with DEG expression profiles and subsequent disease progression. Therefore, more genes were upregulated in koalas with exogenous KoRV subtypes (Group II) than in koalas with endogenous KoRV subtypes (Group I). A similar pattern of gene expression dysregulation has been reported in human hepatocellular carcinoma (HCC) caused by an oncogenic retrovirus, the hepatitis B virus [19, 20, 30, 31].

Using these dysregulated koala genes, we performed a PPI network analysis. PPI network analysis is the most prominent part of the study, as hub gene detection, module analysis, and drug identification depend on the PPI network [21, 24]. According to the PPI network analysis, four major clusters and 17 unique hub genes (of them, 82.3% upregulated and 17.7% downregulated) were identified. Among the hub genes, only four hub nodes (CCNB2, POLR3F, KIF23, and RSL24D1) were predicted to be potential hub proteins using five distinct algorithmic methods (MCC, MNC, Closeness, Degree, and EPC). KIF23 contains genetic instructions for producing mitotic kinesin-like protein-1 (MKLP1), an extremely conserved protein essential for the assembly of the central spindle and midbody during cytokinesis [20, 33]. CCNB2 plays a crucial role in regulating the cell cycle, particularly during the transition from the G2 phase to the M phase (mitosis). Its function involves the activation of cyclin-dependent kinase 1 (CDK1), which forms a cyclin B2/CDK1 complex. High CCNB2 expression is associated with poor prognosis, indicating its potential role as a prognostic marker [34]. *KIF23* and *CCNB2* showed the highest expression in normal koala PBMC in this study, therefore, they might be linked with anti-oncogenic activity.

Changes in POLR3F levels have been observed in various types of cancers. Anomalies in polymerase III activity can lead to the aberrant expression of non-coding RNAs,

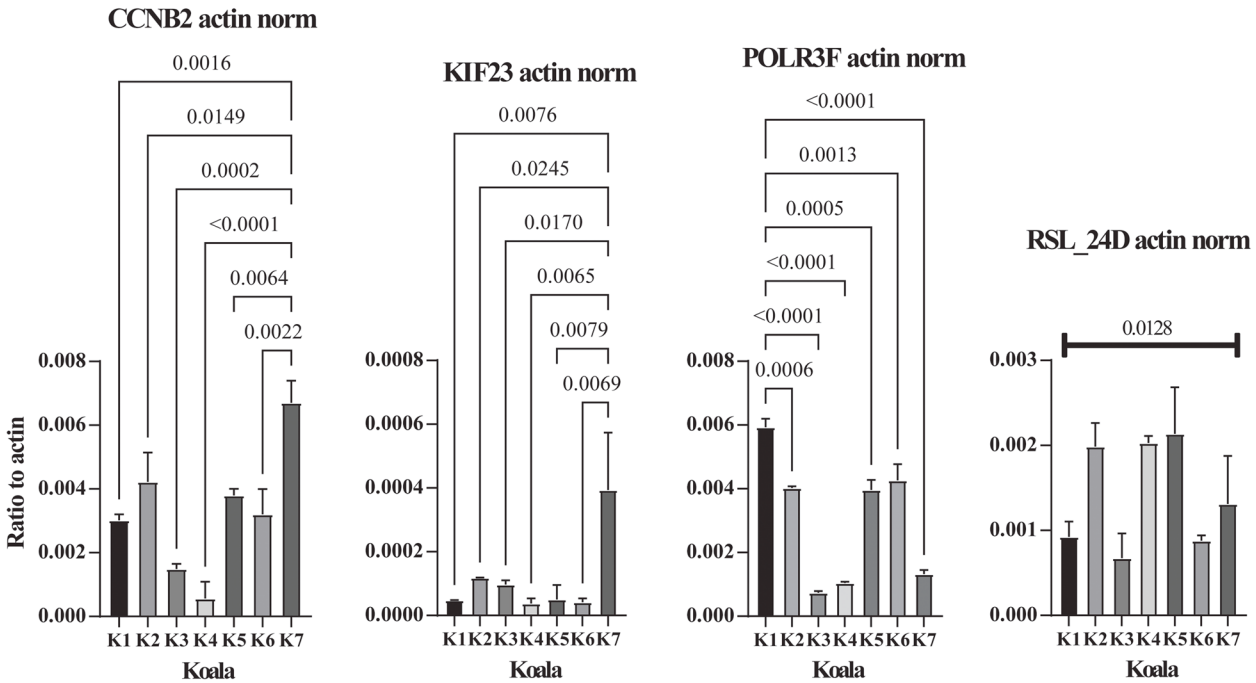


**Fig. 9** Expression profile of hub gene mRNAs in koala tissues. Expression patterns of CCNB2, POLR3F, KIF23, and RSL24D1 mRNAs in the spleen (e.g., K1-K6) and blood (K7) are shown. The transcript levels were normalized against koala actin mRNA. Data are presented as mean ± SD. All samples are statistically examined using One-way ANOVA, and Dunnets analysis (Graph Pad Prism ver. 9.5.1, software). *P* values less than 0.05 are significant and indicated when comparing with healthy control (K7). RSL2D only showed significant value after One-way ANOVA analysis ($p = 0.0128$)

affecting cellular processes such as protein synthesis and cell growth. Furthermore, mutations or dysregulation of Pol III components, such as POLR3F, can contribute to genomic instability, a hallmark of cancer [35]. Consistent with these observations, upregulation in koala lymphoma (K1) of *POLR3F* was observed. The specific function of RSL24D1 likely involves its role in ribosomal assembly. Ribosomal proteins are also involved in various cellular processes beyond protein synthesis, including cell cycle regulation, apoptosis, and cellular stress response [36]. Concerning *RSL24D1*, we have not found a significant link with the pathogenesis of 7 koalas in this study, so far. Hub proteins were identified as potential hub genes owing to their high interaction rates or degree values. Each of the detected hub proteins acts as a prognostic biomarker for various types of cancer, including breast cancer (e.g., KIF23 and BUB1) [37, 38] and esophageal, liver, breast, prostate, lung, ovarian, gastric, renal, and colorectal cancers (e.g., BUB1, SNRPB2, and KIF23) [21, 38, 39]. Several studies have reported that KIF23 is significantly overexpressed in ovarian, stomach, hepatic, and lung cancer cells [20, 40–42]. A recent study on human hepatitis B-infected HCC revealed the overexpression of *KIF14* and *KIF23* mRNAs in the liver [20]. These potential hub proteins PPI network and gene enrichment analyses were related to distinct biological processes, cellular components, and molecular functions.

Oncogenic retroviruses cause cancer in their hosts by transforming the genes involved in the regulation of cell division. The GO analysis revealed that most proteins encoded by the DEGs were involved in cell division, cellular response to DNA damage, DNA replication and repair, mitotic chromosome condensation, transcription using RNA polymerase III, protein catabolic processes, proteasome activator complexes, ATP binding and ATP-dependent microtubule motor activity, RNA/rRNA binding, RNA helicase activity, nucleic acid binding, NAD(P) + nucleosidase activity, threonine-type endopeptidase activity, and phosphatidylinositol phospholipase C activity. We hypothesized that the proteins drive undergo several physiological processes in cells to establish an oncogenic microenvironment. Some metabolites may provide signals to the nuclear pathway by controlling chromatin changes or transcription factor activity. Such signals are subsequently altered by metabolic disturbance in cancer to facilitate carcinogenesis and cancer progression [43]. Further enrichment analysis of different molecular pathways showed 30 KEGG pathways, of which the apelin signaling pathway; glucagon and calcium signaling pathway; starch, sucrose, nicotinate, and nicotinamide

metabolism; salivary and pancreatic secretion; insulin signaling pathway; GnRH and Oxytocin signaling pathways; FoxO and cGMP-PKG signaling pathway; insulin resistance; renin secretion; inflammatory mediator regulation of TRP channels; and chemokine signaling pathway were the most significant pathways.

Gene Ontology and KEGG pathway analyses have been used as tools for identifying prognostic markers of carcinogenesis [20, 31]. Gene expression levels vary according to factors such as tissue type, age, and cancer type. To establish normal reference ranges and identify potential diagnostic and prognostic markers, it is prudent to conduct quantitative surveys on both affected and healthy individuals or animals. Such an approach facilitates a comprehensive assessment of gene expression patterns, aiding the delineation of normative parameters and identification of gene expression alterations associated with disease states [44–46]. Including a larger sample size in our study may help mitigate the observed expression heterogeneity, providing a more accurate representation of prognostic determinants and yielding a cancer gene expression pattern that better reflects the reality of a small captive koala population in Japan. Regarding mapping, while HISAT2 was chosen for its efficient memory usage and compatibility with our pipeline, we recognize that STAR is widely used for its speed and accuracy, especially with complex genomes. We will consider using STAR in future analyses to benefit from its established advantages. To the best of our knowledge, the present study represents the first attempt to investigate DEGs in KoRV-infected koalas, with the aim of identifying potential hub gene markers for carcinogenesis and validating mRNA expression patterns in other multitype KoRV-infected koalas. Our findings suggest that the identified hub genes, such as KIF23, CCNB2, and POLR3F, could serve as potential biomarkers for monitoring KoRV infection and disease progression in koalas, facilitating earlier diagnosis and intervention strategies, for example, using in situ hybridization using spleen tissue. By identifying individual koalas at higher risk for developing severe disease, these biomarkers could be integrated into targeted health management practices, such as selective breeding programs aimed at reducing the prevalence of pathogenic KoRV subtypes in captive populations. Furthermore, understanding the molecular mechanisms underlying KoRV pathogenesis enhances conservation efforts by providing new tools for assessing and mitigating the impact of this virus on both captive and wild koala populations.

## Methods

### Study animals and sampling

For transcriptome analysis of koala PBMCs, three adult koalas (viz., KHB-02, KHL-03, and KHY-04) were selected from Hirakawa Zoological Park, Kagoshima, Japan. They were infected with either the endogenous (KoRV-A) or the exogenous (KoRV-B and/or -C) subtype. Additionally, spleen samples ($n=6$) from six dead koalas (viz., K1–K6), and a blood sample ($n=1$) from an alive koala (K7) was collected from the zoo. The demographic data for each koala and their infection status (infected with KoRV-A, -B, or -C subtypes) are presented in Tables S1 and S2. Whole-blood samples treated with heparin or ethylenediaminetetraacetic acid (EDTA) were collected via venipuncture in November 2020, and PBMCs were isolated as described previously [29]. Genomic DNA (gDNA) was extracted from EDTA-treated whole-blood samples using a Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions, and the extracted gDNA was used as a template for PCR subtyping. KoRV-A, -B, and -C subtypes were determined by subtype-specific PCR targeting *env* (Table S7) and further confirmed by sequencing as described previously [7, 23, 32]. The health status of each koala used in the RNA-seq was determined via routine hematological analyses.

### RNA extraction and high-throughput sequencing

Total RNA was extracted from PBMCs and spleen tissue using the RNeasy Plus Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. The concentration and purity of the extracted RNA were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA quality and integrity were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). In the present study, next-generation sequencing/single-cell RNA-seq was used. RNA samples were converted into cDNA libraries using a TruSeq Standard Total RNA Prep kit (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. Paired-end ($2\times150$ bp) sequencing of the prepared library pool of samples was performed using a HiSeq2000 sequencer (Illumina) at GENEWIZ (South Plainfield, NJ, USA).

### Transcriptome alignment, assembly, and gene annotation

After removing the 5′ adaptor, trimming the 3′ acceptor, filtering low-quality reads with a quality value (Q) < 10 (the quality value was calculated using the following

equation: Q = ASCII character code – 64), and cleaning up contaminated reads [47]. Consequently, the clean reads were obtained. To obtain high-quality reads, raw RNA-seq data were filtered and trimmed using Trimmomatic [48]. High-quality clean reads were aligned against the reference KoRV genome (NC_039228.1) using HISAT2 (hierarchical indexing for the spliced alignment of transcripts) (v2.0.1) [49]. In addition, the read distribution of clean data was obtained using the reference genome, and the read density distribution on each chromosome was calculated as $\log_2$ values of read counts in 1 kb.

To perform de novo assembly, distinct contigs were assembled with Illumina short reads using Trinity v2.2.0 [50]. The reads were aligned to the assembled contigs to detect contigs from the same transcript. Subsequently, scaffolds between the closed contigs were constructed using paired-end mapping. Finally, paired-end reads were used to fill the intra-scaffold gaps, and a set of non-redundant unigenes was constructed with the least number of unaligned reads. Open reading frames (ORFs) were detected using TransDecoder v3.0.0 in Trinity software. For functional annotation, the BLAST program (E-value = $1e^{-5}$) was used based on unigenes and various nucleotide/protein databases, including the Nr, Swiss-Prot, GO, and COG databases. The best alignment results were retained for each database. Specifically, when the results from different databases conflicted, the priority order of NR, Swiss-Prot, KEGG, COG, and GO was used to determine the annotated unigenes. Gene expression was measured using read density; the higher the read density, the higher the level of gene expression. Gene expression was calculated using RSEM (RNA-Seq by Expectation–Maximization) v1.2.4, which measures gene expression in fragments per kilobase of exon per million mapped (FPKM) using the following formula: FPKM = 109 C/NL, where C is the number of fragments uniquely aligned to the gene of interest, N is the total number of fragments uniquely aligned with the reference genome, and L is the base length of the target gene. Finally, gene diversity analysis was performed based on FPKM read counts using HT-seq v 0.6.1 [51].

## Differential gene expression analysis

### *Dataset preparation and identification of differentially expressed genes*

Gene expression analysis was performed to identify the differentially expressed genes (DEGs) using the Galaxy web browser (https://usegalaxy.org/). Using the three RNA-seq datasets as inputs, an RNA-seq workflow (Fig. S6) was used to sequentially perform the analytical steps of the adaptor and quality trimming, mapped to a reference genome, and read the count per identified gene. In the present study, two datasets were prepared from three RNA-seq datasets ($N = 3$) of infected koala PBMC samples, including one RNA-seq dataset ($n = 1$) from koala infected with endogenous (KoRV-A; KHY-04) subtype (Group I) and two RNA-seq datasets ($n = 2$) from koalas infected with exogenous (KoRV-B and/or -C; KHB-02, and KHL-03) subtypes (Group II),. A paired-end sequence was used for each RNA-seq dataset, which was filtered using Trimmomatic [48] for adapter trimming. Subsequently, the trimmed reads were mapped against the NCBI reference koala genome sequence (Accession No.: GCF_002099425.1, Bio Project No.: PRJNA359763) using HISAT2 (hierarchical indexing for spliced alignment of transcripts). The featureCounts [52] program was used to calculate gene expression in the RNA-seq datasets. Several statistical operations were performed on the datasets to determine the DEGs. The Benjamini–Hochberg false discovery rate method was used to balance the discovery of statistically significant genes with the limitation of false positives [24]. The BioJupies Generator online server (https://maayanlab.cloud/biojupies/) was used for RNA-seq raw data analysis. In the present study, genes with absolute p-value < 0.05 and absolute value of $\log_2$ Fold Change > 1.5 were considered DEGs. Besides, we set up the p-value as < 0.05 and log Fold Change > 1.5 for an upregulated gene and log Fold Change < -1.5 for a downregulated gene. Subsequently, unique differentially expressed genes were identified using Venny 2.0 [53]. A custom-built dataset for the koala genome was used due to the absence of specific references for non-human species in BioJupies.

### *Protein–protein interaction network construction, cluster analysis, and selection of hub proteins*

A protein-protein interaction (PPI) network was built using DEGs shared between the two datasets (Group I vs. Group II). PPI networks of the shared DEGs were constructed using the STRING database (https://string-db.org/) [54], and cluster analysis was performed using the ClusterViz [55] plugin of Cytoscape v3.8.2 [56]. The MCODE (Molecular Complex Detection) algorithm was used to find highly interconnected proteins with established "parameters" including degree score threshold:2, node score threshold:0.2, and K-Core:2 [55]. Potential hubs within the PPI network were determined using different local- and global-based methods with the cytoHubba plugin in Cytoscape v3.8.2. Whereas the local method ranks hubs based on the relationship between a node and its direct neighbor, the global method ranks hubs based on the interaction between the node and the entire network. In total, five different methods were considered, including three local rank methods: degree,

maximum neighborhood component (MNC), and maximal clique centrality (MCC), and two global rank methods: edge percolated component (EPC) and closeness centrality [24]. Thereafter, the results were compared, and common nodes were identified as the most potential hubs. The PPI networks were analyzed using Cytoscape v3.8.2. Finally, the common hub proteins were identified by comparing five cytoHubba methods using the jVenn online server (http://jvenn.toulouse. cra.fr/app/example. html) [57].

#### *Functional and pathway enrichment analysis*

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) online server (https://david.ncifcrf.gov/) [58] was used to perform functional and pathway enrichment analyses of potential hub genes using the Gene Ontology (GO) [59] and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [60]. After performing an over-representation analysis, a collection of enriched cell signaling pathways and functional GO keywords was discovered, revealing the biological importance of the previously detected DEGs. Duplicate pathways were deleted, and important paths with $p$-value < 0.05 were evaluated and considered. For functional GO annotations, the biological process (BP), molecular function (MF), and cellular component (CC) datasets of GO were investigated in DAVID, and the most important GO terms were selected based on the set criteria with an adjusted p-value < 0.05. Bubble plots of the upregulated and downregulated common DEGs were drawn using SRplot (https://www.bioinformatics.com.cn/en), and chord plots of the GOs and pathways shared by the four databases were drawn using SRplot.

### Quantification of hub gene expression by RT-qPCR

The mRNA expression levels of koala hub genes (*CCNB2*, *POLR3F*, *KIF23*, and *RSL24D1*) were assessed in koala spleen tissues (K1 – K6) and PBMCs from the blood (K7) of koalas using a one-step RT-qPCR approach with Brilliant III Ultra-Fast SYBR Green RT-qPCR Master Mix (Agilent Technologies) according to the manufacturer's instructions. Primers for amplifying the hub genes were designed using Primer-BLAST (accessed on August 21, 2023, at https://www.ncbi.nlm.nih.gov/tools/primer-blast/) based on the target sequences available in NCBI GenBank (Table S8), and primers for the reference gene (koala actin) have been previously described [6]. The cycling conditions consisted of reverse transcription at 50 °C for 10 min, initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 5 s, and annealing/extension for 10 s. Each reaction was conducted in duplicate within a 20-µL volume, utilizing a 96 micro-well plate and the CFX Connect Real-Time PCR Detection System from Bio-Rad (Hercules, CA, USA). Each reaction included a no-template control and standard curve for each gene. The specificity of each PCR reaction was confirmed through melt curve analysis, with koala actin serving as the endogenous control for the normalization of results.

### Statistical analysis

The RT-qPCR data were analyzed and graphically represented using GraphPad Prism 10.0.2 (https://www.graphpad.com/scientific-software/prism/, Boston, MA, USA). One-way Analysis of Variance (ANOVA) and following Dunnett's multiple comparison tests were performed for the RT-qPCR data, and a p-value of < 0.05 was considered significant.

## Conclusion

In conclusion, our study confirms that KoRV subtypes affect koala health significantly, especially exogenous subtypes such as KoRV-A, -B, and -C. Koalas with multiple KoRV subtypes tend to have higher KoRV plasma loads, which are potential prognostic markers. Gene expression analyses revealed the immunosuppressive and oncogenic characteristics of KoRV, linking it to leukemia, lymphoma, and immunosuppression in koalas. The results of the present study enhance our understanding of KoRV pathogenesis in KoRV-infected koalas and offer insights into their health and conservation. Additionally, the DEGs identified, *CCNB2*, *POLR3F*, *KIF23*, and *RSL24D1*, provide potential therapeutic targets for KoRV-related diseases. Notably, *KIF23* and *CCNB2* show the highest expression in healthy koala PBMCs, and *POLR3F* shows the highest expression in koala with lymphoma (K1).

Although pioneering, further research with a larger cohort of multi-subtype KoRV-infected koalas is required to comprehensively understand the effects of KoRV subtypes on koala health and cancer development.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12917-024-04357-5.

Supplementary Material 1.

#### Acknowledgements

We thank Taiki Eiei, Kyoya Mochizuki, Shinsaku Ochiai, Ayaka Ito, and Nanao Ito for their contributions in caring for koalas and their help with sampling. The authors thank Saitama Children's Park for supplying the koala K3 tissues.

#### Authors' contributions

Lipi Akter, Md Abul Hashem, and Mohammad Enamul Hoque Kayesh: Conceptualization, Methodology, Software, Validation, Data curation, writing – original draft, Visualization, Investigation. Md. Arju Hossain and Fumie Maetani: Data curation, writing – original draft, Visualization, Investigation, Writing – review and editing. Rupaly Akhter, Kazi Anowar Hossain, and Md Haroon Or Rashid:

Visualization, Investigation, Writing – Review and Editing. Hiroko Sakurai and Takayuki Asai: Supervision, Software, Validation. M. Nazmul Hoque, Kyoko Tsukiyama-Kohara: Supervision, Software, Validation, Writing – review and editing.

**Funding**
This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

**Data availability**
All RNA-seq data obtained from 3 koalas' (KHB-02, KHL-03, and KHY-04) PBMC RNAs in this study were uploaded to the Sequence Read Archive (SRA) of NCBI acsession number SRR31061194, SRR31061195, and SRR31061196, respectively under bioproject PRJNA1174993 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1174993).

## Declarations

**Ethics approval and consent to participate**
The Institutional Animal Care and Use Committee of the Joint Faculty of Veterinary Medicine, Kagoshima University, reviewed and approved the study (approval number: 19K001). All the animals were treated in accordance with the protocols of the Institutional Animal Care and Use Committee for scientific purposes. The institutional research protocol was approved by AAALAC International (approval number: 001698); hence, our study was reported in accordance with the ARRIVE guidelines (https://arriveguidelines.org). To collect koala tissue samples, we have permission and the consent decree from Hirakawa Zoological Park.

**Consent for publication**
All authors listed in the manuscript have contributed significantly to the conception, design, execution, or interpretation of the study, and all have approved the final version of the manuscript submitted for publication.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Transboundary Animal Diseases Center, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima 890-0065, Japan. [2]Laboratory of Animal Hygiene, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima 890-0065, Japan. [3]Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [4]Department of Microbiology and Public Health, Patuakhali Science and Technology University, Babugonj Barishal-8210, Bangladesh. [5]Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh. [6]Hirakawa Zoological Park, Kagoshima 891-0133, Japan. [7]Awaji Farm, Park England Hill Zoo, Hyogo 656-0443, Japan. [8]Molecular Biology and Bioinformatics Laboratory, Department of Gynecology, Obstetrics and Reproductive Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur 1706, Bangladesh.

Received: 1 May 2024 Accepted: 24 October 2024
Published online: 30 October 2024

**References**

1. Xu W, Stadler CK, Gorman K, Jensen N, Kim D, et al. An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. Proc Natl Acad Sci. 2013;110:11547–52.
2. Tarlinton R, Meers J, Young P. Endogenous retroviruses: biology and evolution of the endogenous koala retrovirus. Cell Mol Life Sci. 2008;65:3413–21.
3. Waugh CA, Hanger J, Loader J, King A, Hobbs M, et al. Infection with koala retrovirus subgroup B (KoRV-B), but not KoRV-A, is associated with chlamydial disease in free-ranging koalas (Phascolarctos cinereus). Sci Rep. 2017;7:134.
4. Blyton MD, Pyne M, Young P, Chappell K. Koala retrovirus load and non-A subtypes are associated with secondary disease among wild northern koalas. PLoS Pathog. 2022;18: e1010513.
5. Hanger JJ, Bromham LD, McKee JJ, O'Brien TM, Robinson WF. The nucleotide sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. J Virol. 2000;74:4264–72.
6. Tarlinton RE, Meers J, Young PR. Retroviral invasion of the koala genome. Nature. 2006;442:79–81.
7. Hashem MA, Maetani F, Kayesh MEH, Eiei T, Mochizuki K, et al. Transmission of koala retrovirus from parent koalas to a joey in a Japanese zoo. J Virol. 2020;94:e00019-00020.
8. Shojima T, Yoshikawa R, Hoshino S, Shimode S, Nakagawa S, et al. Identification of a novel subgroup of koala retrovirus from koalas in Japanese zoos. J Virol. 2013;87:9943–8.
9. Xu W, Gorman K, Santiago JC, Kluska K, Eiden MV. Genetic diversity of koala retroviral envelopes. Viruses. 2015;7:1258–70.
10. Chappell K, Brealey J, Amarilla A, Watterson D, Hulse L, et al. Phylogenetic diversity of koala retrovirus within a wild koala population. J Virol. 2017;91:e01820–e01816.
11. Hobbs M, King A, Salinas R, Chen Z, Tsangaras K, et al. Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. Sci Rep. 2017;7:15838.
12. Zheng H, Pan Y, Tang S, Pye GW, Stadler CK, et al. Koala retrovirus diversity, transmissibility, and disease associations. Retrovirology. 2020;17:1–23.
13. Quigley BL, Timms P. Helping koalas battle disease–Recent advances in Chlamydia and koala retrovirus (KoRV) disease understanding and treatment in koalas. FEMS Microbiol Rev. 2020;44:583–605.
14. Hashem MA, Kayesh MEH, Maetani F, Eiei T, Mochizuki K, et al. Koala retrovirus (KoRV) subtypes and their impact on captive koala (Phascolarctos cinereus) health. Arch Virol. 2021;166:1893–901.
15. Quigley BL, Ong VA, Hanger J, Timms P. Molecular dynamics and mode of transmission of koala retrovirus as it invades and spreads through a wild Queensland koala population. J Virol. 2018;92:e01871–e01817.
16. MacLean D, Jones JD, Studholme DJ. Application of'next-generation'sequencing technologies to microbial genetics. Nat Rev Microbiol. 2009;7:96–7.
17. Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods. 2008;5:16–8.
18. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.
19. Ashrafi F, Nassiri M, Javadmanesh A, Rahimi H, Rezaee SA. Epigenetics evaluation of the oncogenic mechanisms of two closely related bovine and human deltaretroviruses: A system biology study. Microb Pathog. 2020;139: 103845.
20. Cheng C, Wu X, Shen Y, Li Q. KIF14 and KIF23 promote cell proliferation and chemoresistance in HCC cells, and predict worse prognosis of patients with HCC. Cancer Manag Res. 2020;Volume 12:13241–57.
21. Wu Z, Song Y, Wu Y, Ge L, Liu Z, et al. Identification of KIF23 as a prognostic biomarker associated with progression of clear cell renal cell carcinoma. Front Cell Dev Biol 2022;10:839821. https://doi.org/10.3389/fcell.2022.839821.
22. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18:1509–17.
23. Wilhelm BT, Landry J-R. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009;48:249–57.
24. Hoque MN, Sarkar M, Hasan M, Khan M, Hossain M, et al. Differential gene expression profiling reveals potential biomarkers and pharmacological compounds against SARS-CoV-2: Insights from machine learning and bioinformatics approaches. Frontiers in Immunology: 3875. 2022.
25. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep. 2019;26(1627–1640): e1627.
26. Zamani-Ahmadmahmudi M, Najafi A, Nassiri S. Reconstruction of canine diffuse large B-cell lymphoma gene regulatory network: detection of functional modules and hub genes. J Comp Pathol. 2015;152:119–30.
27. Fu Y-P, Liang Y, Dai Y-T, Yang C-T, Duan M-Z, et al. De novo sequencing and transcriptome analysis of Pleurotus eryngii subsp. tuoliensis (Bailinggu) mycelia in response to cold stimulation. Molecules. 2016;21: 560.

28. Kayesh MEH, Hashem MA, Tsukiyama-Kohara K. Toll-like receptor and cytokine responses to infection with endogenous and exogenous Koala retrovirus, and vaccination as a control strategy. Curr Issues Mol Biol. 2021;43:52–64.
29. Kayesh MEH, Yamato O, Rahman MM, Hashem MA, Maetani F, et al. Molecular dynamics of koala retrovirus infection in captive koalas in Japan. Arch Virol. 2019;164:757–65.
30. Tang Y, Zhang Y, Hu X. Identification of potential hub genes related to diagnosis and prognosis of hepatitis B virus-related hepatocellular carcinoma via integrated bioinformatics analysis. BioMed Res Int. 2020;(1):4251761. https://doi.org/10.1155/2020/4251761.
31. Ashrafi F, Ghezeldasht SA, Ghobadi MZ. Identification of joint gene players implicated in the pathogenesis of HTLV-1 and BLV through a comprehensive system biology analysis. Microb Pathog. 2021;160:105153.
32. Hashem MA, Kayesh MEH, Yamato O, Maetani F, Eiei T, et al. Coinfection with koala retrovirus subtypes A and B and its impact on captive koalas in Japanese zoos. Arch Virol. 2019;164:2735–45.
33. Batman U, Deretic J, Firat-Karalar EN. The ciliopathy protein CCDC66 controls mitotic progression and cytokinesis by promoting microtubule nucleation and organization. PLoS Biol. 2022;20: e3001708.
34. Gu P, Zhang M, Chen X, Du J, Chen L, et al. Prognostic value of cell division cycle-associated protein-3 in prostate cancer. Medicine 2023;102:36(e34655).
35. Watt, KEN, Macintosh J, Bernard G, Paul A. RNA polymerases I and III in development and disease. Semin Cell Dev Biol 2023;136:49–63.
36. Kang J, Brajanovski N, Chan KT, Xuan J, Pearson RB, et al. Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. Signal Transduct Target Ther. 2021;6:323.
37. Salmerón-Hernández Á, Noriega-Reyes MY, Jordan A, Baranda-Avila N, Langley E. BCAS2 enhances carcinogenic effects of estrogen receptor alpha in breast cancer cells. Int J Mol Sci. 2019;20:966.
38. Zhu LJ, Pan Y, Chen XY, Hou PF. BUB1 promotes proliferation of liver cancer cells by activating SMAD2 phosphorylation. Oncol Lett. 2020;19:3506–12.
39. Li Z, Yao Q, Zhao S, Wang Z, Li Y. Protein coding gene CRNKL1 as a potential prognostic biomarker in esophageal adenocarcinoma. Artif Intell Med. 2017;76:1–6.
40. Hu Y, Zheng M, Wang C, Wang S, Gou R, et al. Identification of KIF23 as a prognostic signature for ovarian cancer based on large-scale sampling and clinical validation. Am J Trans Research. 2020;12:4955.
41. Kawai T, Akira S. The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol. 2009;21:317–37.
42. Liang W-T, Liu X-F, Huang H-B, Gao Z-M, Li K. Prognostic significance of KIF23 expression in gastric cancer. World J gastrointestinal oncology. 2020;12: 1104.
43. Campbell SL, Wellen KE. Metabolic signaling to the nucleus in cancer. Mol Cell. 2018;71:398–408.
44. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, et al. Sex differences in cancer mechanisms. Biol Sex Differ. 2020;11:1–29.
45. Singh U, Hernandez KM, Aronow BJ, Wurtele ES. African Americans and European Americans exhibit distinct gene expression patterns across tissues and tumors associated with immunologic functions and environmental exposures. Sci Rep. 2021;11:9905.
46. Yamamoto R, Chung R, Vazquez JM, Sheng H, Steinberg PL, et al. Tissue-specific impacts of aging and genetics on gene expression patterns in humans. Nat Commun. 2022;13:5803.
47. Hoque MN, Istiaq A, Clement RA, Sultana M, Crandall KA, et al. Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. Sci Rep. 2019;9:1–14.
48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
49. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
50. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
51. Fonseca NA, Marioni J, Brazma A. RNA-seq gene profiling-a systematic empirical comparison. PLoS ONE. 2014;9: e107026.
52. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
53. Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams.http://bioinfogp/cnb/csic/es/tools/venny/index/html.
54. Cv M, Huynen M, Jaeggi D, Schmidt S, Bork P, et al. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31:258–61.
55. Wang J, Zhong J, Chen G, Li M, Wu F-x, et al. ClusterViz: a cytoscape APP for cluster analysis of biological network. IEEE/ACM Trans Comput Biol Bioinf. 2014;12:815–22.
56. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, et al. Cytoscape Web: an interactive web-based network browser. Bioinformatics. 2010;26:2347–8.
57. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics. 2014;15:1–7.
58. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007;35:W169–75.
59. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224:iyad031.
60. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2007;36:D480–4.


## Publisher's Note