ABSTRACT

Koala retrovirus is thought to be an underlying cause of high levels of neoplasia and immunosuppression in koalas. While epidemiology studies suggest a strong link between KoRV and disease it has been difficult to prove causality because of the complex nature of the virus, which exists in both endogenous and exogenous forms. It has been difficult to identify koalas completely free of KoRV, and infection studies in koalas or koala cells are fraught with ethical and technical difficulties, respectively. This study uses KoRV infection of the susceptible human cell line HEK293T and RNAseq to demonstrate gene networks differentially regulated upon KoRV infection. Many of the pathways identified are those associated with viral infection, such as cytokine receptor interactions and interferon signalling pathways, as well as viral oncogenesis pathways. This study provides strong evidence that KoRV does indeed behave similarly to infectious retroviruses in stimulating antiviral and oncogenic cellular responses. In addition, it provides novel insights into KoRV oncogenesis with the identification of a group of histone family genes that are part of several oncogenic pathways as upregulated in KoRV infection.

Keywords:

RNAseq, transcriptome analysis, human cell line HEK293T, infectious retroviruses

1. Introduction

Koala retrovirus, a gammaretrovirus, is currently undergoing the transition between being an exogenous (non-inherited and horizontally transmitted) virus to being an endogenous (inherited and vertically transmitted) virus integrated in the host's genome (Tarlinton et al., 2006). The virus is linked to neoplasia and clinical chlamydial disease in koalas with high viral loads (Tarlinton et al., 2005; Legione et al., 2017; Waugh et al., 2017). It was thought to be absent in some southern populations but recent work has demonstrated that at least some of these animals do in fact have defective KoRV transcripts (Tarlinton et al., 2017). It is difficult to demonstrate causality of a disease process to an endogenous virus that is present in all or most animals but transcriptomic analysis comparing animals from a population with a high viral load (Queensland, QLD) to animals from a population with a low viral load and defective KoRV transcripts demonstrated differential expression of pathways concerned with B and T lymphocyte regulation and oncogenesis (unpublished data). As obtaining completely KoRV free primary koala cell lines is technically difficult (and may be impossible given recent findings (Tarlinton

et al., 2017)) and the koala transcriptome not comprehensively annotated, this study has taken the approach of infecting well characterised continuous human cell lines with KoRV to analyse the cellular response to KoRV infection.

KoRV was categorized into several KoRV subgroups based on the phylogenetic relationships and the receptor binding differences of the *env* gene: KoRV-A (Hanger et al., 2000), KoRV-B (Xu et al., 2013b), KoRV-C, KoRV-D (Shojima et al., 2013), KoRV-E, KoRV-F (Xu et al., 2015), KoRV-G, KoRV-H, and KoRV-I (Chappell et al., 2017). For other retroviruses, mutation or recombination events in *env* genes play a significant role in pathogenicity, such as immunosuppression with variants of feline leukaemia virus, but the role of KoRV variants in causing specific types of disease has not yet been resolved (Overbaugh et al., 1988; Anderson et al., 2000; Chandhasin et al., 2005). A recent study hypothesized that KoRV-D and KoRV-E are defective viruses (Hobbs et al., 2017). It has also been hypothesized that KoRV-B is more pathogenic and is thought to play a significant role in neoplasia and chlamydia disease occurrence (Xu et al., 2013b; Waugh et al., 2017; Quigley et al., 2018).

Upon infection of a target cell, retroviruses utilise the host-cellular machinery to finish their life cycles. Following viral entry into the host cell, viral RNA is reverse transcribed and afterwards integrated into the cellular genome and the host's cell machinery is then used for generation of viral transcripts. Regarding this subsequent replication process, viral factors affect the normal cellular functions due to interaction with multiple cellular genes and proteins (Kenyon and Lever, 2011). These virus interactions mediate changes in cellular biological or physiological processes, which can be determined on a transcriptional level as demonstrated in Human immunodeficiency virus (HIV) (Giri et al., 2006), Feline immunodeficiency virus (FIV) (Ertl and Klein, 2014) or Reticuloendotheliosis virus (REV) (Miao et al., 2015). High throughput approaches such as transcriptome analysis by next generation sequencing (RNA-seq) techniques have been used in several *in vivo* or *in vitro* retroviral studies to identify differentially expressed genes from infected host or cell types (Ertl and Klein, 2014; Wu et al., 2018). The types of changes typically seen indicate immune dysregulation, (particularly interferon stimulated genes) apoptosis, viral replication and persistence. Therefore, to analyse factors contributing to KoRV pathogenesis, Human Kidney 293T (HEK293T) cells were co-cultured with KoRV-positive peripheral blood mononuclear cells (PBMCs). PBMCs were chosen as lymphocytes are likely to be producing infectious virus and have been used to isolate KoRV in several previous studies (Hanger et al., 2000; Miyazawa et al., 2011; Xu et al., 2013b). The HEK293T cell line was used as a model for KoRV infection because it gives high titres when used for retroviral production and is permissive to KoRV infection via the KoRV A receptor, phosphate transporter receptor *Pit1*, the human variant of which binds KoRV A and the KoRV B receptor thiamine transporter 1 THTR1 (Mendoza et al., 2006; Miyazawa et al., 2011; Xu et al., 2013a).

This study aimed to compare the transcriptome of KoRV-infected versus uninfected HEK293T cells to determine the impact of KoRV on host cell gene expression. The output of this study will significantly contribute to deeper insights into the complex virus-host interactions network in KoRV pathogenesis.

2. Methods

2.1 Culture of Peripheral Blood Mononuclear Cells

Blood samples were collected from 15 koalas presented at Moggill Koala Hospital (MKH), Australia Zoo Wildlife Hospital (AZWH), RSPCA Wacol and Sea World Paradise Country (SWPC) in South East QLD. After collection, blood was immediately transferred to an EDTA vacutainer tube (BD Biosciences, Australia) and transported at +4°C to the University of Queensland, Gatton Campus. Ethical approval for this study was granted by the University of Queensland (UQ) Animal Ethics Committee, permit number ANFRA/SVS/461/12 and ANRFA/SVS/445/15, the Queensland Government Department of Environment and Heritage Protection permit number WISP11989112.

PBMCs were separated from whole blood using Ficoll-Paque PLUS (GE Healthcare). The blood was diluted with an equal volume of sterile HBSS (without calcium or magnesium) (Sigma Aldrich). Diluted blood samples were layered onto the Ficoll-Paque solution at 3:4 ratio. The tube was then centrifuged at 400 g for 30 min at 18°C. The mononuclear cell layer was removed and resuspended in 3 volumes of Hanks solution and centrifuged at 100 g for 10 min at 18°C. This step was repeated twice. Samples that were observed to contain a red clump within the washed PBMC pellet were treated to remove RBC contamination. The pellet was suspended with 1 ml Red cell lysis buffer (Sigma) and incubated at 37°C for 10 mins, followed by 3 washes with cell culture media. Otherwise the pellet was directly suspended in 1 ml RPMI1640 growth medium containing 20% heat inactivated fetal calf serum (FCS), 1 mM sodium pyruvate, 1% MEM non-essential amino acid, 0.2% primocin and 100 units/ml recombinant interleukin-2. A cell count was performed through gentle mixing of cells with 0.4% trypan blue solution at 1:1 ratio and counts were done in triplicate using hemocytometer under a 40× objective following the standard methodology (Louis and Siegel, 2011). Cells were diluted to a concentration of 10^6 cells/ml using RPMI1640 growth media. The presence of lymphocytes was checked using a cytospin and automated GIEMSA stainpack, and examined using a light microscope. The percentage of lymphocytes was counted manually.

2.2 Cell line

Human embryonic kidney (HEK) 293T cells (kindly provided by Professor Alex Khromykh, School of Chemistry & Molecular Biosciences, University of Queensland, Australia) were cultured in growth medium comprising Dulbecco's modified eagle's medium (DMEM) containing L-glutamine, glucose,

sodium pyruvate (Thermofisher Scientific) supplemented with 5% heat-inactivated FCS and 0.25% primocin (InvivoGen). These cells were cultured at 37°C in a humidified atmosphere of 5% CO_2 in air.

2.3 KoRV passage in HEK293T cell cultures

HEK293T cells (~10^6 cells/ml) were added into growth medium in a T75 culture flask (SARSTEDT) and incubated for 16 - 20 hrs at 37°C in a humidified atmosphere of 5% CO_2 in air. The medium was removed, and suspensions of freshly isolated and non-stimulated PBMCs with ~10^6 cells/ml concentration were directly introduced at 40 - 50% cellular density of HEK293T cells and incubated for 1 hr at 37°C for cell adsorption. Then growth medium was added, and the cultures incubated at 37°C until HEK293T cells had reached 100% cell confluency. The confluency was checked using an inverted microscope every 8 hrs. After 100% confluency was attained, culture medium was removed from the flask, centrifuged at 500g for 5 min and the supernatant was introduced into a fresh 24hr HEK293T cell culture flask and passaged in this manner for up to five cycles. At each passage, cells adhering to the culture flask were harvested through trypsinization, centrifuged and the pellet stored for RNA extraction. In each co-culture and passage, non-infected HEK293T cell line were used as a control.

Cells were harvested using a cell scraper (Thermofisher Scientific). Following centrifugation, the cell pellet of each passage was lysed using trizol and RNA was extracted using a Qiagen miRNeasy kit according to manufacturer's instructions with on-column DNase digestion (Qiagen) procedure to remove residual DNA. Quality and concentration of total RNA was assessed by NanoDrop2000 spectrophotometer and 2% agarose gel electrophoresis.

2.4 Quantification of viral RNA

RT-qPCR targeting the *pol* gene was used to determine the KoRV viral RNA copy number in the cell pellet (HEK293T infected culture). The previously published real time PCR primers and probe for KoRV *pol* gene were used (Tarlinton et al., 2005). Reaction mix for the RT-qPCR consisted of 12.5 µl of 2x reaction mix, 0.5 µl of SuperScript® III RT/Platinum® Taq Mix (SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase, Invitrogen), 1 µl of primers (forward and reverse 10 µM final concentration), 0.5 µl of probe (10 µM final concentration), 5.5 µl of ultrapure water, and 5 µl of template. Cycling conditions were: reverse transcription at 50°C for 30 min, Superscript platinum taq activation at 95°C for 2 min, followed by 45 cycles of denaturation at 95°C for 15 sec and annealing/extension at 60°C for 30 sec. Standard curves were prepared for each run using RNA standards, and results for each run measured against these standard curves. The samples were run in triplicate on a BioRad CFX 96 system.

2.5 RNA sequencing through Next Generation sequencing platform

Among 15 cultured samples only five maintained growth until passage five. Five samples of KoRV infected HEK293T culture pellets with a KoRV viral load $>10^5$ copies/ml were taken from passage 4 cells for RNA extraction and were sent to the Ramaciotti Centre for Genomics at the University of NSW for processing and sequencing. RNA extracted from cell pellets of five non-infected HEK293T cell cultures was used as a control. RNA quality was checked through an Agilent 2100 Bioanalyzer. The mRNA library was prepared for sequencing from 1 µg of total RNA using the Illumina TruSeq stranded mRNA library prep kit. These steps included poly-A pulldown and adapter ligation at both ends. The paired-end (PE) libraries were sequenced as 2X 76 bp PE using the Illumina NextSeq500 platform.

2.6 RNA-seq Analysis through Bioinformatics tools

The NextSeq500 platform generated four lane reads, which were uploaded into the Galaxy server of University of Queensland for all subsequent analyses. Forward and Reverse reads over four lanes were joined using *Text manipulation - Concatenate datasets/tail-to-head* options. The joined paired end reads were run through FastQC to check the quality of raw unpaired sequence reads. *Trimmomatic* (Bolger et al., 2014) was used to remove TrueSeq3 adapter sequence and low-quality reads to prevent noise from nonsense reads in downstream analysis. Individual *Fastq* files were checked for read quality with FastQC software (version 0.10.1). The trimmed paired reads were mapped to the human reference genome *hg19* using HISAT2 (Kim et al., 2015). Duplicate reads were then removed using *PicardTools - Mark Duplicates - Clean SAM*. Mapping statistics were tabulated using *SAMTools/Flagstat* (Li et al., 2009). Gene read counts were generated with *HTSeq-count* (http://www-huber.embl.de/users/anders/HTSeq/doc/index.html) using the 'union' option and each gene FPKM (expected number of fragments per kilobase of transcript sequence per million base pairs sequenced) was calculated based on gene length and counts of paired-end reads mapped to the gene. These counts were used to identify differential expression of genes (DEGs) among samples with Differential count models of R/Bioconductor package, EdgeR (Robinson et al., 2010). KoRV infected and uninfected HEK293T cell line samples were designated as treatment and control group, respectively. Genes having expression values less than one count per million were removed from both libraries. Ordinary deviance estimates of Robust dispersion method in the EdgeR package were used to estimate each feature's dispersion parameter and false discovery rate (FDR), a value of 0.05 was set to correct the threshold of *p*-value. Significantly differentially expressed genes were further filtered with Log 2-fold change ≥ 2. Gene ontology analyses on biological processes (BP) were performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang da et al., 2009a; Huang da et al., 2009b). The significant enrichment score analysed based on hypergeometric distribution. Following Benjamaini-Hochberg methodology, the *p*-value was corrected and p value ≤ 0.05 was considered

representative of significant gene enrichment. To further investigate the biological pathways involving the identified DEGs, a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using Cytoscape installed ClueGo plug-in. Multiple enrichment tools were available but at the time of analysis, the ClueGo plug-in KEGG database was the most recently updated and thus was selected for use. In this enrichment analysis, a two-sided hypergeometric test was used following the Human GO database and Bonferroni correction for multiple testing-controlled p-values and GO with corrected $p < 0.05$ was considered significantly enriched.

2.7 Genetic diversity of KoRV

To identify KoRV RNA presence and quantify transcript level in the cell line (HEK 293T) after infection, each infected sample was mapped using HISAT2 tools against KoRV-A (AF151794), KoRV-B (KC779547.1), KoRV-C (AB828005.1), KoRV-D (AB828004.1), KoRV-E (KU533853.1), KoRV-F (KX588025.1), KoRV-G (KX587961.1), KoRV-H (KX588036.1) and KoRV-I (KX587976.1). The mapped reads were quantified using SALMON tools (Patro et al., 2017), quantification output generates transcript Length, transcripts per Million (TPM) and estimated number of reads (an estimate of the number of reads drawn from this transcript given the transcript's relative abundance and length).

3. RESULTS

3.1 RNA sequencing and assembly

KoRV *pol* RNA was amplified from the co-culture pellet of PBMC-HEK293T cells from 10 koalas out of a total 15, consisting of six healthy and four diseased animals. Results are shown in Table 1. The cultures from five of these koalas remained KoRV RNA positive at the 4th passage in HEK293T cells. There was no cytopathic effect observed in the KoRV-positive HEK293T cells compared to uninfected control cells. Ultimately 10 samples were selected for RNA-seq analysis to characterize the host transcriptional response to KoRV infection, where five were 4th passage KoRV infected HEK293T cells, designated as "treated" and five were non-infected HEK293T cells, designated as "control". Extracted RNA yields varied between samples giving between 1500-3500 ng/µl total RNA. The A260/280 ratio of all purified RNA samples was between 2.0 and 2.2 and the A260/230 ratio values of all samples were greater than 2.0. Gel electrophoresis confirmed that the majority of extracted RNA samples were present in intact form with no DNA contamination. Using NextSeq500 2x 76 bp Illumina platform, at least 76 million reads per sample were obtained. After removal of adapter sequences and low quality reads, samples were mapped against the human reference genome *hg19*, with at least 85.30 % paired end reads mapped. The mapping percentages are detailed in Table 2. Sequence reads are available in NCBI database, accession number PRJNA514936.

The replication of KoRV in the cell line was further confirmed by mapping the sequences against the full genome of KoRV-A (Accession no AF151794). All "treated" samples were KoRV positive when mapped against the KoRV-A genome, values are presented as Transcript level per million (TPM) in Table 3.

3.2 Differentially expressed genes and functional enrichment analysis

To determine the effects of KoRV infection on the host transcriptome, differentially expressed genes were analysed by comparing KoRV infected and non-infected HEK293T cells. In total, 4701 genes were significantly differentially expressed (p- value ≤ 0.05). When these were further filtered to identify genes that were differentially expressed by at least a two-fold log change (up or down), 1458 genes were identified where 973 genes were up-regulated, and 485 genes were down-regulated. Of these genes, only 526 up-regulated and 234 downregulated genes were annotated with *Homo sapiens*, the full list of genes is presented in supplementary file 1. The distribution of DEGs is presented in Figure 1.

Many immune associated genes were up-regulated in the KoRV infected cells with significantly high levels of enrichment. For example, genes that showed increased expression included those related to interferon responses such as *IFITM1* (interferon induced transmembrane protein 1; fold change 2), *SECTM1* (secreted and transmembrane 1; fold change 3.1), *OAS2* (2'-5'-oligoadenylate synthetase 2; fold change 2.9) and *TNFSF10* (tumor necrosis factor (ligand) superfamily, member 10; fold change 4.2). Other up-regulated immune related genes included chemokine C-C motif ligand (*CCL1*, *CCL2*, *CCL5*, *CCL10* and certain interleukins and their receptors, including *IL6*, *IL8*, *IL17F*, *IL32*, interleukin 2 receptor alpha (*IL2RA*) and interleukin 2 receptor gamma (*IL2RG*).

In contrast, genes showing decreased expression were Micro(mi)RNA, such as microRNA 4442(MIR4442), microRNA 1289-1(MIR1289-1), microRNA 3657(MIR3657), microRNA 548l(MIR548L), and microRNA 221(MIR221). MiRNAs are involved in the regulation of biological processes, including immune response and apoptosis and their pathogenic role, when dysregulated, has been widely studied in the malignant diseases and autoimmune disorders (Jakymiw et al., 2006; Long et al., 2016). Some downregulated genes related to homeostatic processes, such as *ARRDC3* (arrestin domain containing 3; fold change -2.11), *SLC24A4* (solute carrier family 24 member 4; fold change -2.49) and *SLC26A4* (solute carrier family 26 member 4; fold change -2.36) were also identified.

To further characterize the biological function of the DEGs after KoRV infection, a functional enrichment GO and KEGG pathway analysis was performed. Significantly up-regulated DEGs were involved in 74-biological processes (BP). These included 'negative regulation of viral genome replication', 'nucleosome assembly', 'defense response to virus', 'type I interferon signaling pathway', 'telomere organization' and 'immune response'. Meanwhile, down-regulated DEGs involved 13 BP.

Some processes were related to cell communication or signalling such as adenylate cyclase-inhibiting G-protein coupled receptor signalling pathway, G-protein coupled purinergic nucleotide receptor signalling pathway and phospholipase C-activating G-protein coupled receptor signalling pathway. The top 10 upregulated and downregulated biological processes are illustrated in Figure 2. The full list is presented in supplementary file 2 and 3.

The KEGG pathway enrichment results (Figure 3) showed that the DEGs were involved mostly in viral disease or response pathways with significant overlap in "cytokine-cytokine receptor interaction, "influenza A", "herpes simplex infection", "NOD-like receptor signalling pathway". Similarly, there was significant overlap between "viral carcinogenesis, "systemic lupus erythematosus" and "transcriptional misregulation in cancer". The alcoholism pathway can be explained by the overlap with the Systemic lupus erythematosus pathway due to the presence of the histone family gene clusters and may not be important by itself. Only up-regulated genes were involved in the biological pathways and none of the down-regulated genes had any association with biological pathways. Genes involved in the KEGG pathways are listed in supplementary file 4.

3.3 Assessment of KoRV subgroup expression levels

Before proceeding to further analysis, KoRV expression after cell line integration was further confirmed using mapping against KoRV-A, KoRV-B, KoRV-C , KoRV-D, KoRV-E, KoRV-F, KoRV-G, KoRV-H and KoRV-I. All five infected samples were KoRV-A positive with high TPM values and 3 samples were additionally KoRV-B positive, with lower TPM values compared to their KoRV-A levels. Results are depicted in table 3.

4. Discussion

This study was designed to gain insight into host response to KoRV infection and to investigate pathogenesis upon KoRV infection using RNA-seq based transcriptome analysis of infected and uninfected cell lines. Our results indicated that there were stark differences in the levels of expression of certain genes in the host after KoRV infection. The majority of the up-regulated and down-regulated genes are involved with immune functions and are reflective of an immune response to the viral challenge.

It had been previously reported that Koala PBMCs could be isolated from whole blood following Ficoll-Hypaque isolation procedure and growth maintained with 9% FCS and 1% koala serum (Miyazawa et al., 2011). In the current study, PBMCs were isolated from blood following Ficoll-Hypaque isolation using 20 % FCS instead of koala serum. KoRV was successfully grown in HEK293T cells and virus replication was steady with each passage, growth didn't increase markedly with each passage although

there was a trend to higher copy number in the 4th passage of the cell cultures from 3 koalas, perhaps suggesting virus was adapting to growth in these cells.

Host response to the initial entry of KoRV into cells was characterised by significant upregulation of genes involved in the innate immune response and the inhibition of viral replication which was not unexpected. The innate immune response plays a significant role in viral invasion resistance during early infection. This is triggered when specific pattern recognition receptor (PRRs) families detect microbial pathogens and act to increase expression of type I interferons and other cytokines (Kawai and Akira, 2009). Toll-like receptor 2 (TLR2), which is involved in recognition of viruses and their products, was characterized as a significantly up-regulated gene (other TLR genes were not differentially regulated). The up-regulation of toll-like receptors and subsequent effect on inflammatory response has already been demonstrated in other retrovirus infections such as HIV-1 (Hernandez et al., 2012).

Negative regulation of viral genome replication was characterised as the most significantly up-regulated biological process. This is not surprising, because activation of this part of the body's first line of defence occurs when a virus invades a host. Several interferon (IFN)-stimulated genes (ISG's) were significantly up-regulated in this process. Interferon induced transmembrane protein 1 and 3 (*IFITM1* and *IFITM3*) and *IFNB1* are part of the network of ISGs, and have previously been shown to have antiviral properties and an ability to suppress HIV virus replication (Lu et al., 2011). Among other up regulated genes, *TRIM22* (tripartite motif protein 22) is known to inhibit Gag assembly (Barr et al., 2008) and *APOBEC3A* (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A) is known to limit HIV-1 replication through hypermutation at reverse transcription (Sheehy et al., 2002; Malim and Emerman, 2008). The JAK-STAT (Janus kinase/signal transducer and activator of transcriptions) signaling pathway is employed by multiple cytokines and interferons and is crucial for the development of both innate and adaptive immunity (Yan et al., 2018). However, the up-regulation of ISGs, which is the hallmark of host response against viral infection, suggests continual, but controlled expression of viral replication association genes in KoRV infection of HEK293T cells.

As expected, the Cytokine-cytokine receptor interaction pathway was significantly up-regulated in response to KoRV infection. The 21 significant differently expressed up-regulated genes detected in this pathway (*CCL2, CCL26, CCL5, CSF3R, CXCL1, CXCL10, CXCL8, GH1, GH2, IFNB1, IFNL1, IFNL2, IFNL3, IL17F, IL2RA, IL2RG, IL6, INHBA, INHBE, TNFRSF14, TNFSF10*) may be a set of genes related to early general host response against infections (Perry et al., 1997; Maher et al., 2014). Among other chemokines, *CCL2* is unique in having enhancing effects on viral replication and pathogenesis and higher levels are associated with counterproductive immune activation and inflammation of HIV infected subjects (Covino et al., 2016). *IL-6* modulates naive T cells

differentiation into *Th1* or *Th2* cells (Diehl and Rincon, 2002; Rodriguez et al., 2010) and has involvement with increasing inflammation and promoting chlamydial growth (Rodriguez et al., 2010). The significant up-regulation of *IL-6* and *CCL2* genes suggests an antiviral immune response which is expected at earlier infection.

An association between KoRV infection and development of different types of neoplasia including lymphoid leukaemia, mesothelioma, craniofacial tumor and osteochondroma has been demonstrated (Tarlinton et al., 2005). The current study investigated the effect of KoRV infection on a cell line to better understand KoRV pathogenesis. Multiple oncogenes and tumor genes were differentially expressed. For example, *CSF3R* which is highly related with leukaemogenesis (Ikewaki et al., 2012) was upregulated as was *MMP9* which contributes significantly to tumour growth and angiogenesis and is overexpressed in lymphocytic leukaemia (Casabonne et al., 2011). Several oncogenes including REL (v-rel reticuloendotheliosis viral oncogene homolog (avian)), members of RAS oncogene family, *RAB17* and *RAB38*, *TLX1* (T-cell leukemia homeobox 1) were also identified as being upregulated. Transcriptional misregulation in cancer and viral carcinogenesis pathways were upregulated, the histone family genes (*HIST1H3B*, *HIST1H3D*, *HIST1H3E*, *HIST1H3H*) involved in these pathways are linked to cell cycle progression defects and DNA damage (Mei et al., 2017). The pathway level analysis of differentially expressed genes in the network milieu provides novel insights into the pathogenesis of neoplasia in koalas. However, the results need to be interpreted with some caution as the studied cell line was not of koala origin and the cells were only infected only for a short time. Tumor development begins when a cell sustains a genetic mutation that increases its propensity to divide and decreases its responsiveness to normal growth control mechanisms. Depending on the mutation and cell type this can lead to one or a series of the following: hyperplasia, dysplasia, and benign, in-situ or invasive cancer. Clearly, further mechanistic studies to clarify the particular role of the oncogenes and their interaction with viral proteins in KoRV diseases are warranted.

It would have been preferable to perform this study in a koala cell line. However, there are several practical difficulties with this approach. There are currently no available continuous cell lines of koala origin. The cell line reported by (Girjes et al., 2003) proved in the end not to be of koala origin (unpublished data). Hence all koala cell cultures are currently primary cell lines with limited ability to sustain passage in cell culture. It has also proven virtually impossible to isolate koala cells that do not contain KoRV. Animals previously thought to be KoRV-free in southern Australia have been shown in our recent work (Tarlinton et al., 2017) to have defective variants of the virus missing the polymerase gene. It is likely that through superinfection interference, these defective KoRV variants would interact with infectious KoRV in the cell cultures. Thus, without access to koala cell lines or KoRV-free primary cell cultures, this work was performed in the model system of human HEK293T cells.

Animals from the QLD populations that these samples were derived from, have been shown to all possess KoRV-A, with a more variable number also having KoRV B usually present at a 10 fold lower expression level than KoRV A, consistent with the virus variants seen in this study (Sarker et al., 2019).. The receptors for KoRV A and B (PiT 1 and THTR1) are both present in the HEK293T cell line so it is unsurprising that both these variants have been passaged in this cell line. What is more interesting is the absence of other variants (KoRVs C-I) which are known to be present in the blood of QLD animals (Chappell et al., 2017; Sarker et al., 2019). In particular the KoRV D variant has been shown to be present in all animals in this population, in some cases at a 10 fold higher level in the blood than KoRV A (Quigley et al., 2019). Those variants of KoRV D that have been fully sequence characterised are replication defective (Hobbs et al., 2017) and may either arise *de-novo* within animals or require a helper virus to transmit between animals. The receptors for KoRVs C-I are not known. As subgroups C-I have not sub-passaged in the HEK293T cells despite almost certainly being present in the original samples, these appear to be incapable of sustained replication in this human cell line. Whether this is due to the absence of necessary receptors or the viruses themselves not being fully replication competent (or a combination of both) remains to be determined. This study certainly supports existing evidence that the KoRV A and B subgroups are transmissible (Fiebig et al., 2006; Shojima et al., 2013; Xu et al., 2013a) in a variety of non-koala cell lines and animal models. To date, no studies had been conducted on the association of LTR (long terminal repeat) genetic changes with KoRV disease pathogenicity although, along with the *env* gene, LTR also contributes to disease pathogenicity in feline leukaemia virus (FeLV) infection of cats (Matsumoto et al., 1992) and avian leukosis virus (ALV) in chickens (Zavala et al., 2007). The LTR of FeLV contains two or three tandem direct repeats of enhancer elements in lymphoma positive cats, while non-neoplastic strains contain only a single copy of the enhancer (Miura et al., 1989; Fulton et al., 1990; Matsumoto et al., 1992). A previous study highlighted the presence of four 17 bp tandem repeats in a KoRV-B isolate LTR while three 37 bp tandem repeats were found in a KoRV-J isolate LTR and one tandem repeat in KoRV-A (Shimode et al., 2014). It is possible that differential expression of KoRV env subtypes might be linked with LTR enhancer or promoter activity.

This study clearly demonstrates that active KoRV infection induces antiviral and oncogenic responses in a naive human cell line, supporting the hypothesis that KoRV is, like HIV an immunosuppressive and oncogenic virus. It lends support to the epidemiological evidence that KoRV is the underlying cause of the very high rates of leukaemia, lymphoma and immunosuppression (leading to chlamydial disease) seen in koalas (Hanger et al., 2000; Waugh et al., 2017). This study also provides insight into possible pathogenic mechanisms for neoplasia development in KoRV infected koalas. The study also provides support to the idea that the KoRV A and B subgroups of virus are fully replication competent and transmissible with subgroups C-I not demonstrating the ability to be sub-passaged in the human cell line used in this study. In the future, the approach and methodology of this study could be applied to

koala specific cell lines, if those that are completely KoRV free can be identified and maintained in culture.

Supplementary Information:

Supplementary file 1: Complete list of differentially expressed genes

Supplementary file 2: GO enrichment analysis of biological processes illustrated by up-regulated differentially expressed genes

Supplementary file 3: GO enrichment analysis of biological processes illustrated by down-regulated differentially expressed genes

Supplementary file 4: Differentially expressed genes involved in KEGG pathway

Acknowledgment

This project and scholarship for NS were funded by the Queensland Department of the Environment and Heritage Koala Research Grant Programme 2012. NS was also supported by a Keith Mackie Lucas travel scholarship from the University of Queensland. Koalas for post mortem were accessed through the Moggill Wildlife hospital (QLD Department of the Environment and Heritage Protection) and Fauna Rescue of South Australia and we are extremely grateful for the staff and volunteers that work with these organisations for their work in koala rescue

AUTHOR CONTRIBUTIONS

N.S. performed RNA extraction, laboratory experiments, bioinformatics analysis and drafted manuscript. J. Meers, J.S., G.S. and H.O. helped in laboratory experiment set up, data interpretation and manuscript preparation. R.D.E and R.T reviewed the manuscript and bioinformatics analysis. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interest.

Reference

391 Anderson, M.M., Lauring, A.S., Burns, C.C. and Overbaugh, J., 2000. Identification of a cellular cofactor
392 required for infection by feline leukemia virus. Science 287, 1828-30.
393 Barr, S.D., Smiley, J.R. and Bushman, F.D., 2008. The interferon response inhibits HIV particle
394 production by induction of TRIM22. PLoS Pathog 4, e1000007.
395 Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence
396 data. Bioinformatics 30, 2114-20.
397 Casabonne, D., Reina, O., Benavente, Y., Becker, N., Maynadié, M., Foretová, L., Cocco, P., González-
398 Neira, A., Nieters, A., Boffetta, P., Middeldorp, J.M. and de Sanjose, S., 2011. Single nucleotide
399 polymorphisms of matrix metalloproteinase 9 (MMP9) and tumor protein 73 (TP73) interact
400 with Epstein-Barr virus in chronic lymphocytic leukemia: results from the European case-
401 control study EpiLymph. Haematologica 96, 323-327.
402 Chandhasin, C., Coan, P.N. and Levy, L.S., 2005. Subtle mutational changes in the SU protein of a
403 natural feline leukemia virus subgroup A isolate alter disease spectrum. J Virol 79, 1351-60.
404 Chappell, K.J., Brealey, J.C., Amarilla, A.A., Watterson, D., Hulse, L., Palmieri, C., Johnston, S.D., Holmes,
405 E.C., Meers, J. and Young, P.R., 2017. Phylogenetic Diversity of Koala Retrovirus within a Wild
406 Koala Population. J Virol 91.
407 Covino, D.A., Sabbatucci, M. and Fantuzzi, L., 2016. The CCL2/CCR2 Axis in the Pathogenesis of HIV-1
408 Infection: A New Cellular Target for Therapy? Curr Drug Targets 17, 76-110.
409 Diehl, S. and Rincon, M., 2002. The two faces of IL-6 on Th1/Th2 differentiation. Mol Immunol 39, 531-
410 6.
411 Ertl, R. and Klein, D., 2014. Transcriptional profiling of the host cell response to feline
412 immunodeficiency virus infection. Virol J 11, 52.
413 Fiebig, U., Hartmann, M.G., Bannert, N., Kurth, R. and Denner, J., 2006. Transspecies transmission of
414 the endogenous koala retrovirus. J Virol 80, 5651-4.
415 Fulton, R., Plumb, M., Shield, L. and Neil, J.C., 1990. Structural diversity and nuclear protein binding
416 sites in the long terminal repeats of feline leukemia virus. J Virol 64, 1675-82.
417 Giri, M.S., Nebozhyn, M., Showe, L. and Montaner, L.J., 2006. Microarray data on gene modulation by
418 HIV-1 in immune cells: 2000-2006. J Leukoc Biol 80, 1031-43.
419 Girjes, A.A., Lee, K.E. and Carrick, F.N., 2003. Establishment and characterization of a new epithelial
420 cell line, KC-1, from koala (Phascolarctos cinereus) conjunctiva. In Vitro Cell Dev Biol Anim 39,
421 110-3.
422 Hanger, J.J., Bromham, L.D., McKee, J.J., O'Brien, T.M. and Robinson, W.F., 2000. The nucleotide
423 sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus
424 related to Gibbon ape leukemia virus. J Virol 74, 4264-72.
425 Hernandez, J.C., Stevenson, M., Latz, E. and Urcuqui-Inchima, S., 2012. HIV type 1 infection up-
426 regulates TLR2 and TLR4 expression and function in vivo and in vitro. AIDS Res Hum
427 Retroviruses 28, 1313-28.
428 Hobbs, M., King, A., Salinas, R., Chen, Z., Tsangaras, K., Greenwood, A.D., Johnson, R.N., Belov, K.,
429 Wilkins, M.R. and Timms, P., 2017. Long-read genome sequence assembly provides insight
430 into ongoing retroviral invasion of the koala germline. Sci Rep 7, 15838.
431 Huang da, W., Sherman, B.T. and Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward
432 the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1-13.
433 Huang da, W., Sherman, B.T. and Lempicki, R.A., 2009b. Systematic and integrative analysis of large
434 gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57.
435 Ikewaki, J., Kawano, R., Sato, T., Imamura, T., Kohno, K., Ogata, M., Ohtsuka, E. and Kadota, J., 2012.
436 An acquired CSF3R mutation in an adult chronic idiopathic neutropenia patient who
437 developed acute myeloid leukaemia. Br J Haematol 157, 264-6.
438 Jakymiw, A., Ikeda, K., Fritzler, M.J., Reeves, W.H., Satoh, M. and Chan, E.K., 2006. Autoimmune
439 targeting of key components of RNA interference. Arthritis Res Ther 8, R87.
440 Kawai, T. and Akira, S., 2009. The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol
441 21, 317-37.

Kenyon, J.C. and Lever, A.M., 2011. The molecular biology of feline immunodeficiency virus (FIV). Viruses 3, 2192-213.
Kim, D., Langmead, B. and Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357-60.
Legione, A.R., Patterson, J.L., Whiteley, P., Firestone, S.M., Curnick, M., Bodley, K., Lynch, M., Gilkerson, J.R., Sansom, F.M. and Devlin, J.M., 2017. Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease. J Med Microbiol 66, 236-244.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-9.
Long, H., Yin, H., Wang, L., Gershwin, M.E. and Lu, Q., 2016. The critical role of epigenetics in systemic lupus erythematosus and autoimmunity. J Autoimmun 74, 118-138.
Louis, K.S. and Siegel, A.C., 2011. Cell viability analysis using trypan blue: manual and automated methods. Methods Mol Biol 740, 7-12.
Lu, J., Pan, Q., Rong, L., He, W., Liu, S.L. and Liang, C., 2011. The IFITM proteins inhibit HIV-1 infection. J Virol 85, 2126-37.
Maher, I.E., Griffith, J.E., Lau, Q., Reeves, T. and Higgins, D.P., 2014. Expression profiles of the immune genes CD4, CD8beta, IFNgamma, IL-4, IL-6 and IL-10 in mitogen-stimulated koala lymphocytes (Phascolarctos cinereus) by qRT-PCR. PeerJ 2, e280.
Malim, M.H. and Emerman, M., 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe 3, 388-98.
Matsumoto, Y., Momoi, Y., Watari, T., Goitsuka, R., Tsujimoto, H. and Hasegawa, A., 1992. Detection of enhancer repeats in the long terminal repeats of feline leukemia viruses from cats with spontaneous neoplastic and nonneoplastic diseases. Virology 189, 745-9.
Mei, Q., Huang, J., Chen, W., Tang, J., Xu, C., Yu, Q., Cheng, Y., Ma, L., Yu, X. and Li, S., 2017. Regulation of DNA replication-coupled histone gene expression. Oncotarget 8, 95005-95022.
Mendoza, R., Anderson, M.M. and Overbaugh, J., 2006. A putative thiamine transport protein is a receptor for feline leukemia virus subgroup A. Journal of virology 80, 3378-3385.
Miao, J., Bao, Y., Ye, J., Shao, H., Qian, K. and Qin, A., 2015. Transcriptional Profiling of Host Gene Expression in Chicken Embryo Fibroblasts Infected with Reticuloendotheliosis Virus Strain HA1101. PLoS One 10, e0126992.
Miura, T., Shibuya, M., Tsujimoto, H., Fukasawa, M. and Hayami, M., 1989. Molecular cloning of a feline leukemia provirus integrated adjacent to the c-myc gene in a feline T-cell leukemia cell line and the unique structure of its long terminal repeat. Virology 169, 458-61.
Miyazawa, T., Shojima, T., Yoshikawa, R. and Ohata, T., 2011. Isolation of koala retroviruses from koalas in Japan. J Vet Med Sci 73, 65-70.
Overbaugh, J., Donahue, P.R., Quackenbush, S.L., Hoover, E.A. and Mullins, J.I., 1988. Molecular cloning of a feline leukemia virus that induces fatal immunodeficiency disease in cats. Science 239, 906-10.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417-419.
Perry, L.L., Feilzer, K. and Caldwell, H.D., 1997. Immunity to Chlamydia trachomatis is mediated by T helper 1 cells through IFN-gamma-dependent and -independent pathways. J Immunol 158, 3344-52.
Quigley, B.L., Ong, V.A., Hanger, J. and Timms, P., 2018. Molecular Dynamics and Mode of Transmission of Koala Retrovirus as It Invades and Spreads through a Wild Queensland Koala Population. J Virol 92.
Quigley, B.L., Phillips, S., Olagoke, O., Robbins, A., Hanger, J. and Timms, P., 2019. Changes in endogenous and exogenous Koala Retrovirus (KoRV) subtype expression over time reflects koala health outcomes. J Virol.

492 Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential
493 expression analysis of digital gene expression data. Bioinformatics 26, 139-40.
494 Rodriguez, N., Dietrich, H., Mossbrugger, I., Weintz, G., Scheller, J., Hammer, M., Quintanilla-Martinez,
495 L., Rose-John, S., Miethke, T. and Lang, R., 2010. Increased inflammation and impaired
496 resistance to Chlamydophila pneumoniae infection in Dusp1(-/-) mice: critical role of IL-6. J
497 Leukoc Biol 88, 579-87.
498 Sarker, N., Fabijan, J., Seddon, J., Tarlinton, R., Owen, H., Simmons, G., Thia, J., Blanchard, A.M.,
499 Speight, N., Kaler, J., Emes, R.D., Woolford, L., Trott, D., Hemmatzadeh, F. and Meers, J., 2019.
500 Genetic diversity of Koala retrovirus env gene subtypes: insights into northern and southern
501 koala populations. J Gen Virol 100, 1328-1339.
502 Sheehy, A.M., Gaddis, N.C., Choi, J.D. and Malim, M.H., 2002. Isolation of a human gene that inhibits
503 HIV-1 infection and is suppressed by the viral Vif protein. Nature 418, 646-50.
504 Shimode, S., Nakagawa, S., Yoshikawa, R., Shojima, T. and Miyazawa, T., 2014. Heterogeneity of koala
505 retrovirus isolates. FEBS Lett 588, 41-6.
506 Shojima, T., Yoshikawa, R., Hoshino, S., Shimode, S., Nakagawa, S., Ohata, T., Nakaoka, R. and
507 Miyazawa, T., 2013. Identification of a novel subgroup of Koala retrovirus from Koalas in
508 Japanese zoos. J Virol 87, 9943-8.
509 Tarlinton, R., Meers, J., Hanger, J. and Young, P., 2005. Real-time reverse transcriptase PCR for the
510 endogenous koala retrovirus reveals an association between plasma viral load and neoplastic
511 disease in koalas. J Gen Virol 86, 783-7.
512 Tarlinton, R.E., Meers, J. and Young, P.R., 2006. Retroviral invasion of the koala genome. Nature 442,
513 79-81.
514 Tarlinton, R.E., Sarker, N., Fabijan, J., Dottorini, T., Woolford, L., Meers, J., Simmons, G., Owen, H.,
515 Seddon, J.M., Hemmatzedah, F., Trott, D., Speight, N. and Emes, R.D., 2017. Differential and
516 defective expression of Koala Retrovirus reveal complexity of host and virus evolution.
517 bioRxiv, 211466.
518 Waugh, C.A., Hanger, J., Loader, J., King, A., Hobbs, M., Johnson, R. and Timms, P., 2017. Infection with
519 koala retrovirus subgroup B (KoRV-B), but not KoRV-A, is associated with chlamydial disease
520 in free-ranging koalas (Phascolarctos cinereus). Sci Rep 7, 134.
521 Wu, X., Zeng, Y., Lu, R., An, Y., Yu, S., Zhao, J., Wu, Y., Wu, B., Wang, Q. and Huang, Y., 2018.
522 Transcription analysis of the interaction between chicken thymus and recombinant avian
523 leukosis virus isolate FJ15HT0. Virus Res 244, 147-152.
524 Xu, W., Gorman, K., Santiago, J.C., Kluska, K. and Eiden, M.V., 2015. Genetic diversity of koala retroviral
525 envelopes. Viruses 7, 1258-70.
526 Xu, W., Stadler, C.K., Gorman, K., Jensen, N., Kim, D., Zheng, H., Tang, S., Switzer, W.M., Pye, G.W. and
527 Eiden, M.V., 2013a. An exogenous retrovirus isolated from koalas with malignant neoplasias
528 in a US zoo. Proceedings of the National Academy of Sciences 110, 11547-11552.
529 Xu, W., Stadler, C.K., Gorman, K., Jensen, N., Kim, D., Zheng, H., Tang, S., Switzer, W.M., Pye, G.W. and
530 Eiden, M.V., 2013b. An exogenous retrovirus isolated from koalas with malignant neoplasias
531 in a US zoo. Proc Natl Acad Sci U S A 110, 11547-52.
532 Yan, Z., Gibson, S.A., Buckley, J.A., Qin, H. and Benveniste, E.N., 2018. Role of the JAK/STAT signaling
533 pathway in regulation of innate immunity in neuroinflammatory diseases. Clin Immunol 189,
534 4-13.
535 Zavala, G., Cheng, S. and Jackwood, M.W., 2007. Molecular epidemiology of avian leukosis virus
536 subgroup J and evolutionary history of its 3' untranslated region. Avian Dis 51, 942-53.

537