Differential and defective expression of Koala Retrovirus indicate complexity of host and virus evolution

R.E Tarlinton[1, 8], A.R. Legione[6], N. Sarker[2], J. Fabijan[3], J. Meers[2], L. McMichael[2], G.Simmons[2], H.Owen[2], J.M.Seddon[2], G. Dick[3], J.S. Ryder[4], F. Hemmatzedah[5], D.J., Trott[5], N. Speight[5], N. Holmes[7], M. Loose[7], R.D.Emes[1]

1. School of Veterinary Medicine and Science, University of Nottingham, UK
2. School of Veterinary Science, The University of Queensland, Australia
3. Longleat Safari Park, United Kingdom, Durrel Wildlife Conservation Trust, United Kingdom
4. Garston Veterinary Group, Somerset, United Kingdom
5. School of Animal and Veterinary Sciences, University of Adelaide, Australia
6. Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Australia
7. School of Life Sciences, University of Nottingham, UK
8. Corresponding Author: Rachael.tarlinton@nottingham.ac.uk

Word Count Abstract:

Word Count Manuscript:

Abstract

Koala retrovirus (KoRV) is unique amongst endogenous (inherited) retroviruses in that its incorporation to the host genome is still active, providing an opportunity to study what drives this fundamental process in vertebrate genome evolution. Animals in the southern part of the natural range of koalas were previously thought to be either virus free or to have only exogenous variants of KoRV with low rates of KoRV induced disease. In contrast, animals in the northern part of their range universally have both endogenous and exogenous KoRV with very high rates of KoRV induced disease such as lymphoma. This paper uses a combination of sequencing technologies, Illumina RNA sequencing of "southern" (south Australian) and "northern" (SE QLD) koalas and CRISPR enrichment and nanopore sequencing of DNA of "southern" (South Australian and Victorian animals) to retrieve full length loci and intregration sites of KoRV variants. We demonstrate that koalas that tested negative to the KoRV *pol* gene qPCR, used to detect replication competent KoRV, are not in fact KoRV free but harbour defective, presumably endogenous, "RecKoRV" variants that are not fixed between animals. This indicates that these populations have historically been exposed to KoRV and raises questions as to whether these variants have arisen by chance or whether they provide a protective effect from the infectious forms of KoRV. This latter explanation would offer the intriguing prospect of being able to monitor and selectively breed for disease resistance to protect the wild koala population from KoRV induced disease.

Introduction

Koalas (*Phascolarctos cinereus*) are an iconic marsupial species listed as vulnerable on the IUCN 'red list' of threatened species [1]. While a large part of their ongoing population decline is due to habitat loss, two major disease threats, chlamydial infection and Koala Retrovirus (KoRV), are additionally limiting population viability [2]. These infections are particularly prevalent in the northern regions of Australia, namely

the states of Queensland and New South Wales, and less so in the south (South Australia, Victoria) [3,4].

Following European settlement, large koala populations across Australia declined significantly due to hunting in the 1890's to 1920's, with southern populations nearing extinction. During this time, small refuge populations were established on offshore Victorian islands and these koalas have been used subsequently to restock most of their former southern range. This southern population is genetically distinct from the northern animals [5] with a more limited genetic diversity[6]. The history of translocations in southern animals is complex but the original founder populations of French and Phillip Islands are thought to have been the source for most mainland Victorian animals with potential remnant populations of greater diversity in the Strzelecki ranges [5]. The mainland Mount Lofty Ranges koala population in South Australia originates from koalas from both the Kangaroo Island population, populated by koalas from French Island [7] as well as koalas from Queensland and New South Wales [5,8].

Endogenous retroviruses (ERVs) are those that have become incorporated into their host's genome. They are ubiquitous in vertebrate genomes and in some cases constitute up to 10% of total genome content [9]. They are usually not functional viruses due to the accumulation of mutations and deletions but are often expressed at an RNA level, where they are thought to play a role in genomic regulation [9 10,11]. They are known in some cases to provide essential functions to their hosts, such as the syncytin genes responsible for placental fusion in many species [12,13] as well as their role in stem cells, reproductive tissue and early embryos [14]. However their effects on the host upon initial entry to the host genome are not clear. KoRV is part of a small group of unusual "modern" endogenous retroviruses (including Murine leukaemia virus, Feline leukaemia virus and Jaagsietke sheep retrovirus). These modern ERVs are replication competent and display considerable overlap with their exogenous infectious counterparts, including swapping of gene segments [15,16].

KoRV is one of the most recent entrants into any known mammalian genome, with estimates of integration time somewhere between 200 and 49,000 years ago [17,18].It

is thought to have arisen from a recent species jump as its closest relatives are endogenous viruses in two subspecies of *Melomys burtoni* (the grassland mosaic tailed rat) in northern Australia and Indonesia [19,20], Gibbon ape leukaemia virus (GALV), a pathogenic exogenous virus that most likely arose as a spill over event from south east Asian rodents in the late 1960s [21] and Flying fox retrovirus, a very recently described exogenous virus of black flying foxes (*Pteropus alecto*) [22] [23].

KoRV was originally identified during investigations into the high rates of lymphoid neoplasia (lymphoma and leukaemia) in Queensland koalas [17]. Koalas with lymphoid neoplasia have significantly higher KoRV viral loads [24] [25] and some strains of KoRV also influence the cytokine response profile of koala lymphocytes [26]. Recent studies have indicated that somatic insertions of KoRV perturb oncogenes and underlie the very high rate of cancer in KoRV A positive animals [27]. Multiple studies also indicate that high KoRV viral loads (in northern populations) or positive PCR status (in southern populations) [28-32] are linked to clinical chlamydial disease, probably as a factor of retroviral induced immunosuppression.

KoRV has been found in 100% of Queensland and New South Wales koalas but appears to have a lower prevalence in southern populations [4,28,30,32-34]. The virus displays a high diversity in proviral copy number and integration sites between individuals and populations, with southern animals having lower copy numbers in their DNA [24,35] [4]. Somatic insertions are also apparent against a background of endogenous insertions in northern animals [27].

A number of sequence variants of the *env* gene region, which encodes the surface unit (SU) of the envelope protein (Env), have also been identified (Figure 1). These vary between individuals and resemble the viral quasispecies common to infectious retroviruses, with clades referred to as A to J [34,36]. The originally identified virus is now known as KoRV A and appears to be present in all individuals that are KoRV-positive [27,28,30,33,37]. Various koala genome sequencing studies indicate that only KoRV A is endogenised in northern animals with other variants present at lower than one copy/genome equivalent, indicating that they are not present in all tissues or cells of an

animal [27,35,38]. A recent study indicated that there may be one KoRV A locus shared amongst most (perhaps all) northern animals, which perhaps represents the original endogenisation event [27]. KoRV A infections in southern animals may represent genuine exogenous (infectious) virus as these are in many cases also present at less than one copy per genome equivalent [4]. The non-A variants may also represent genuine exogenous (infectious) virus in both northern and southern animals, circulating independently with these present as low copy number/somatic insertions[27,38,39], not detected in all animals [29,30,34,40-42] and display a pattern of detection in family groupings consistent with a maternally transmitted infection [42 40 33 43]. Some caution is necessary in interpreting this however as phylogenetic analysis of the envelope variants from a variety of sequencing studies do not clearly indicate chains of transmission [34,36,41] and, by analogy with infectious retroviruses in other species (for instance FeLV in cats), many envelope sequence variants may arise from KoRV A within individual infections rather than transmitting from animal to animal [15] or may be transmitted as a co-infection with KoRV-A. This is particularly likely for many of the "D" group of variants that do not appear to be replication competent [34,44,45].

There has been much debate as to whether the B/J variant, which displays a different receptor usage to KoRV A is more pathogenic as these variants have been epidemiologically linked with clinical disease in some studies but not others [46 47 29 33]. This may however be a factor of the sensitivity of diagnostic methods used as at least one study has demonstrated that koalas with higher viral loads display greater quasispecies diversity and are more likely to test positive on PCR based tests for non-A variants [36]. That study also demonstrated that viral diversity is much higher in RNA (actively replicating virus) than DNA (copies inserted either endogenously or from initial infection) from the same animal.

Genomic sequencing studies have also demonstrated that there are a number of other older endogenous retroviruses and transposable elements within the koala genome [6,38,48,49]. One of these, Phascolarctid endogenous retroelement (PhER), is found frequently in northern koala genomes in recombination with KoRV. These recombinant

KoRV "RecKoRV" structures typically consist of the 5′ LTR and 5′ end of the KoRV *gag* gene, approximately 5 Kb of the 3′ end of PhER and its LTR, followed by the 3′ end of the KoRV env gene and KoRV 3′LTR [38,49] (Figure 1). There appear to be multiple variants of these that arise from very similar recombinations at particular points in the KoRV/PhER genomes. They are not shared between all animals but do display some geographical clustering in loci that are shared between individuals and may be absent in some populations [49]. Variants of KoRV A with large indels or "Solo LTRs" (where the middle part of the virus is spliced out during cellular DNA replication) are also seen [38].

This study reports the presence of RecKoRV variants in Southern animals that do not carry KoRV A. These variants appear to be a different genetic lineage to that present in northern animals and to be present (though not fixed) in all animals tested from multiple Victorian and South Australian populations, including the founder population on French Island.

Methods

Ethics

Ethical approval for this study was granted by the University of Queensland Animal Ethics Committee, permit number ANFRA/SVS/461/12, the Queensland Government Department of Environment and Heritage Protection permit number WISP11989112, the University of Adelaide Animal Ethics Committee permit number S-2013-198 and the South Australian Government Department of Environment, Water and Natural Resources Scientific Research Permit Y26054, the University of Nottingham School of Veterinary Medicine and Science Clinical Ethics Research Panel, and Department of Environment and Primary Industries (Victoria, Australia) (Research Permit 10006924).

Samples for DNA sequencing

DNA sequencing was performed on samples from five southern koalas (Table 1 B). Spleen samples from three wild Victorian animals were collected at necropsy as outlined in [30]. Liver samples were collected from one 3 year old female South Australian koala

166 housed in a zoological park in the UK that had been recently imported from an Australian
167 captive population derived from the Mt Lofty ranges and Kangaroo Island population in
168 SA. This animal died of the kidney disease oxalate neprosis with samples of liver
169 collected at post mortem and stored at -20°C until DNA extraction and sequencing.
170 Lymph node samples were collected from one wild Mt Lofty (SA) that died as a result of
171 dog attack as described in [25].
172
173 **Samples for RNA seq**
174 Samples were collected from wild-rescued koalas euthanised for clinical reasons
175 and submitted for post-mortem examinations from South East Queensland (Greater
176 Brisbane) (n=10) and South Australia (Mount Lofty Ranges) (n=19). Age was
177 determined by dentition and the amount of wear on the upper premolar [50] (Table 1 A).
178 Full details of these animals are presented in [6]. Submandibular lymph nodes were
179 collected within 2-6 hours of death into RNALater® and stored at -80°C. Where possible,
180 blood was collected into EDTA prior to euthanasia (BD vacutainer) with whole blood and
181 plasma added to RNA later as per previous studies [51] kept at -80°C. Of the ten koalas
182 from South East Queensland (QLD), six were male and four female and all were adults,
183 with a tooth wear class (TWC) 4 or 5. Nineteen koalas were sampled from the Mount
184 Lofty Ranges, South Australia (SA); seven female and 12 male. Six were juvenile (TWC 1
185 or 2) and 13 were adults (TWC 3 or 4).
186 **Nanopore Sequenced animals**
187 **DNA extraction for nanopore sequencing**
188 Genomic DNA was extracted from frozen liver/spleen tissue that had been ground into a
189 fine powder under liquid nitrogen. The Qiagen Genomic Tip (100/G) kit (Qiagen; 10243)
190 was used to extract DNA from 100 mg of tissue powder. DNA was quantified using the
191 Qubit Fluorometer (Thermo Fisher Scientific) and the Qubit dsDNA BR Assay Kit (Thermo
192 Fisher Scientific; Q32853) and the molecular weight was assessed using the Agilent
193 TapeStation 4200 and the Agilent Genomic DNA ScreenTape Assay (Agilent; 5067-5365
194 and 5067-5366). A sequencing library was prepared using the Genomic DNA by Ligation

Kit (Oxford Nanopore Technologies; SQK-LSK109) and run on a PromethION flow cell (Oxford Nanopore Technologies; FLO-PRO002) for 72 hours on a PromethION beta sequencer (Oxford Nanopore Technologies).

Nanopore Sequencing for KoRV insertions

Cas9-mediated PCR-free enrichment was performed to identify individual KoRV insertion sites. Genomic DNA was also extracted as described above or was extracted from spleen tissue, that had been stored in RNAlater (ThermoFisher) at -80°C, using the Qiagen PureGene DNA extraction Kit (Qiagen; 158445).

Genomic DNA was dephosphorylated to inhibit binding of Oxford Nanopore sequencing adapters to non-specific DNA fragments. Six custom Alt-R CRISPR-Cas9 crRNA (Integrated DNA Technologies) were used to form Cas9 ribonucleoprotein complexes (RNPs) that would facilitate strand-specific cleavage at target sites within KoRV. Cleaved ends were simultaneously dA-tailed to facilitate directional ligation of sequencing adapters and enrich for reads initiating at these crRNA cleavage sites. Lyophilized crRNA were reconstituted to 100 uM TE (pH7.5) and pooled in equimolar amounts. Cas9-mediated enrichment, sequencing library preparation and sequencing were then performed according Oxford Nanopore Technologies Cas-mediated PCR-free enrichment protocol (Version: ENR_9044_v1_xxxx_08Aug18); and each library was run on a separate MinION flow cell (Oxford Nanopore Technologies; FLO-MIN106 R9.4.1) on the GridION X5 Mk1.

Nanopore Sequencing of PCR amplicons

PCR amplification was conducted using the primer set KRV R2 forward (ATCTACCCGGAGACGGACAG) and reverse (GCCGGTACCTATACCTGCTG) [25] to amplify an approximately 6kb fragment of the KoRV genome from extracted genomic DNA from the SA koala K15-012. A sequencing library was prepared using the Rapid Sequencing Kit (Oxford Nanopore Technologies; SQK-RAD004) and run on a MinION flow cell (Oxford

Nanopore Technologies; FLO-MIN106D) for 36 hours on a MinION sequencer (Oxford Nanopore Technologies).

Sequence Assembly and Mapping

Nanopore sequences were basecalled using guppy and reads that passed the default read filtering metrics were obtained. Reads for each koala were mapped to the KoRV retrovirus reference genome (Genbank Accession number: AF151794) using minimap2 [52] and samtools [53]. Read mapping was visualised using Geneious Prime software (Biomatters, New Zealand) and reads were truncated to retain regions upstream and downstream of the KoRV genome. These truncated reads were then mapped against the koala reference genome assembly (Genbank Accession number: GCA_002099425.1) using minimap2 with no secondary hits allowed. The mapped reads were visualised in Genious Prime to identify the directionality of the insert, whether the insert potentially interrupted coding regions of the koala genome, and identify upstream genes that could be influenced by insertion. Additionally, reads were mapped to a sequence of PhER [54].

All reads mapping to KoRV for each koala were assembled using flye [55] in order to obtain a consensus insert assembly. Additionally, reads that mapped to individual contigs of the koala reference genome, representing individual insert sites, were extracted and assembly was also attempted using flye.

Animals for RNAseq

RNA preparation for RNAseq

Total RNA was extracted from lymph nodes using an RNeasy Mini kit with on column DNAase1 digestion (Qiagen). RNA quantity and quality were assessed via anXpose spectrophotometer (Bioke) and Agilent 2100 Bioanalyzer. mRNA was prepared for sequencing using the Illumina TruSeq stranded mRNA library prep kit and 100 base pair, paired end sequencing was performed on an Illumina HiSeq. Details of the koalas, sample quality and read quantity are provided in Supplementary data 1.

RNA and DNA extraction for qPCR/PCR (SA and QLD animals)

252 DNA was extracted from 100 µL of EDTA blood using a DNeasy blood and tissue kit
253 (Qiagen). Where available RNA was extracted from plasma Using the QIAMp Viral RNA
254 mini kit with on-column Qiagen RNase free DNAse digestion. The extracted RNA and DNA
255 was stored at -80°C for RT-PCR (RNA) and PCR (DNA) as required.

256 **KoRV qPCR**

257 The presence of KoRV provirus for individual gene segments was assessed by
258 qPCR for the KoRV A *pol* gene (the standard KoRV diagnostic assay)[31] on DNA extracted
259 from whole blood as reported in [25].

260 **KoRV genome coverage**

261 To reduce mis-mapping due to the abundance of highly repetitive long terminal
262 repeat sequences, the adapter-trimmed fastq files were first mapped using Hisat2 [56] to
263 the isolated Long Terminal Repeat (LTR) region of the koala KoRV type sequence
264 (accession AF151794). LTR depleted reads were then mapped to representative
265 sequences of KoRV A and RecKoRV derived from the koala reference genome (KoRV45
266 and RecKoRV6 Supplementary data 2 and 3) [56]. Per-base coverage was determined from
267 bam files for each isolate using samtools version 1.3.1 depth (with parameters –aa –q
268 10 –d 20000).

269 **KoRV envelope variant gene expression**

270 To quantitate the expression of KoRV envelope variants, LTR depleted reads for
271 individual koalas were pseudoaligned to the *gag*, *pol* and *env* genes of KoRVA
272 (accessions AAF15097.1_1, AAF15097.1_2 and AAF15097.1_3 respectively) and the first
273 575 nucleotides of the envelope variants of the non-A KoRV variants B-I (accessions
274 AB822553.1, AB828005.1, AB828004.1, KX588043.1, KX587994.1, KX587961.1,
275 KX588036.1 and KX588021.1 respectively) and the 3' overlap of PhER/KoRV in RecKoRV
276 using Kallisto [57]. These nucleotides correspond to the hypervariable region of the *env*
277 gene that is used in KoRV envelope variant classification.

278

279 **Data Availability**

KoRV sequence data (as fasta formatted data) are available from adac figshare [https://figshare.com/authors/Adac_uon_Adac_uon/566308]. Raw RNA sequence reads available in FASTQ format at ENA with the accession number PRJEB21505. Nanopore sequence data is available via accession number PRJNA770362.

Results

RNA from submandibular lymph nodes from 10 QLD and 19 SA animals was subjected to paired end illumina sequencing (HiSeq 100bp) and was mapped to representative KoRVA and RecKoRV sequences from the koala reference genome (Figure 3) . Demographic data for individual animals are presented in Supplementary data set 1.

When normalised for total mapped read depth, coverage was very similar for both the SA and QLD groups of koalas across the ends of the KoRV genomes (LTR-*gag*, and *env*-LTR). However, between positions 1389 and 7124 of the KoRV A sequence the SA group showed a mean coverage of < 10% of the QLD group suggesting that part of *gag*, all of *pro-pol* and part of the *env* genes were largely missing in the RNA transcripts, with six SA koalas not expressing this region at all (Figure 3 A) . The target site of the standard KoRV *pol* qPCR used in most studies is contained within this missing region [51]. Data from other publications from this sample cohort indicate that some of SA animals were KoRV PCR positive for the proviral *pol* gene (and other genes) suggesting that at least partial proviruses for this region were present but were expressed at levels undetectable in the transcriptome[25].

The higher number of reads in the *env* and LTR regions of the QLD animals can be explained by the presence of spliced *env* transcripts in addition to full length genomic transcripts as has been reported by other groups [54], although these are not detected as complete individual transcripts by the mapping methods used in this study.

Mapping of the reads to RecKoRV demonstrated relatively even coverage from the QLD animals, However there was little to no coverage of the 3' portion of the PhER segment of RecKoRV in the SA animals, indicating that while there are RecKoRV

sequences in the SA animals these likely differ in sequence from those in the genome animal (Figure 3 B)

Pseudomapping of the sequence reads to the KoRV A genome (complete *gag, pro-pol* and *env* genes) and type sequences of the hypervariable region of the *env* gene (base pairs 6000-6575 of KoRV A) of each of the previously identified KoRV envelope variants (KoRV A to I as per the classification scheme used in Chappell *et.al.* 2016 [34]) demonstrated that while QLD koalas had multiple envelope variants within individuals, SA animals had far lower KoRV envelope variant diversity. Significantly higher expression was observed for KoRV A,B,D,E and G variants in QLD compared to SA samples (unpaired t-test with unequal variance) (Figure 4 and Supplementary data 4). It was observed that QLD animals were older (mean tooth wear class 4.22 95% CI 3.88-4.56) than SA (mean tooth wear class 3.05 95% CI 2.58-3.52) and so age may confound KoRV expression comparisons. When the same test was repeated for samples from koalas with the same tooth class 4 (7 QLD 8 SA samples), expression of A,B,E and G variants remained significantly different between locations (Supplementary data 5), supporting the finding that KoRV env expression is significantly higher in the QLD than the SA populations. Eleven out of nineteen SA animals (58%) had KoRV A. Six of these koalas had only KoRV A reads (Figure 4 and Table 2). Four animals had reads for KoRV A and one other variant only (D or E). Two animals had reads for KoRV E but no detectable reads for any other variant (including KoRV A). Only one SA koala (Z Table 2) had counts comparable to the QLD cohort with a similar range of variants (A, B, C, D, E, F, G, I), while the rest had counts that were <10% of the QLD koalas. *Pol* gene counts were also similarly considerably lower in the SA koalas than the QLD group. Relative expression as estimated count values for individual animals for each gene region and KoRV envelope variant are presented in Supplementary data 4.

Mapping of CRISPR enriched nanopore sequences from koala samples to the KoRV reference genome identified a clear drop in coverage across the main portion of the genome. This went from base 450 in the *gag* coding region to base 1134 in the *env*

337 coding region, or bases 1411 - 7040 across the KoRV-A reference genome (Figure 5).
338 Mapping to a PhER assembly identified improved coverage, but there were still clear
339 regions of near zero coverage in the mid-region of the reference (Figure 5). Importantly
340 the three samples that had previously tested negative to KoRV using conventional PCR
341 targeting the *pol* gene (Koalas 01 03 and 04) both had reads mapping to KoRV, but no
342 coverage in the region of the PCR targets. Alignment of PhER and KoRV A from the
343 (northern) reference genome animal and the sequence variants found in the southern
344 animals is presented in Supplementary data 6. Assembly of reads that mapped to the
345 koala reference genome generated 17 contigs containing RecKoRV variants (8 from K01-
346 SA1-CRISPR, 7 from K01-SA1-WG, and 1 each from K03-Vic23 and K04-Vic31). The
347 general structure of these inserts were similar across the assemblies besides a ~579 bp
348 gap at the 5′ end at the interruption of the KoRV gag gene. Aligning all read sets back to
349 one of the RecKoRV variants from Koala 01 showed that this insert was present across
350 all koala samples (Figure 5).
351
352 Mapping of CRISPR enriched nanopore sequences from four koala samples identified
353 potential KoRV insert locations on 30 koala reference genome contigs (filtering this to
354 require at least five reads mapping at the same site in at least one koala to constitute an
355 insertion point). The data from koala 5 could not be mapped in this way as the PCR and
356 sequencing strategy excluded the insertion sites. A summary of insert sites and read
357 mapping is available in Table 3. Of the predicted insertion points (Figure 6), eight were
358 shared between samples, with koala 1 sharing insert sites with koalas 3 (2 contigs) and
359 4 (1 contig), and koala 3 sharing sites with koala 2 (1 contig) and 4 (4 contigs). No
360 insertion sites were shared between all koalas
361
362 An outline of interrupted genes, or genes downstream of KoRV insert sites, is presented
363 in Table 4. Of the 30 insertion sites determined by mapping reads to the koala reference
364 genome, 10 occurred within annotated genes, typically in predicted introns.
365

Discussion

The findings of the current study suggest that KoRV infection involves a more complex host-viral relationship than previously recognised, particularly in SA and Victorian koalas. Other studies have shown differences between northern and southern koala populations in the prevalence of KoRV infection, levels of KoRV proviral and viral loads and disease burden[25,58]. This study has revealed additional viral factors that indicate these population differences are more complicated than merely presence or absence of virus and virus load.

The results of this study were unexpected. Instead of these southern animals having demonstrably no KoRV as expected from a preliminary PCR based KoRV pol screen it was evident in the RNAseq study that they do in fact have at least partial KoRV sequences. Long read nanopore based DNA sequencing subsequently demonstrated that these sequences are a variant of the "RecKoRV" recombinant retroelements demonstrated in northern animals[49]. These are a recombination between the middle portion of an older retrotransposon in the koala genome and partial sequences of the 5′ and 3′ ends of KoRV (with the structure LTR-partial gag- central portion of PhER, - partial TM unit of env and LTR). The southern koala sequences are apparently of a different lineage to those found in the northern animals with the substitution of an unidentified piece of DNA between the KoRV and PhER sequences that is not present in the reference genome animal.

A comparison of differing sequencing methods (whole genome nanopore sequencing), the CRISPR enrichment and a PCR and nanopore sequencing strategy demonstrates that the CRISPR method produced greater read coverage and depth to resequencing the entire genome from the same animal and has the distinct advantage of being considerably cheaper (c £1000 cw £20,000). The PCR and long read sequencing in comparision was both challenging to get a PCR that worked and produced a lower read coverage and poorer homology. These sequences were also shorter than the expected

6000 Bp and likely represent mis-priming and amplification of the KoRV sequences in the PCR. This strategy also does not produce sequence information on the insertion site of the sequences. The PCR mispriming is not unexpected as the repetitive nature of the LTRs frequently results in poor PCR amplification from genomic DNA (where there are multiple copies of these ERVs) with many other studies also failing to amplify full length KoRV proviruses from koala DNA with PCR [17,40,59]. Partial segment PCRs of the KoRV genome (LTR-*gag, gag*, part of *pol*, *env* in two parts) on DNA extracted from blood samples from SA (results presented in [25]) demonstrated that many SA animals that test negative on the standard KoRV qPCR have at least some of the missing KoRV segments in their DNA. This indicates that there may be low copy number (likely somatic) infections of KoRV present in addition to these high copy number germline RecKoRV sequences.

Koalas with these RecKoRV variants would have been identified as KoRV negative in previous studies as the standard tests for the virus are conventional PCR or qPCR assays targeting the portion of the *pol* gene that is missing in these sequences [4,30,51]. Other studies using KoRV *pol* PCR tests for proviral loci in DNA have also indicated that at least some southern animals have this gene but at much lower copy numbers than in QLD animals [4]. The pattern of deletion for more ancient retroviral loci is one of loss of the *env* genes with maintenance of the *gag-pol* genes to facilitate spread within invidual cells [60]. The replication defective variants missing their *pro-pol* genes in the current study indicate that the drivers of retroviral endogenisation in the face of an infectious virus challenge are very different to the long term ones in well adapted virus/host systems.

These RecKoRV variants are clearly replication defective and are unlikely to have colonised the genome by themselves. They may have originally arisen by being "carried" along with replication competent viruses as occurs for other retroviruses such as Rous Sarcoma Virus [61]. It seems likely that these variants along with infectious KoRV were present before the southern animals were genetically isolated in the 1920's and that infectious KoRV alleles either never integrated into the genome of these animals or were

lost due to the genetic bottlenecks in the Southern animals[6]. The presence of the RecKoRV variants in the Victorian animals, particularly in the animal from the founder population of French Island indicates that it is likely that all southern animals have these, calling into question whether genuinely KoRV free animals exist. Examining further animals in these populations for these variants alongside genomic KoRV A is a priority. Intriguingly these insertions do not appear to be fixed between animals or populations with only a few loci shared (and none between all animals). This is comparabile to the KoRV insertion patterns seen in the northern animals [27] and indicates multiple colonisation events over time. It may indicate ongoing intracellular transposition as has been hypothesised as the mechanism for the proliferation of defective variants in older endogenised retroviruses in other species [60]. It is also possible that depth of coverage in some animals has missed some loci and follow up studies, including a larger number of animals will be essential to confirm the distribution of these defective loci across the southern koala population.

The host genetic restriction in the SA population may also have resulted in animals with viral receptor allelles that are unable to bind infectious KoRV, restricting infectious virus replication and transmission and preventing endogenisation of infectious KoRV. This situation occurs in several mouse strains resistant to certain murine leukaemia virus strains [62], though to date there are no known variations between southern and northern koalas for the KoRV A and B receptors, Pit1, and THTR1 and our transcriptomics screen of the two populations did not highlight these genes as varying between northern and southern animals [6 44,47]. It is also possible that mutations in other genes important in retroviral replication (such as retroviral restriction factors) differ between the two populations resulting in restricted replication in the SA animals, although these were not obvious in our genomic screen [6] and this remains to be explored.

Blockade of infectious retroviruses by defective variants has been reported for several other mammalian endogenous/exogenous retroviruses. Receptor blockade by defective Env proteins occurs in Jaagsietke sheep retrovirus (JSRV)[63], in part explaining

453 the tissue tropism of the exogenous virus for tissues where the endogenous variants are
454 not expressed. Endogenous JSRV loci also exert a further block on exogenous viral
455 replication at the viral assembly stage, where defective Gag proteins from the ERV loci
456 are packaged along with infectious variants preventing the viral particles from being
457 packaged and transported correctly for viral release from the cell. Receptor blockade by
458 endogenous Env proteins has also been reported in Murine Leukaemia virus variants in
459 mice, along with a Gag mediated block at the pre-integration step of viral replication [64].
460 In this respect a number of lncRNAs were identified downstream of KoRV inserts
461 that may play a regulatory function in expression of genes in the reverse orientation of
462 KoRV insert sites. However the distance between each of these inserts and the
463 associated genes is notable. One example of this is the XPR1 gene (Xenotropic and
464 polytropic retrovirus receptor 1) which is a receptor for certain gammaretroviruses, at
465 which two koalas (Koala 01 and Koala 03) have inserts (K01 – 48 reads; K03 – 47
466 reads) 500 kb upstream from the lncRNA.
467 While we do not yet know which of these scenarios is responsible for the marked
468 difference in KoRV profiles between northern and southern animals, they raise the
469 intriguing possibility that these replication defective transcripts may be interfering in
470 some way with the full length virus variants completing their replication cycle. Future
471 work will need to include in vivo studies of the truncated variants identified here and
472 whether these variants do (and at what stage) blockade infectious virus replication.
473 It is also possible that as the southern animals (at least the ones in this study)
474 are not born with endogenised KoRV A, they are not immune tolerised to the virus and
475 are more able to mount an effective immune response to it. This would potentially
476 explain the variations in antibody profiles against KoRV A evident between northern and
477 southern animals and the very much lower KoRV induced disease prevalence between
478 the two populations[25,65,66].
479 This study does not resolve the issue of which (if any) of the identified KoRV
480 envelope variants is the transmissible version of the virus. As has been reported in many
481 other studies [34,40,44,46] our northern animals display considerable variation in their KoRV

envelope variants as would be expected for a infectious replicating retrovirus. Our SA animals (with the exception of one animal), display a much more limited env variant diversity (where there are detectable reads at all) with animals expressing env genes limited to variants A, D,and E. Animal Z was the only SA animal with reads other than these three variants. We have previously reported that SA animals (whether KoRV A positive or not) display a reduced viral load and diversity compared with their QLD counterparts [36]. It may be that these KoRV positive animals represent those with exogenous rather than endogenous KoRV as has been posited several times [67]and are better able to control virus replication.

The discovery of these replication defective KoRV sequences in SA animals has opened up a number of intriguing implications for both controlling disease in koala populations and the drivers of retroviral endogenisation in their hosts. The hypothesis that the replication defective variants may blockade infectious KoRV replication, if substantiated, opens up the option to use selective breeding to re-introduce this trait into the KoRV susceptible northern population, though this would need to be done with caution given the presence of other deleterious genetic mutations such as those responsible for the high incidence of oxalate nephrosis [68] in southern animals.

Acknowledgements

This project was 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 Mogill Wildlife hospital (QLD Department of the Environment and Heritage Protection) and the Adelaide Koala and Wildlife Hospital, Plympton, South Australia and Fauna Rescue of South Australia Inc. AL was supported by the VESKI Victoria Fellowship 2018. Long leat Safari park has also received support through the non-profit Koala Life foundation for work with South Australian koalas.

Conflict of Interest Statement

The authors declare no conflict of interest

Figures and Tables

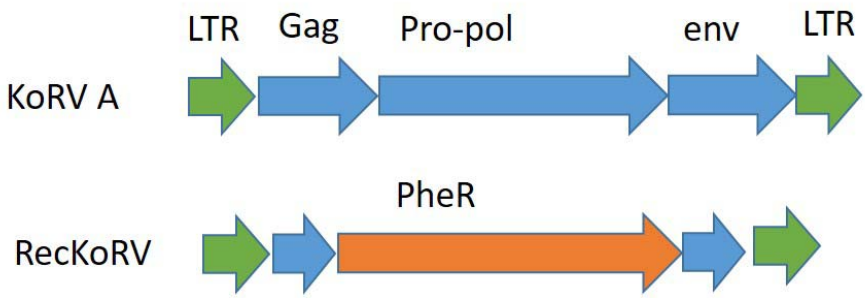
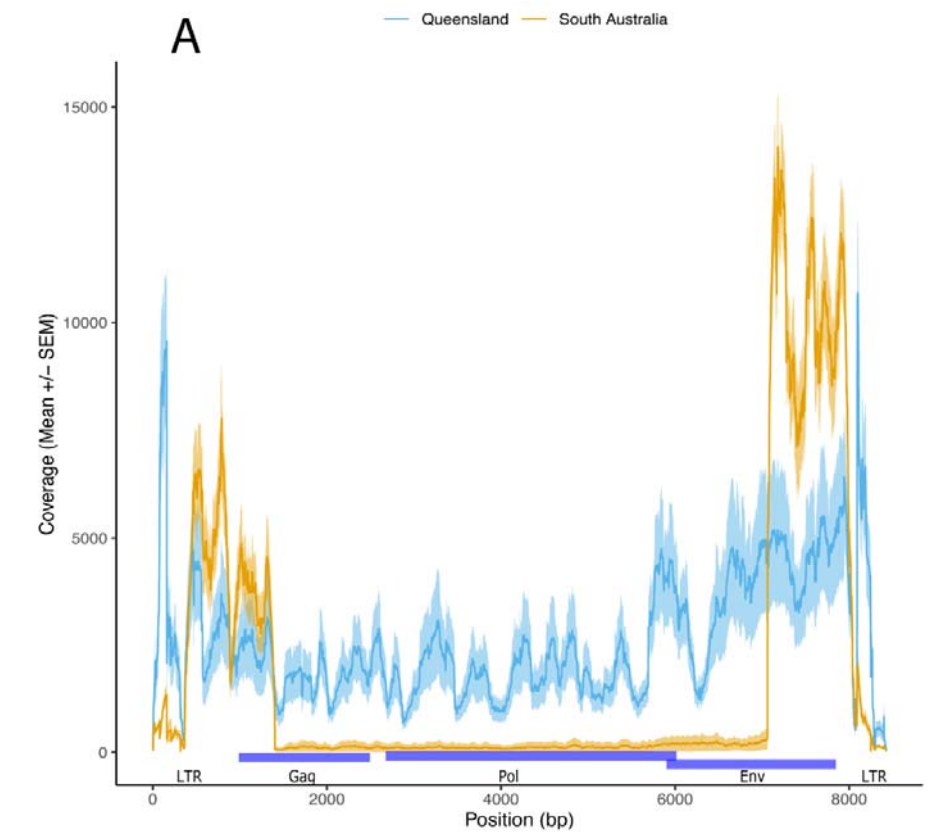
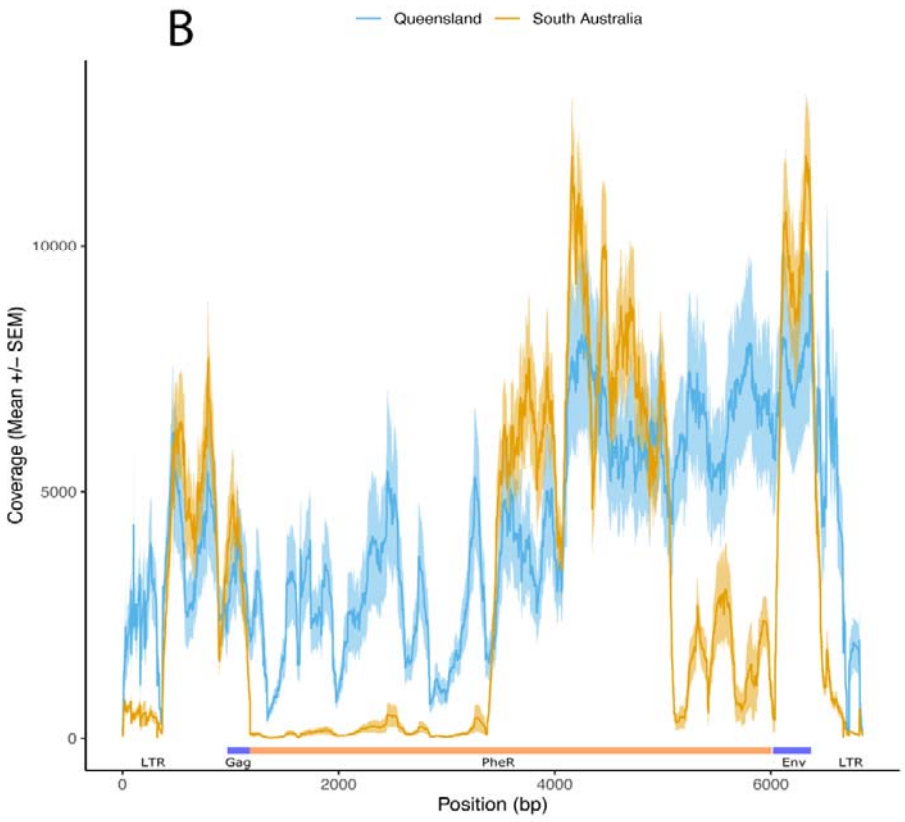


Figure 1: Cartoon of KoRVA and RecKoRV genetic sequence. KoRV LTRs are marked in green, KoRV sequences in blue, PhER sequences in orange.

Figure 2: Map of the locations of the animals sampled in this study. Mt Lofty Ranges orange drop (SA), Cape Otway blue drop, French Island purple drop, Strezlecki ranges green drop (VIC), SE QLD brown drop (QLD) (map created with Google maps)





539

Figure 3: Panel A: Coverage of reads mapped to a representative sequence of KoRV A from the koala reference genome . For each group the mean normalised coverage ((per position coverage/total coverage) x $1x10^6$) is represented by a line and +/- the standard error is shaded around the mean. QLD samples in blue, SA in orange. KoRV genomic regions are marked underneath the read maps with blue bars, these regions are: 5′ LTR, gag, pol, env, LTR for KoRV A . **Panel B:** Coverage of reads mapped to a representative sequence of RecKoRV from the koala reference genome . For each group the mean normalised coverage ((per position coverage/total coverage) x $1x10^6$) is represented by a line and +/- the standard error is shaded around the mean. QLD samples in blue, SA in orange. RecKoRV genomic regions are marked underneath the read maps with blue bars, these regions are: 5′ LTR, gag portion, PhER , env portion, 3′ LTR .

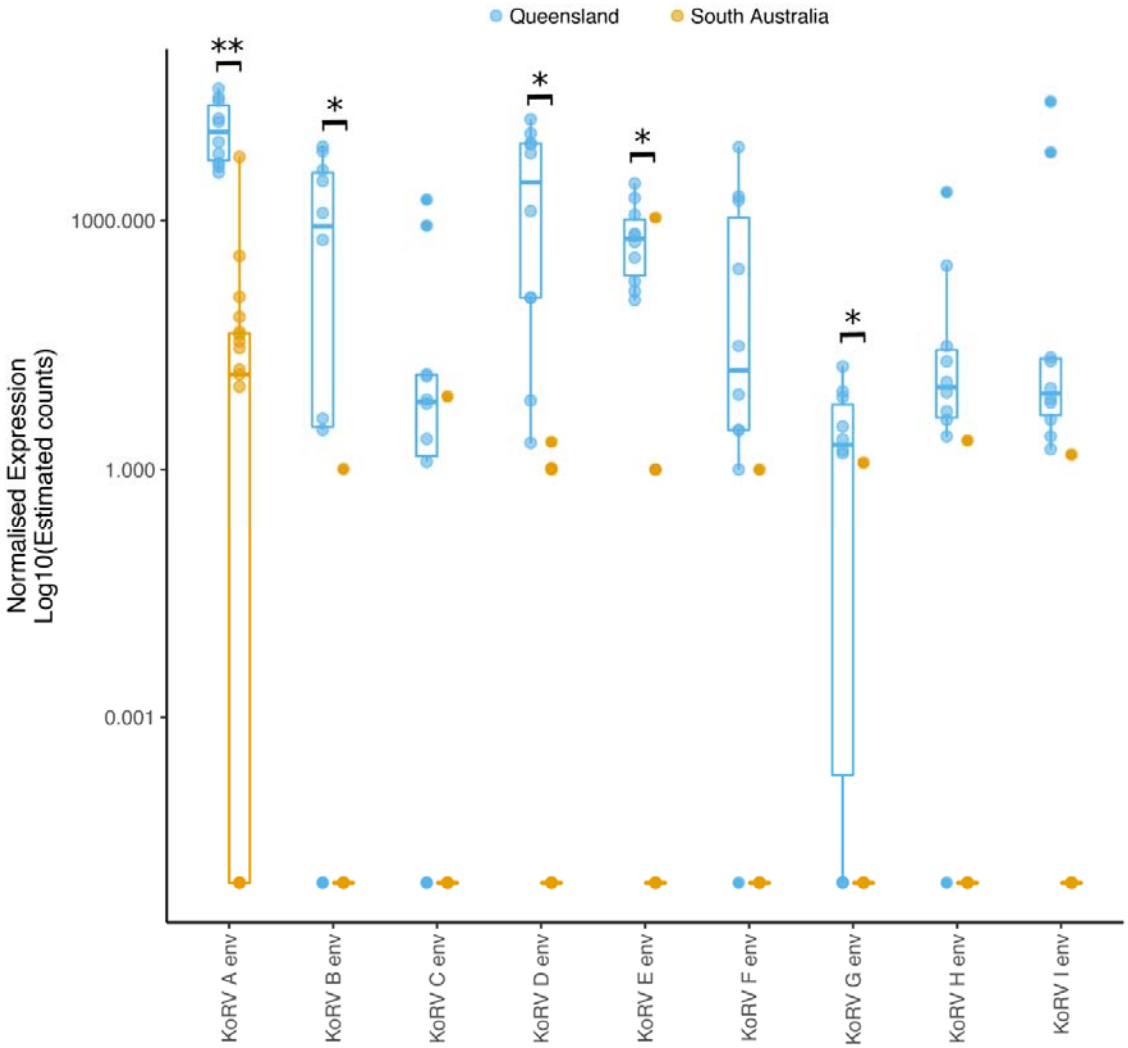


Figure 4: Normalised expression Log10(estimated counts) of KoRV A complete *env* gene and the 575 nucleotides of the hypervariable region of the envelope variants (B-I). Box and whisker plots show the median and interquartile ranges (box) and minimum/maximum expression (whiskers) of groups. Data for individual animals within a group are shown by circles. QLD animals in blue and SA animals in orange. *Env* variants with significantly different expression between QLD and SA groups marked with black bars (** = P<0.001, *= P<0.005)

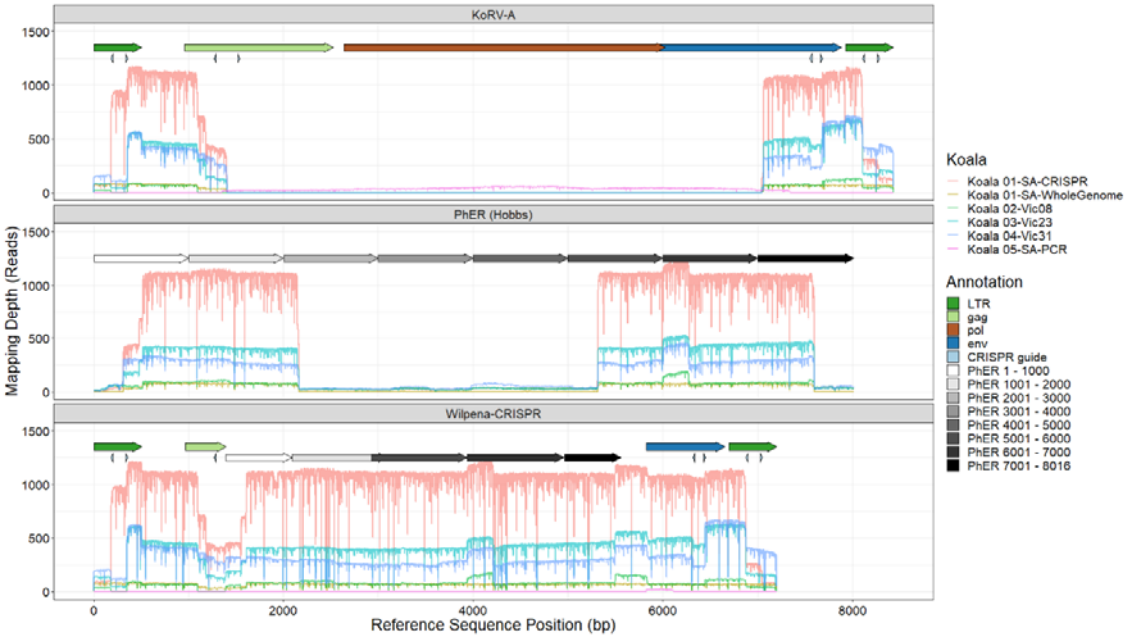


Figure 5. Coverage maps of Nanopore reads mapped to three different reference sequences (KoRV-A, PhER (Hobbs), Wilpena-CRISPR) using minimap2. Annotation arrows represent locations of coding domains from KoRV-A (in colour), CRISPR guide oligos, and genome regions of PhER in 1000 bp increments (greyscale) to highlight the insertion within the recKoRV assembly Wilpena-CRISPR. Note that the genomes do not align in the figure and base positions are relative to the reference genome in each plot.

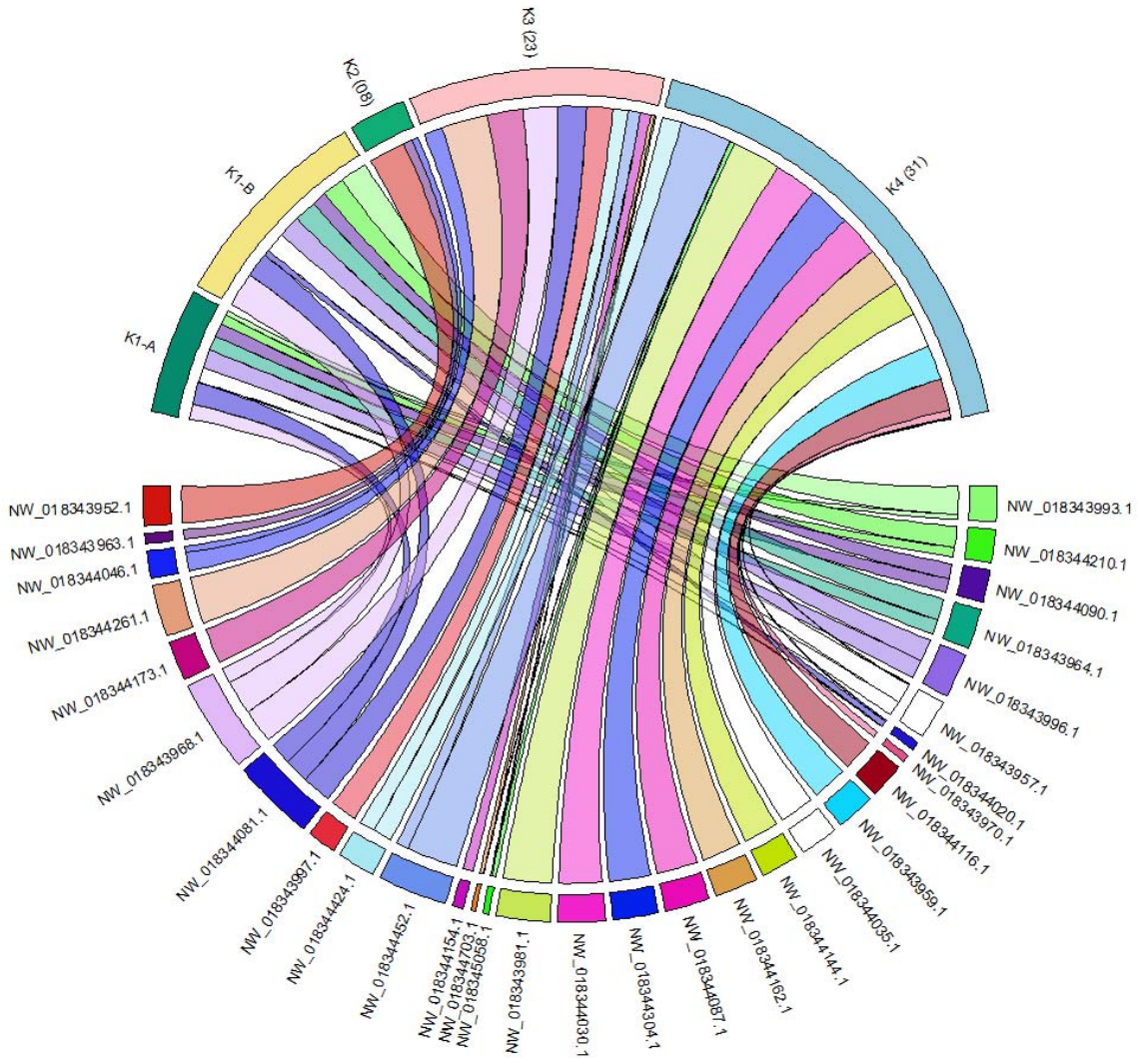


Figure 6: Circos plot of the number of reads mapping to koala reference genome contigs, highlighting the shared insert points between koalas 1 – 4.

Table 1: Details of the koalas used in this study

A: Summary of animals used in the RNA seq study

	QLD	SA
Sex		
Male	6	12
Female	4	7
Age		
Adult	10	13
Juvenile	0	6
Disease Status		
Chlamydia	9[a]	6[a]

Neoplasia	1[b]	0
Oxalate Nephrosis	0	4
Scoliosis and kyphosis	0	2
Healthy	0	5
Miscellaneous	1[c]	3[d]
Total	**10**	**19**

[a] some animals had more than one disease syndrome on post mortem
[b] osteochondroma
[c] non neoplastic hepatic mass
[d] one each of unknown, osteomyelitis secondary to trauma and non-chlamydial reproductive tract disease

B: Animals used for nanopore sequencing

Koala	Genetic population	KoRV A status	Tissue sample	Sequencing strategy	Cause of death
K01 (Wilpena)	SA (Mt Lofty ranges)	negative	Liver	Whole genome CRISPR enrichment	Oxalate nephrosis
K02 (08)	VIC (Cape Otway)	Positive	Spleen	CRISPR enrichment	Euthanased as part of population management
K03 (23)	VIC (French Island)	Negative	Spleen	CRISPR enrichment	Cystic thyroid/thymic mass
K04 (31)	VIC (Strezlecki ranges)	Negative	Spleen	CRISPR enrichment	Trauma
K05 (K15-012)	SA(Mt Lofty ranges)	Positive only for LTR and TM unit of env gene	lymph node	Long range PCR and nanopore	Dog attack

602 **Table 2:** KoRV variant expression of individual animals

ID	Location[a]	Sex	Age[b]	Provirus PCR	KoRV variants[c]
A	QLD	M	4	+	ALL
B	QLD	M	4	+	ALL
C	QLD	F	4	+	ALL
D	QLD	F	4	+	ALL
E	QLD	F	>3	+	A, C, D, E, F, G, H, I
F	QLD	M	4	+	A, D, E, I
G	QLD	M	5	+	A, B, C, D, E, F, H, I
H	QLD	M	4	+	A, B, D, E, F, G, H, I
I	QLD	F	4	+	ALL
J	QLD	M	5	+	A, B, C, D, E, F, H, I
K	SA	F	4	+	A
L	SA	M	3	+	A
M	SA	M	2	+	N[d]
N	SA	M	4	-	N
O	SA	M	3	+	N
P	SA	F	3	+	N
Q	SA	M	4	+	A
R	SA	M	2	+	A
S	SA	M	2	+	N
T	SA	M	2	+	A, D
U	SA	F	4	+	E
V	SA	M	4	+	E
W	SA	F	4	+	N
X	SA	F	3	-	A, D
Y	SA	F	4	-	A
Z	SA	M	3	+	A, B, C, D, E, F, G, I

A1	SA	F	1	+	A, E
A2	SA	M	2	-	A, D
A3	SA	M	4	-	A

[a] Population location: QLD – Queensland; SA – South Australia

[b] Age determined my dentition and the degree of wear of the upper pre-molar (Martin *et al*. 1999)

[c] KoRV variants determined by KoRV transcripts; ALL = all published variants (A to I)

[d] N= no env hypervariable region detected.

603

604 **Table 3**. Summary information of total Nanopore reads matching to the koala reference

605 genome

606

Sample	Reads mapped to KoRV	Reads mapped to koala genome	Insertion Sites	Median (range) reads mapped per site
Koala 1 – whole genome	156	152	14	10 (1 – 26)
Koala 1 – CRISPR enrichment	2488	272	16	14.5 (1 – 47)
Koala 2	275	72	3	13 (11 – 48)
Koala 3	1512	323	18	10 (1 – 63)
Koala 4	1699	609	25	5 (1 – 70)
Koala 5	156	NA	NA	NA
Total	6286	1428	56	8 (1 – 116)

607 NA – Koala 5 nanopore reads generated using long-range PCR of KoRV primers, and did

608 not overlap the koala genome

609

610 **Table 4**. Summary of insertion sites in the koala reference genome (Genbank Accession number: GCA_002099425.1) identified by

611 mapping reads with minimap2

Contig	Reads mapped	K01-WG	K01-CR	K2 (08)	K3 (23)	K4 (31)	Comment on Insert Site
NW_018343952.1	48	0	0	48	0	0	Insert within MAP2K5 gene
NW_018343957.1	49	20	28	0	0	1	Insert at different locations on contig between K1 and K4; K1 insert ~86 kb upstream of XR_002328485.1 Lnc RNA (potential interaction with VCAN gene)
NW_018343959.1	47	0	0	0	0	47	Insert within MPP4 gene
NW_018343963.1	13	0	0	13	0	0	
NW_018343964.1	50	20	30	0	0	0	~36 kb upstream of RAB3GAP2 gene
NW_018343968.1	116	21	47	0	48	0	~150 kb upstream of STX6 gene
NW_018343970.1	5	0	0	0	0	5	~60 kb upstream of LOC110207063 gene
NW_018343981.1	68	0	0	0	0	68	Insert within LOC110209428 gene
NW_018343993.1	45	7	38	0	0	0	~18 kb upstream of LOC110211657 gene
NW_018343996.1	56	26	30	0	0	0	~7 kb upstream of LOC110212362 gene
NW_018343997.1	32	0	0	0	32	0	~113 kb upstream of COMMD6
NW_018344020.1	11	1	8	0	0	2	~789 kb gap between mapping of upstream and downstream regions
NW_018344030.1	58	0	0	0	0	58	Insert within LOC110217408 gene
NW_018344035.1	54	0	0	0	0	54	Insert within LOC110217834 LncRNA
NW_018344046.1	32	0	0	11	21	0	Insert in CPA6 gene
NW_018344081.1	91	20	34	0	37	0	Insert within TSPAN5 gene
NW_018344087.1	54	0	0	0	0	54	~18 kb upstream of BLOC1S6 gene
NW_018344090.1	40	19	21	0	0	0	
NW_018344116.1	45	0	0	0	0	45	~1 kb upstream of LOC110193889 LncRNA
NW_018344144.1	49	0	0	0	0	49	~80 kb upstream of HOOK3
NW_018344154.1	15	0	0	0	15	0	Insert within PITPNM2 gene
NW_018344162.1	53	0	0	0	0	53	~60 kb upstream of TM4SF20 gene
NW_018344173.1	47	0	0	0	47	0	~17, 20, 21, & 22 kb, upstream of tRNA-GCC, tRNA-GUC, tRNA-CUC, and LOC110197942 gene (predicted to encode heat shock 70 kDa protein 6-like) respectively
NW_018344210.1	42	13	29	0	0	0	~134 kb upstream of CXXC4 gene

NW_018344261.1	63	0	0	0	63	0	Insert in PPFIBP1 gene
NW_018344304.1	58	0	0	0	0	58	Insert within DCLK1 gene
NW_018344424.1	50	0	0	0	21	29	Insert sites ~7 kb apart in different koalas
NW_018344452.1	88	0	0	0	18	70	~18 kb gap between mapping of upstream and downstream regions
NW_018344703.1	6	0	0	0	5	1	Small genome contig (~41 kb)
NW_018345058.1	10	0	0	0	4	6	Small genome contig (~31 kb)
NW_018345540.1	5	0	0	0	5	0	Small genome contig (~14 kb)
Totals	1400	147	265	72	316	600	

612

613

Supplementary Data

Supplementary Data 1: Details of the animals included in this study: N/A = sample not available for testing, region of origin, QLD= Queensland, SA= South Australia, sex, M= Male, F= female, tooth wear class (age classification on a 7 point scale [69]), KoRV proviral status (pol gene PCR on DNA from whole blood), lymph node RNA quality and NGS read details: concentration, A260/280 and A260/230 ratios, RIN value, number of paired raw reads and number of trimmed reads from each sample.

Supplementary Data 2: Table of sequence location and naming of KoRV and RecKoRV sequences from the koala reference genome [70]

Supplementary Data 3: Fast file of sequences of KoRV and RecKoRV from the koala reference genome [70]

Supplementary Data 4: Estimated counts of KoRV envelope variants for individual animals, Column A = Sequence read archive (SRA) identifier,Column B = Koala ID as per Supplementary file 1, Column C, D, E: TPM for KoRV A, *gag, pro*/*pol* and *env* (Genbank number AF151794), Column F-M = KoRV envelope variant based on the first 575 nucleotides of the envelope variants (B-I). Column N= State of origin, Column O= KoRV pol gene PCR on whole blood DNA.

Supplementary Data 5: Normalised expression Log10(estimated counts) of KoRV A complete *env* gene and the 575 nucleotides of the hypervariable region of the envelope variants (B-I) for animals from tooth wear (age) class 4. Box and whisker plots show the median and interquartile ranges (box) and minimum/maximum expression (whiskers) of groups. Data for individual animals within a group are shown by circles. QLD animals in

642 blue and SA animals in orange. Envelope variants with significantly different expression
643 between QLD and SA groups marked with black bars (** = P<0.001, *= P<0.005)
644
645 **Supplementary Data 6:** Sequence similarity alignment generated using EasyFig [71].
646 Representative assemblies from each of Koala 1, Koala 3, and Koala 4 were compared
647 using Blast, with regions with an identity of at least 75% between sequences connected
648 and coloured by identity value. The location in the koala genome for each of the four
649 assemblies is denoted by the koala reference genome contig accession number in the
650 title for each sequence. Annotated fragments of sequence regions (PhER 5′ and PhER 3′)
651 or incomplete genes (*gag*, *env*) are denoted with jagged lines at the 5′ or 3′ end of the
652 annotation. K01 SA1 NW018344210 has a deletion seen in 50% of the assembled inserts
653 further truncating the gag gene compared to the other representative RecKoRV
654 assemblies. This deletion ranged from ~400 – 500 bases, depending on the assembly.
655
656
657
658 References

659 1 Woinarski, J. & Burbidge, A. A. Phascolarctos cinereus. *The IUCN Red List of*
660 *Threatened Species 2016* **e.T16892A21960344** (2016).
661 2 Committee, E. a. C. R. (ed Senate Committee) (Senate Printing Unit,
662 Parliament House Canberra, 2011).
663 3 Polkinghorne, A., Hanger, J. & Timms, P. Recent advances in understanding the
664 biology, epidemiology and control of chlamydial infections in koalas. *Vet Microbiol*
665 **165**, 214-223, doi:10.1016/j.vetmic.2013.02.026 (2013).
666 4 Simmons, G. S. *et al.* Prevalence of koala retrovirus in geographically diverse
667 populations in Australia. *Aust Vet J* **90**, 404-409, doi:10.1111/j.1751-
668 0813.2012.00964.x (2012).
669 5 Neaves, L. E. *et al.* Phylogeography of the Koala, (Phascolarctos cinereus), and
670 Harmonising Data to Inform Conservation. *PLoS One* **11**, e0162207,
671 doi:10.1371/journal.pone.0162207 (2016).
672 6 Tarlinton, R. *et al.* Transcriptomic and Genomic Variants Between Koala
673 Populations Reveals Underlying Genetic Components to Disorders in a
674 Bottlenecked Population. *Conservartion Genetics* **In Press**, 10.1007/s10592-
675 10021-01340-10597 doi:10.1007/s10592-021-01340-7 (2021).
676 7 Robinson, A. C. in *The koala: proceedings of the Taronga symposium on koala*
677 *biology, management and medicine.* 132-143 (Zoological Parks Board, Sydney).
678 8 lindsay, H. A. Re-establishing the koala in SA. *Wild Life* **12**, 257-262 (1950).
679 9 Hayward, A., Grabherr, M. & Jern, P. Broad-scale phylogenomics provides insights
680 into retrovirus-host evolution. *Proc Natl Acad Sci U S A* **110**, 20146-20151,
681 doi:10.1073/pnas.1315419110 (2013).

10 Manghera, M. & Douville, R. N. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* **10**, 16, doi:10.1186/1742-4690-10-16 (2013).
11 Chuong, E. B., Rumi, M. A., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature genetics* **45**, 325-329, doi:10.1038/ng.2553 (2013).
12 Chen, C. P. *et al.* Functional Characterization of the Human Placental Fusogenic Membrane Protein Syncytin 2. *Biology of reproduction* **79**, 815-823 (2008).
13 Dupressoir, A. *et al.* Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc Natl Acad Sci U S A* **102**, 725-730 (2005).
14 Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221-225, doi:10.1038/nature14308 (2015).
15 Watanabe, S. *et al.* Phylogenetic and structural diversity in the feline leukemia virus env gene. *PLoS One* **8**, e61009, doi:10.1371/journal.pone.0061009 (2013).
16 Evans, L. H. *et al.* Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. *J Virol* **83**, 2429-2435 (2009).
17 Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M. & Robinson, W. F. The Nucleotide Sequence of Koala (*Phascolarctos cinereus*) Retrovirus: a Novel Type C Endogenous Virus Related to Gibbon Ape Leukemia Virus. *Journal of Virology* **74**, 4264-4272 (2000).
18 Avila-Arcos, M. C. *et al.* One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol Biol Evol* **30**, 299-304, doi:10.1093/molbev/mss223 (2013).
19 Simmons, G., Clarke, D., McKee, J., Young, P. & Meers, J. Discovery of a novel retrovirus sequence in an Australian native rodent (Melomys burtoni): a putative link between gibbon ape leukemia virus and koala retrovirus. *PLoS One* **9**, e106954, doi:10.1371/journal.pone.0106954 (2014).
20 Alfano, N. *et al.* Endogenous Gibbon Ape Leukemia Virus Identified in a Rodent (Melomys burtoni subsp.) from Wallacea (Indonesia). *J Virol* **90**, 8169-8180, doi:10.1128/jvi.00723-16 (2016).
21 Brown, K. & Tarlinton, R. Is Gibbon Ape Leukaemia Virus still a threat? . *Mammal Review* **47**, 53-61 (2017).
22 Hayward, J. A. *et al.* Infectious KoRV-related retroviruses circulating in Australian bats. *Proc Natl Acad Sci U S A* **117**, 9529-9536, doi:10.1073/pnas.1915400117 (2020).
23 McMichael, L. *et al.* A novel Australian flying-fox retrovirus shares an evolutionary ancestor with Koala, Gibbon and Melomys gamma-retroviruses. *Virus genes* **55**, 421-424, doi:10.1007/s11262-019-01653-3 (2019).
24 Tarlinton, R. E., Meers, J. & Young, P. R. Retroviral invasion of the koala genome. *Nature* **442**, 79-81, doi:10.1038/nature04841 (2006).
25 Sarker, N. *et al.* Koala retrovirus viral load and disease burden in distinct northern and southern koala populations. *Scientific reports* **10**, 263, doi:10.1038/s41598-019-56546-0 (2020).
26 Maher, I. E. & Higgins, D. P. Altered Immune Cytokine Expression Associated with KoRV B Infection and Season in Captive Koalas. *PLoS One* **11**, e0163780, doi:10.1371/journal.pone.0163780 (2016).
27 McEwan, G. *et al.* Retroviral integrations contribute to elevated host cancer rates during germline invasion. *Nature communications* **In Press** (2021).
28 Quigley, B. L. *et al.* Changes in endogenous and exogenous Koala Retrovirus (KoRV) subtype expression over time reflects koala health outcomes. *J Virol*, doi:10.1128/jvi.00849-19 (2019).
29 Waugh, C. A. *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). *Scientific reports* **7**, 134, doi:10.1038/s41598-017-00137-4 (2017).

30 Legione, A. R. *et al.* Koala retrovirus (KoRV) genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease. *Journal of medical microbiology*, doi:10.1099/jmm.0.000416 (2016).

31 Tarlinton, R., Meers, J., Hanger, J. & Young, P. Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *Journal of General Virology* **86**, 783-787, doi:10.1099/vir.0.80547-0 (2005).

32 Fabijan, J. *et al.* Prevalence and clinical significance of koala retrovirus in two South Australian koala (Phascolarctos cinereus) populations. *Journal of medical microbiology* **68**, 1072-1080, doi:10.1099/jmm.0.001009 (2019).

33 Quigley, B. L. *et al.* Koala Retrovirus in Northern Australia Shows a Mixture of Stable Endogenization and Exogenous Lineage Diversification within Fragmented Koala Populations. *J Virol*, doi:10.1128/jvi.02084-20 (2021).

34 Chappell, K. J. *et al.* Phylogenetic diversity of Koala Retrovirus within a Wild Koala Population. *J Virol*, doi:10.1128/jvi.01820-16 (2016).

35 Cui, P. *et al.* Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembled reference genome. *PeerJ* **4**, e1847, doi:10.7717/peerj.1847 (2016).

36 Sarker, N. *et al.* Genetic diversity of Koala retrovirus env gene subtypes: insights into northern and southern koala populations. *J Gen Virol* **100**, 1328-1339, doi:10.1099/jgv.0.001304 (2019).

37 Sarker, N. *et al.* Koala retrovirus viral load and disease burden in distinct northern and southern koala populations. *Scientific reports* **In press** (2019).

38 Hobbs, M. *et al.* Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. *Scientific reports* **7**, 15838, doi:10.1038/s41598-017-16171-1 (2017).

39 Quigley, B. L., Wedrowicz, F., Hogan, F. & Timms, P. Phylogenetic and geographical analysis of a retrovirus during the early stages of endogenous adaptation and exogenous spread in a new host. *Molecular ecology*, doi:10.1111/mec.15735 (2020).

40 Xu, W. *et al.* An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc Natl Acad Sci U S A* **110**, 11547-11552, doi:10.1073/pnas.1304704110 (2013).

41 Quigley, B. L., Ong, V. A., 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* **92**, doi:10.1128/jvi.01871-17 (2018).

42 Zheng, H. *et al.* Koala retrovirus diversity, transmissibility, and disease associations. *Retrovirology* **17**, 34, doi:10.1186/s12977-020-00541-1 (2020).

43 Hashem, M. A. *et al.* Transmission of Koala Retrovirus from Parent Koalas to a Joey in a Japanese Zoo. *J Virol* **94**, doi:10.1128/jvi.00019-20 (2020).

44 Xu, W., Gorman, K., Santiago, J. C., Kluska, K. & Eiden, M. V. Genetic diversity of koala retroviral envelopes. *Viruses* **7**, 1258-1270, doi:10.3390/v7031258 (2015).

45 Shimode, S., Nakagawa, S., Yoshikawa, R., Shojima, T. & Miyazawa, T. Heterogeneity of koala retrovirus isolates. *FEBS letters* **588**, 41-46, doi:10.1016/j.febslet.2013.10.046 (2014).

46 Xu, W. & Eiden, M. V. Koala Retroviruses: Evolution and Disease Dynamics. *Annual review of virology* **2**, 119-134, doi:10.1146/annurev-virology-100114-055056 (2015).

47 Shojima, T. *et al.* Identification of a novel subgroup of Koala retrovirus from Koalas in Japanese zoos. *J Virol* **87**, 9943-9948, doi:10.1128/jvi.01385-13 (2013).

48 Yu, T. *et al.* The piRNA Response to Retroviral Invasion of the Koala Genome. *Cell* **179**, 632-643.e612, doi:10.1016/j.cell.2019.09.002 (2019).

49 Lober, U. *et al.* Degradation and remobilization of endogenous retroviruses by recombination during the earliest stages of a germ-line invasion. *Proc Natl Acad Sci U S A* **115**, 8609-8614, doi:10.1073/pnas.1807598115 (2018).

50 Martin, R. & Handasyde, K. *The koala: Natural history, conservation and management*. (University of New South Wales Press, Sydney, 1999).

51 Tarlinton, R., Meers, J., Hanger, J. & Young, P. Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *Journal of General Virology* **86**, 783-787 (2005).

52 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

53 Li, H. *et al.* The Sequence alignment/map (SAM) format and SAMtools,. *Bioinformatics* **25**, 2078-2079 (2009).

54 Hobbs, M. *et al.* A transcriptome resource for the koala (Phascolarctos cinereus): insights into koala retrovirus transcription and sequence diversity. *BMC genomics* **15**, 786, doi:10.1186/1471-2164-15-786 (2014).

55 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology* **37**, 540-546, doi:10.1038/s41587-019-0072-8 (2019).

56 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).

57 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).

58 Fabijan, J. *et al.* Pathological Findings in Koala Retrovirus-positive Koalas (Phascolarctos cinereus) from Northern and Southern Australia. *J Comp Pathol* **176**, 50-66, doi:10.1016/j.jcpa.2020.02.003 (2020).

59 Shojima, T. *et al.* Construction and characterization of an infectious molecular clone of Koala retrovirus. *J Virol* **87**, 5081-5088, doi:10.1128/jvi.01584-12 (2013).

60 Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. & Belshaw, R. Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci U S A* **109**, 7385-7390, doi:10.1073/pnas.1200913109 (2012).

61 Rubin, H. The early history of tumor virology: Rous, RIF, and RAV. *Proc Natl Acad Sci U S A* **108**, 14389-14396, doi:10.1073/pnas.1108655108 (2011).

62 Nethe, M., Berkhout, B. & van der Kuyl, A. C. Retroviral superinfection resistance. *Retrovirology* **2**, 52, doi:10.1186/1742-4690-2-52 (2005).

63 Spencer, T. E., Mura, M., Gray, C. A., Griebel, P. J. & Palmarini, M. Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *J Virol* **77**, 749-753 (2003).

64 van der Kuyl, A. C. HIV infection and HERV expression: a review. *Retrovirology* **9**, 6, doi:10.1186/1742-4690-9-6 (2012).

65 Olagoke, O. *et al.* Induction of neutralizing antibody response against koala retrovirus (KoRV) and reduction in viral load in koalas following vaccination with recombinant KoRV envelope protein. *NPJ vaccines* **3**, 30, doi:10.1038/s41541-018-0066-4 (2018).

66 Olagoke, O., Quigley, B. L., Hemmatzadeh, F., Tzipori, G. & Timms, P. Therapeutic vaccination of koalas harbouring endogenous koala retrovirus (KoRV) improves antibody responses and reduces circulating viral load. *NPJ vaccines* **5**, 60, doi:10.1038/s41541-020-0210-9 (2020).

67 Fabijan, J. *et al.* Chlamydia pecorum prevalence in South Australian koala (Phascolarctos cinereus) populations: Identification and modelling of a population free from infection. *Scientific reports* **9**, 6261, doi:10.1038/s41598-019-42702-z (2019).

68 Speight, K. N. *et al.* Plasma biochemistry and urinalysis variables of koalas (Phascolarctos cinereus) with and without oxalate nephrosis. *Veterinary clinical pathology* **43**, 244-254, doi:10.1111/vcp.12145 (2014).

69 Martin, R. W. Age-specific fertility in three populations of the koala, Phascolarctos cinereus Goldfuss, in Victoria. . *Australian Wildlife Research* **8**, 275-283 (1981).

70 Johnson, R. N. *et al.* Adaptation and conservation insights from the koala genome. *Nature genetics* **50**, 1102-1111, doi:10.1038/s41588-018-0153-5 (2018).

71 Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009-1010, doi:10.1093/bioinformatics/btr039 (2011).