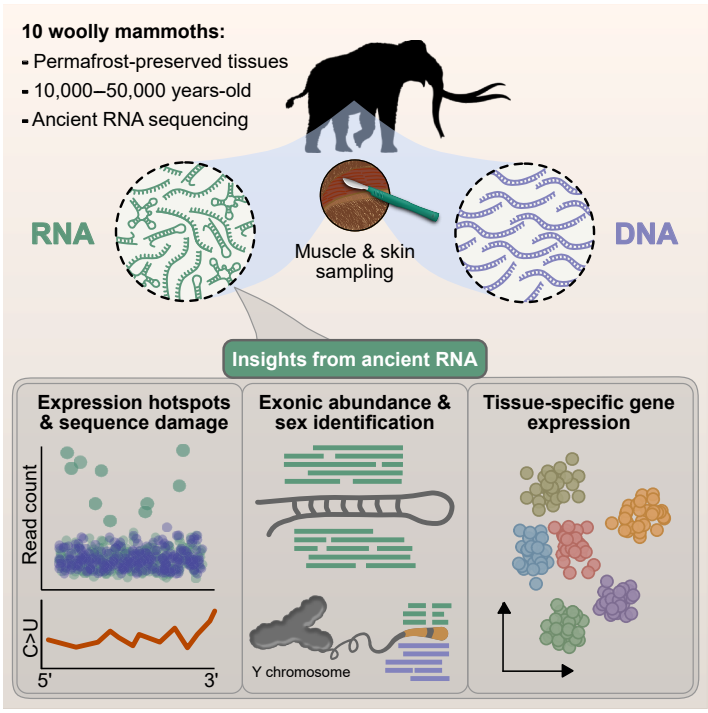
Cell

Ancient RNA expression profiles from the extinct woolly mammoth

Graphical abstract



Highlights

- Ancient RNA molecules are preserved in Late Pleistocene woolly mammoth tissues
- Transcriptional data allow the annotation of novel noncoding loci
- Tissue-specific gene expression patterns are preserved through time
- Ancient RNA expression profiles undercover missing insights from extinct paleofauna

Authors

Emilio Mármol-Sánchez, Bastian Fromm, Nikolay Oskolkov, ..., Kevin J. Peterson, Marc R. Friedländer, Love Dalén

Correspondence

emilio.marmol.sanchez@gmail.com (E.M.-S.), marc.friedlander@scilifelab.se (M.R.F.), love.dalen@zoologi.su.se (L.D.)

In brief

Ancient RNA profiles from Late Pleistocene woolly mammoths were sequenced from permafrost-preserved mummified tissues, revealing tissue-specific transcriptional functions and gene expression regulatory dynamics. These findings highlight the potential of ancient RNA molecules to provide additional biological information, thus advancing toward integrative paleogenetic research beyond the limits of DNA sequencing alone.

Mármol-Sánchez et al., 2026, Cell *189*, 52–69
January 8, 2026
https://doi.org/10.1016/j.cell.2025.10.025

CellPress

CellPress
OPEN ACCESS

Cell

Article

Ancient RNA expression profiles from the extinct woolly mammoth

Emilio Mármol-Sánchez,[1,2,3,19,*] Bastian Fromm,[4] Nikolay Oskolkov,[5] Zoé Pochon,[2,6] Marianne Dehasque,[2,7] Morteza Aslanzadeh,[1] Elif Bozlak,[2,8,9] Katherine Brown,[10] Tom van der Valk,[2,11] Panagiotis Kalogeropoulos,[1] J. Camilo Chacón-Duque,[2,12] Inna Biryukova,[1] Peter D. Heintzman,[2,13] Cecilia Furugård,[14] Valeri Plotnikov,[15] Albert Protopopov,[15] Björn Andersson,[16] Erik Ersmark,[2,11] Kevin J. Peterson,[17] Marc R. Friedländer,[1,18,*] and Love Dalén[2,11,12,18,*]

[1]Science for Life Laboratory, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, 106 91 Stockholm, Sweden
[2]Centre for Palaeogenetics, 106 91 Stockholm, Sweden
[3]Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen 1353, Denmark
[4]The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Tromsø 9006, Norway
[5]Department of Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Lund University, 222 42 Lund, Sweden
[6]Department of Archaeology and Classical Studies, Stockholm University, 11418 Stockholm, Sweden
[7]Human Evolution Program, Department of Organismal Biology, Uppsala University, 752 36 Uppsala, Sweden
[8]Department for Biological Sciences and Pathobiology, Animal Breeding and Genetics, University of Veterinary Medicine Vienna, 1210 Vienna, Austria
[9]Vienna Graduate School of Population Genetics, University of Veterinary Medicine Vienna, 1210 Vienna, Austria
[10]Division of Virology, Department of Pathology, Addenbrooke's Hospital, University of Cambridge, CB2 1QP Cambridge, UK
[11]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 104 05 Stockholm, Sweden
[12]Department of Zoology, Stockholm University, 106 91 Stockholm, Sweden
[13]Department of Geological Sciences, Stockholm University, 106 91 Stockholm, Sweden
[14]Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 10044 Stockholm, Sweden
[15]Academy of Sciences of Sakha Republic, 677007 Yakutsk, Russia
[16]Department of Cell and Molecular Biology, Karolinska Institute, 171 77 Stockholm, Sweden
[17]Department of Biological Sciences, Dartmouth College, 03755 Hanover, NH, USA
[18]Senior author
[19]Lead contact
*Correspondence: emilio.marmol.sanchez@gmail.com (E.M.-S.), marc.friedlander@scilifelab.se (M.R.F.), love.dalen@zoologi.su.se (L.D.)
https://doi.org/10.1016/j.cell.2025.10.025

SUMMARY

Ancient DNA has revolutionized the study of extinct and extant organisms that lived up to 2 million years ago, enabling the reconstruction of genomes from multiple extinct species, as well as the ecosystems where they once thrived. However, current DNA sequencing techniques alone cannot directly provide insights into tissue identity, gene expression dynamics, or transcriptional regulation, as these are encoded in the RNA fraction. Here, we report transcriptional profiles from 10 Late Pleistocene woolly mammoths. One of these, dated to be ~39,000 years old, yielded sufficient detail to recover tissue-specific regulatory mechanisms and biological functions essential for skeletal muscle metabolism, representing the oldest ancient RNA sequences recorded to date. We showcase the potential to study ancient RNA molecules beyond preconceived limitations, providing an analytical framework for validating and decoding preserved transcriptomes through time. With our findings, we anticipate the emergence of integrative paleo-studies combining genomics, proteomics, and transcriptomics.

INTRODUCTION

RNA molecules play crucial and varied roles in cell metabolism, from carrying coding information to producing functional proteins,[1] regulating gene expression,[2] or facilitating protein assembly[3] and intron splicing[4] by virtue of their catalytic properties.[5] Dynamic transcriptional changes shape and regulate cell metabolism across different developmental stages, or in response to environmental stimuli contributing to adaptation,[6] as well as fine-tuning genome organization and gene expression.[7] Such a trove of information is not directly available from genomic DNA alone and only partially captured by proteomic analyses.

Extinct species pose a particular problem in recovering transcriptional data, as living sources of fresh RNA are no longer

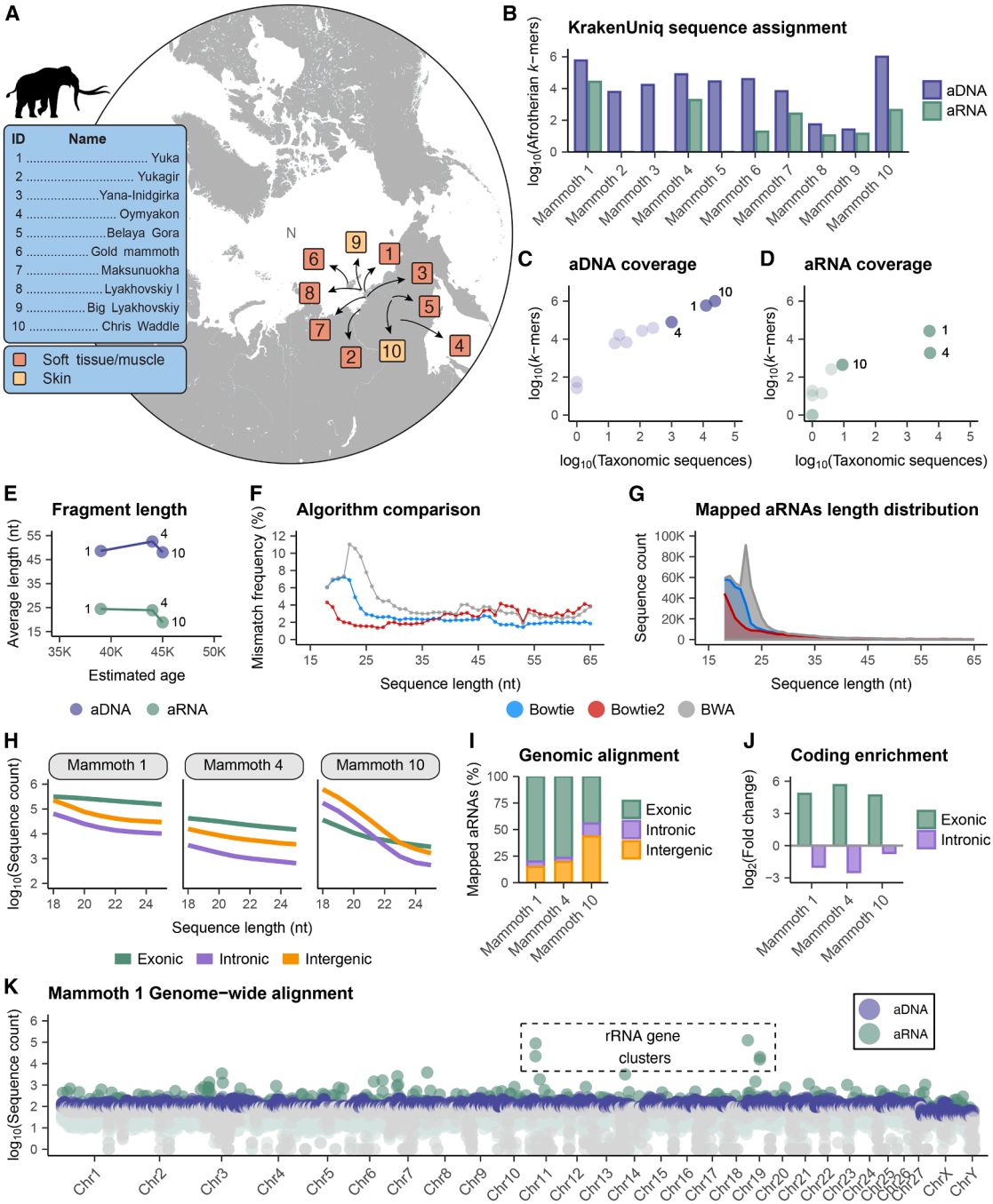


Figure 1. Assessment of the authenticity of aDNA and RNA sequences

(A) Geographical distribution and tissue identity of the woolly mammoth samples.

(B) Mammoth-like content of each tissue sample according to metatranscriptomic and metagenomic analyses based on *k*-mers (length ≥ 31) assigned to the Afrotheria superorder.

(C and D) Mammoth content quality inferred from (C) DNA and (D) RNA sequencing data based on *k*-mers assigned to the Afrotheria superorder and taxonomic sequences according to KrakenUniq.

(E) Average fragment length of sequences from mammoths 1, 4, and 10 show age-dependent shortening in aRNA data but not in aDNA data. Age is expressed in calibrated YBP (IntCal20) for mammoths 1 and 4; mammoth 10's age is taken from radiocarbon and mitochondrial phylogenetic dating.[23]

(F) aRNA alignment efficiency assessed by mismatch frequency across mapped sequences to the Asian elephant assembly (mEleMax1).

(G) aRNA alignment efficiency assessed by the number of mapped sequences and their length.

(H) Number of aRNA sequences mapped to exonic, intronic, and intergenic regions using minimum fragment length thresholds of 18–25 nt for mammoths 1, 4, and 10.

(legend continued on next page)

available. While ancient DNA (aDNA) and proteins have demonstrated to be relatively stable after an organism's death, in some cases surviving for millions of years,[8–10] RNA is considered fragile due to its rapid degradation by RNases,[11] making it unlikely to persist after death without immediate fixation. As a result, ancient RNA (aRNA) studies have remained rare, despite being at the forefront of the paleogenetic field during its early years.[12] Nevertheless, there have been successful efforts in using formalin-fixed and paraffin-embedded (FFPE) tissues for sequencing historical RNA viral genomes, such as the H1N1 "Spanish flu"[13–15] or HIV.[16,17] Besides, recent studies have demonstrated that RNA can be recovered and sequenced from both ancient[18,19] and historical animal specimens[20–22] with sufficient resolution to reveal tissue-specific gene expression patterns. Despite these groundbreaking results, the aRNA field is still in its infancy.

In this study, we sought to extend the aRNA record to the study of ancient transcriptomes in extinct paleofauna, for which the woolly mammoth serves as a renowned example. We present a methodological framework for the isolation of aRNA molecules, leveraging and expanding on previous studies using historical and ancient specimens.[18,19,22] In addition, we outline quality control measures via metagenomic and metatranscriptomic approaches, along with sequencing depth considerations, mapping strategies, and sequence analyses aimed at characterizing features found in ancient transcriptional data. These analyses account for the intrinsic mapping properties and fragmentary nature of aRNA molecules to exclude sources of DNA contamination biasing gene abundance profiling. Moreover, we propose standards for determining the transcriptional abundance, tissue identity, and endogenous origin of aRNA sequences. Applying these approaches to our mammoth samples, we identified several muscle-specific mRNAs from one of the specimens, as well as putative novel microRNA candidate loci in the woolly mammoth genome assembly[23] based on gene expression evidence using mammoth-derived aRNA sequences.

Taken together, our findings highlight the possibility to extract, sequence, analyze, and validate transcriptional profiles from ancient specimens up to around 40,000 years old, thus opening the door to generating aRNA profiles from a wide range of Pleistocene remains from both extinct and extant species.

RESULTS

aRNA molecules are preserved in woolly mammoths

We collected permafrost-preserved tissues from 10 woolly mammoths dated to the Late Pleistocene and found in northeastern Siberian paleontological fields extending from the Central Indigirka region to the mainland Oyogos Yar coast and the New Siberian Islands (Figure 1A; Table S1, #1). These regions are well known for their rich permafrost-preserved paleofauna,[24,25] which are frequently exposed through thermo-erosion and ivory collection. Such processes reveal not only ancient subfossil bones but also occasionally frozen, mummified carcasses with soft tissues still preserved, such as muscle, skin, and hair. We then extracted and sequenced both aDNA and aRNA from each sample, using methods tailored to highly degraded nucleotide fragments[22] (see STAR Methods). The RNA yield ranged from undetectable to concentrations near the lowest detection limits (Table S1, #2), while mammoth library profiles showed a broad sequence length distribution indicative of time-dependent transcript fragmentation (Figure S1).

Both aDNA and aRNA sequences were initially analyzed using metagenomic and metatranscriptomics approaches (*k*-mer length $\geq$ 31, see STAR Methods). This aimed to detect preserved mammoth-like sequences and identify any additional sources of contamination, whether modern or ancient, present in the mammoth tissues. We detected abundant afrotherian-like aDNA sequences, with mammoths 1, 4, and 10 showing significantly higher and reliable yields (Figures 1B and 1C), something also found in aRNA metatranscriptomics analyses (Figure 1D). When analyzing aRNA data to the species level (*M. primigenius*), mammoths 1 and 4 outperformed the rest (Table S1, #3). An interactive visualization of metagenomics and metatranscriptomics composition of each of the ten mammoths analyzed in this study can be found online at https://github.com/emarmolsanchez/aRNA. No evidence of mammoth-specific ancient viral RNA sequences was detected. In agreement with metagenomic and metatranscriptomics analyses, an abundant endogenous aDNA content was detected for mammoths 1, 4, and 10 (Figure 1C) after competitive mapping against the Asian elephant and human genomes (see STAR Methods). Out of the sequences mapped, a negligible source of human contamination (<0.1%) was found for these and mammoth 6 (Table S1, #4). Interestingly, mammoths 1, 4, and 10 showed a decreasing fragment length of aRNA sequences as the estimated age increased (Figure 1E).

Following these results, we considered that only mammoths 1, 4, and 10 were set to contain any reliable sources of ancient endogenous RNA molecules, and hence they are the focus of subsequent analyses.

Mammoth 1 tissues belong to an exceptionally well-preserved presumed female[26,27] juvenile specimen called Yuka, found on the mainland bluff of Oyogos Yar[28] and dated to ~39,000 calibrated years before present (YBP, GrA-53289).[28] Genomic analyses revealed that Yuka's muscle tissue nuclei may show partial activity when implanted in mouse oocytes.[29] Mammoth 4 corresponds to the Oymyakon woolly mammoth, a male calf found in the Ol'chan mine (Oymyakon, Yakutia), and dated to ~44,000 calibrated YBP (GrA-30727).[30] Mammoth 10, also named "Chris Waddle," is a female specimen that lived >44,900 years ago (OxA-38763) found near Belaya Gora, in the Sakha Republic. Phylogenetic dating based on mitochondrial sequence refined its age to ~52,000 YBP.[23] Remarkably, it has recently been reported as the first instance where three-dimensional chromatin

(I) Fraction of aRNA sequences ($\geq$23 nt) mapped to Asian elephant exonic, intronic, and intergenic regions.
(J) Fold enrichment/depletion of aRNA sequences ($\geq$23 nt) mapped to exonic and intronic regions.
(K) Genome-wide mapping of aRNA and aDNA sequences from Yuka (mammoth 1) show uneven aRNA distribution and expression hotspots absent in aDNA data. Each dot represents a non-consecutive genomic window of 0.5 Mb pairs. See Figure S2 for genome-wide mapping of mammoths 4 and 10.
See also Figure S2.

structures were reconstructed from permafrost-preserved tissues.[23] Additional data from these and other mammoths included in this study can be found in Table S1, #1.

An optimized mapping strategy for aRNA

We assessed alignment strategies for aRNA sequences using mammoth 1 (Yuka) data, given its good performance in metagenomic/metatranscriptomic analyses. While aDNA studies rely on tools optimized for short, damaged sequences[31] (e.g., Bowtie2[32] and BWA[33]), methods for aRNA alignment remain less explored, though Bowtie2[32] and Bowtie[34] have been used.[18,22] Our samples (dated to 10,000–50,000 YBP) are older than previous aRNA datasets, thus higher fragmentation and damage would be expected. Moreover, the presence of repetitive paralogous loci increases the frequency of multimapping sequences in RNA compared with DNA data, which renders strict aDNA-style filters problematic.[31] We thus compared Bowtie,[34] Bowtie2,[32] and BWA[33] tools (see STAR Methods) against the Asian elephant (mEleMax1) genome. Bowtie2 showed comparable mismatch rates (~3%) to BWA and Bowtie for sequences > 30 nucleotides (nt), but lower (~2%) for ultrashort (<30 nt) sequences (Figure 1F), though with overall fewer mapped sequences (Figure 1G). Given the <1% divergence between elephants and mammoths,[30] the excess of ultrashort mapped sequences captured by Bowtie/BWA might be false positives, which is supported by their elevated mismatch rates (Figure 1F). We therefore recommend Bowtie2 for aRNA alignment in ancient samples within the thousands-of-years-old range. Alignment statistics throughout this study are based on the use of Bowtie2 for mapping aRNA sequences (Table S2, #1; see STAR Methods).

We also tested the use of the recently reported woolly mammoth genome[23] against the modern Asian elephant (mEleMax1) as a reference assembly of choice. To do so, elephant gene annotations were first converted to genomic coordinates in the mammoth assembly (see STAR Methods; Table S2, #2, annotation available at https://github.com/emarmolsanchez/aRNA). We then mapped aRNA sequences from mammoth 1 (Yuka) to both assemblies. Mapping to the mammoth assembly resulted in slightly lower mismatches (Figure S2A) but fewer mapped sequences compared with mEleMax1 (Figure S2B), likely due to scaffold gaps present in the mammoth assembly. It is important to note, however, that the mammoth genome[23] does not represent a true *de novo* assembly but a reference-assisted reconstruction using the African elephant genome (LoxAfr3) as a template. We thus opted for the mEleMax1 assembly in downstream analyses. Nevertheless, both assemblies behaved similarly with our aRNA data (Table S2, #3), making them rather interchangeable. Likewise, an 8-fold increase in Yuka library sequencing effort only marginally increased the number of captured transcripts (Figure S2C; Table S2, #1), suggesting limited benefit from deep sequencing of aRNA data. Unless otherwise specified, subsequent analyses refer to the deep sequencing dataset.

aRNA sequences show signatures of spliced transcripts

aRNA sequences from all ten mammoths analyzed in this study were aligned to the Asian elephant genome assembly and the derived transcriptome, including coding and noncoding loci, as well as to the woolly mammoth mitochondrial DNA[35] (see STAR Methods; Table S2, #1). Mammoth 1 (Yuka) was again the specimen showing the highest number of confidently captured transcripts (breadth of coverage ≥ 5%) both at the protein-coding and noncoding fraction (Figure S2D). A threshold of at least 18 nt was set as the minimum mapped transcript length allowed (see STAR Methods). Previous reports have highlighted the advantages of including ultrashort sequences when analyzing aRNA data[18,19] (down to 15–20 nt, i.e., below 30 nt as typically used for aDNA studies). Here, 18 nt was chosen given the level of transcript fragmentation in our data (Figure 1G), while limiting excessive ultrashort random multimapping sequences, and allowing certain nucleotide loss at the sequence ends for the smallest type of transcripts considered in this study (microRNAs, typically ~22 nt long). Genome-wide alignment showed more sequences mapping to exonic than to intronic or intergenic regions for mammoths 1 and 4 (Figure 1H), as typically found for intact fresh RNA captured by RNA sequencing (RNA-seq),[36] where spliced transcripts devoid of intronic sequences map to exonic loci unevenly across the genome (Table S2, #1). This trend was consistent across all fragment lengths tested for mammoths 1 and 4 (Figure 1H). In contrast, mammoth 10 showed more sequences mapped to intergenic and intronic regions when including ultrashort sequences (Figure 1H). However, more aRNAs mapped to exonic regions when only considering sequences ≥ 23 nt, resembling what was detected in mammoths 1 and 4 (Figures 1H and 1I). The increased number of ultrashort intergenic and intronic sequences detected in mammoth 10, and by extension in mammoths 1 and 4, might stem from spurious mapping of contaminant and/or mammoth-like endogenous RNAs. This aligns with mammoth 10 being the oldest specimen of all mammoths analyzed in this study, followed by mammoths 4 and 1 (Table S1, #1), given shorter and more damaged fragments would be expected as age increases (Figure 1E). When we considered sequences with a length ≥ 23 nt, we found an exonic enrichment and intronic depletion when comparing mapping patterns between aRNA and aDNA data from the same mammoth samples (Figure 1J).

Genome-wide distribution of aRNA sequences revealed expression hotspots compatible with highly abundant transcripts, something that was not found in aDNA data (Figures 1K, S2E, and S2F). Expression hotspots included ribosomal RNA (rRNA) loci and, to a lesser extent, small nuclear RNAs (snRNAs), transfer RNAs (tRNAs), and protein-coding genes (Table S2, #1). These observations, at least for mammoth 1 (Yuka), and to a lesser extent for mammoth 4, are linked to an increased average sequence length for aRNA sequences mapped to exonic regions compared with those mapped to intronic and intergenic regions, with many exonic sequences genome-wide being well above the minimum mapped sequence length allowed (18 nt, Figure S3A). This could indicate the presence of a non-trivial number of falsely aligned aRNAs to intronic and intergenic loci genome-wide, as discussed above when comparing the reference assemblies of choice and the ratio between exonic, intronic, and intergenic mapped sequences.

As detected for exonic aRNA sequences, Yuka showed longer fragment lengths for all types of transcripts compared with mammoths 4 and 10 (Figure S3B). The longest mammoth-like

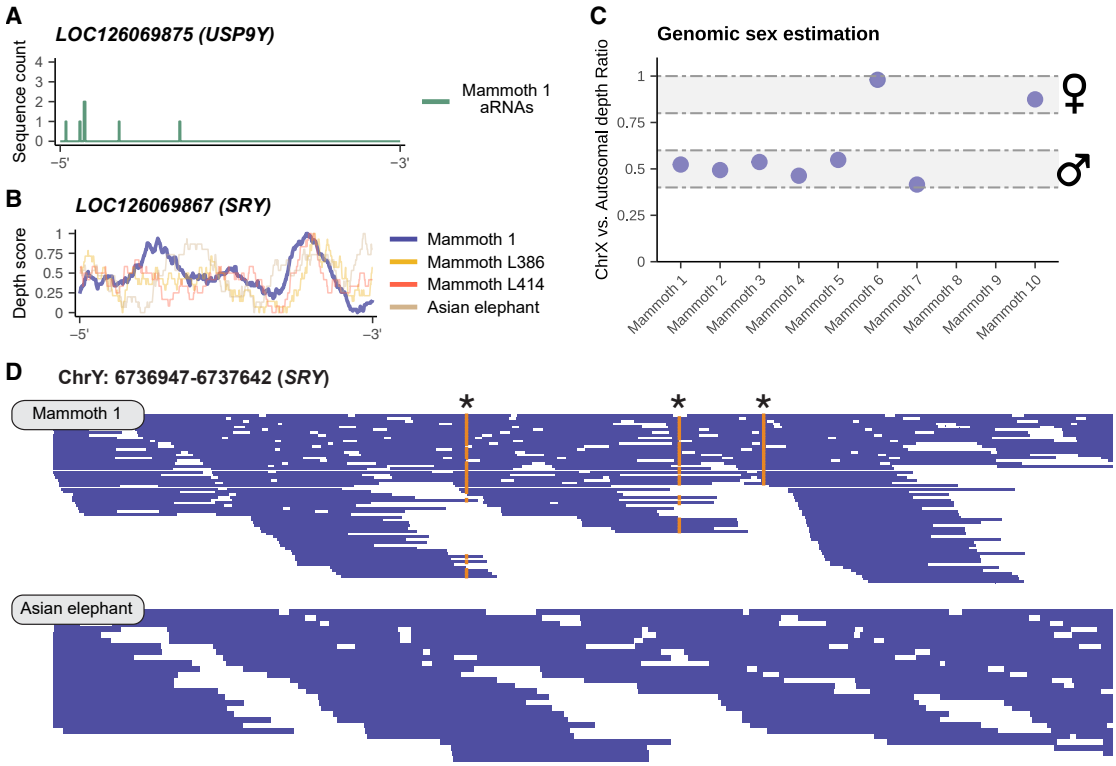


Figure 2. Sexing of mammoth 1 (Yuka) based on aRNA and DNA

(A and B) Yuka's genome shows evidence of the presence of Y-located protein-coding loci at the (A) transcriptional (*USP9Y*) and (B) genomic (*SRY*) level.

(C) Genomic sexing based on the ratio between X chromosome and autosomal depth of coverage using aDNA sequencing data. A ratio between 1 and 0.8 was considered as female (XX), while a ratio between 0.4 and 0.6 was considered as male (XY).

(D) Mammoth-specific SNVs (*) are highlighted in orange and detected in aDNA sequences mapped to the sex-determining region Y (*SRY*) gene in the Asian elephant (mEleMax1) assembly.

endogenous aRNA detected in our sequencing data corresponded to a 97 nt-long fragment mapped to the 28S rRNA (Table S2, #4, see STAR Methods). Despite highly divergent sequencing depth, the distribution and average length of mapped aRNA sequences was overall equivalent in either single-end and paired-end sequencing data from mammoth 1 (Yuka, Table S2, #4). Interestingly, sequences mapped to tRNA loci were mostly ~25–35 nt long, which is compatible with the presence of tRNA-derived small RNA fragments[37] (Figure S3B). This pattern was previously detected in historical RNA data.[22] Yuka also showed the highest number of aRNA sequences spanning exon-exon junctions (see STAR Methods; Table S2, #1). Should these represent evidence of sequences mapped to mature spliced transcripts and not mere alignment artifacts, the genes with the highest detected abundance may concentrate the majority of exon-exon junction-spanning aRNA sequences. Indeed, we found a high correlation between the number of aRNAs spanning exon-exon junctions and the total number of sequences mapped for each transcript (breadth of coverage $\geq$ 5%, Figure S3C). Mammoths 4 and 10 did not yield enough data to explore such a relationship. It is important to note, however, that the reduced number of exon-exon spanning sequences found, even for Yuka, might be explained by the overabundance of aRNA sequences mapped to rRNA transcripts (Table S2, #1), which do not harbor exon-exon junctions as they lack intronic regions.

Yuka is an XY male

Upon external examination, Maschenko et al.[26] reported Yuka as a female mammoth, noting the presence of skin folds in the genital area compatible with *labia vulvae* structures in modern elephants, and the absence of male-specific muscle structures. However, we detected aRNA sequences that mapped to genes located within the Y chromosome (Table S3, #1), including the ubiquitin-specific peptidase 9 (*LOC126069875* or *USP9Y*, Figure 2A). The presence of Y-specific coding loci was also evidenced at the aDNA level, in agreement with genomic data from other known male mammoths and one modern elephant (Figure 2B; Table S3, #2). To corroborate our findings, we conducted genotypic sexing on all the ten mammoths analyzed in this study using aDNA data generated in parallel to our aRNA profiles. Mammoths 1–5 and 7 showed half the depth of coverage for the X chromosome compared with autosomes, thus indicating the presence of only one X copy, and thereby an XY (male) genotype. Mammoths 6 and 10 had an average X depth equivalent to other autosomal chromosomes, indicative of an XX (female) genotype (Figure 2C). This agrees with previous known sexing data available (Table S1, #1), except for Yuka (mammoth 1). To further verify the sample identity of this

mammoth and rule out incidental misclassification, we independently re-analyzed all the available and previously reported Yuka's aDNA sequencing data from multiple independent tissue samples.[26,29] After genotypic sexing, all of them resulted in an XY genotype (Table S3, #3). Moreover, we detected mammoth-specific single-nucleotide variants (SNVs) in Yuka's aDNA sequences mapped to the sex-determining region Y (*SRY*) gene (Figure 2D; Table S3, #4), thus confirming their endogenous mammoth-like origin.

The mismatch between Yuka's genital phenotype and genotypical sex may result from XY gonadal dysgenesis due to altered gene expression in sex-determining loci,[38] though no genetic evidence related to altered gonadal development was found. Whether Yuka's genitalia are indeed female-like, or external sex assignment was the result of misidentification, still remains unclear. Further investigation is needed to clarify Yuka's phenotypic-genotypic sex discrepancy.

aRNA profiles show minimal DNA contamination and GC bias

The presence of expression hotspots and exon-exon spanning sequences supports the authenticity of these fragments as aRNA from mammoth tissues. However, DNA spillover during RNA extraction remains a concern, especially since no DNAse treatment was applied during laboratory work, and cDNA synthesis for PCR amplification makes it virtually impossible to distinguish DNA from RNA unless mapping patterns clearly differ. Indeed, contrary to aRNA, aDNA data should map evenly across the genome and generally be less fragmented, yielding longer sequences (Table S1, #4). Still, undetected DNA contamination could bias our results. To evaluate this, we compared aRNA and aDNA datasets (at equivalent depth of coverage) by mapping them to genomic bins in the mEleMax1 assembly (see STAR Methods). Bins with $\leq$2-fold difference in breadth of coverage and/or number of mapped aRNAs compared with aDNA data (depth of coverage) were flagged as potentially DNA-contaminated. Mammoths 1, 4, and 10 had 1.63%, 10.88%, and 0.12% of their sequences flagged, respectively (Table S3, #5). Most of these sequences had lengths similar to aRNA data (Table S1, #4), suggesting they might be genuine transcriptional fragments. Only a small fraction showed lengths $\geq$ 40 nt consistent with possible DNA spillover (Table S3, #5). Another DNA contamination indicator would be the presence of exon-intron spanning sequences, reflecting fragments mapping to genomic DNA rather than to mature intronless transcripts. We detected 788, 6, and 1,326 of such sequences in mammoths 1, 4, and 10, respectively (Table S2, #1). Few of them overlapped DNA-contaminated bins, but for those that did, their fragment length matched the expected aDNA size and they were significantly fewer in aRNA compared with aDNA datasets (Table S3, #5). These results support minimal DNA contamination in our aRNA data, possibly alongside unspliced nuclear transcripts.

PCR amplification bias against GC- and AT-rich sequences is another known issue, especially at high cycle numbers.[39,40] In our case (23 cycles; see STAR Methods), this may have led to GC distribution bias, introducing artifacts affecting transcript quantification. The GC content distribution of mapped aRNAs showed most sequences falling between 25% and 75% GC, along with a low correlation between GC content and transcript coverage (Figures S3D and S3E). Moreover, genomic regions where aRNAs were mapped had a GC content positively correlated ($r = 0.568$) with unmapped regions (Figure S3F), suggesting minimal to no influence of sequence composition bias on aRNA preservation, isolation, or amplification efficiency.

aRNA deamination patterns are similar to single-stranded aDNA

Mammoths 1, 4, and 10 showed characteristic increased DNA cytosine deamination patterns toward the 5′ sequence end (see STAR Methods; Figures 3A–3C). In agreement with DNA deamination profiles found in our mammoth data, aDNA fragments typically show biased cytosine deamination, with single-stranded overhangs more prone to damage.[41] This is attributed to the protective properties of double-stranded DNA, producing a dual-speed deamination pattern in double- vs. single-stranded fragments, which is particularly relevant in ultrashort aDNA fragments.[42] As done for aDNA, deamination profiles were estimated in aRNA sequences from mammoths 1, 4, and 10 that mapped to the Asian elephant transcriptome. Elevated damage was detected in all three of them (Figures 3D and 3F), with increased mismatch frequencies across the full length of the aRNA sequences. This aligns with damage patterns previously found in ancient/historical RNA,[18,22] and with deamination modeling for single-stranded aDNA libraries[43] equivalent to single-stranded protocols used here for aRNA sequencing (see STAR Methods).

Additionally, we applied a probabilistic model to classify aRNA sequences as ancient or modern contamination,[43] which confirmed that most mapped aRNAs from Yuka (mammoth 1) exhibit damage profiles consistent with their antiquity (Figure 3G).

aRNA sequence variation reveals mammoth genetic signatures

We retrieved 20 high-quality mammoth genomes (Table S3, #2) and identified mammoth-like SNVs relative to the Asian elephant (mEleMax1) reference assembly (see STAR Methods). This enabled detection of variant sites within aRNA sequences mapped to coding and noncoding loci. Fewer than 1% of mapped aRNA sequences contained mammoth-specific variants, with Yuka being the most representative ($\sim$45% of aRNAs with mammoth-specific SNVs mapped to exonic regions, Table S4, #1). Such low percentages are expected, as exons are under strong purifying selection compared with intronic and intergenic regions, even for synonymous mutations.[44] Deep aDNA sequencing data from Yuka[29] revealed SNVs of high (e.g., stop gained, frameshift), moderate (e.g., missense), and low (e.g., synonymous) predicted impact. A total of 407 aRNA sequences harbored mammoth-like variant sites across 73 protein-coding genes, 78 of which ($\sim$20%) matched mammoth-specific variants found in the 20 reference genomes (Table S4, #2).

We also examined SNVs in microRNA loci using Yuka genomic data. Due to their evolutionary conservation,[45] microRNA variants serve as valuable tools for species identification and validation and may affect microRNA biogenesis or target binding.[46] Of the 359 microRNA loci annotated in the Asian elephant (see STAR Methods; Table S4, #3), 60 were detected in Yuka, supported by 572 mapped aRNA sequences (Table S4, #4). Similar

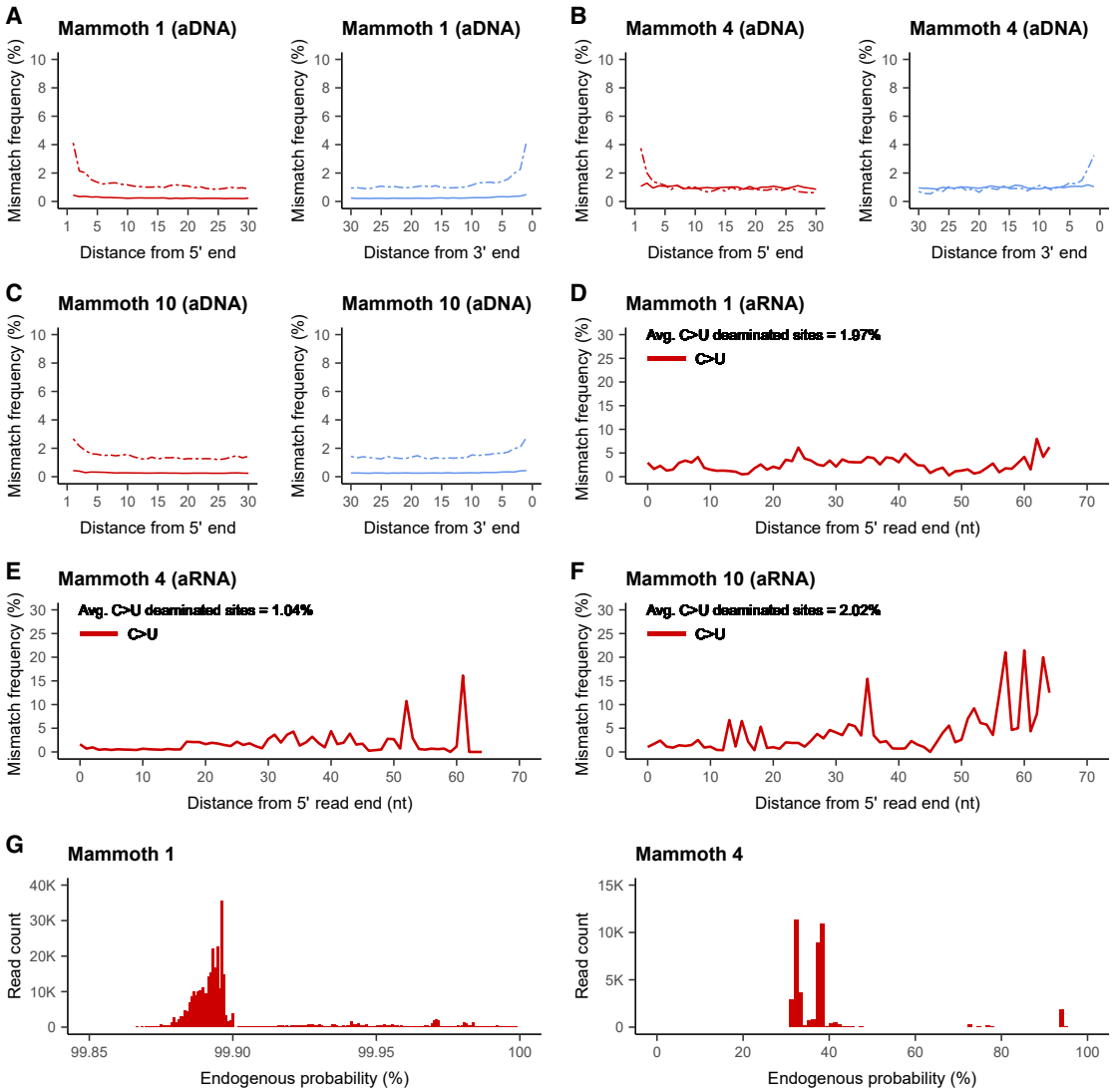


Figure 3. aDNA and RNA damage profiling

(A–C) Cytosine deamination damage analysis (C>U, read as C>T by the sequencer) based on aDNA sequencing data from (A) mammoth 1, (B) mammoth 4, and (C) mammoth 10 after competitive mapping to the Asian elephant genome assembly (mEleMax1), the woolly mammoth mitogenome (see STAR Methods), and the human genome (hg19). Deamination profiles were inferred from CpG islands to circumvent USER treatment implemented during aDNA purification (see STAR Methods).

(D–F) Cytosine deamination damage analysis (C>U, read as C>T by the sequencer) based on aRNA sequencing data from (D) mammoth 1, (E) mammoth 4, and (F) mammoth 10 after mapping the Asian elephant genome assembly (mEleMax1) and the woolly mammoth mitogenome (see STAR Methods).

(G) Probability of ancient endogenous content in mammoths 1 and 4 using aRNA sequences mapped to the Asian elephant genome assembly (mEleMax1). Mammoth 10 did not yield enough reliable information to compute endogenous probabilities.

microRNA profiles were observed using the reference-assisted 3D mammoth assembly[23] (Table S4, #5). In contrast, microRNA detection in the remaining mammoths was limited or absent (Table S4, #4).

Notably, a G variant was detected at the end of the 3′ (3p) arm of Mir-1-P3, supported by 94 of 99 mapped aRNA sequences (Table S4, #6), and also found in modern elephant blood microRNAs (Figure 4A), as well as in Mir-1 sequences of redworm (*E. fetida*), hagfish (*E. burgeri*), sea lamprey (*P. marinus*), and oikopleura (*O. dioica*) (Figure S4A). No contamination from these species was detected in our meta-analyses (https://github.com/emarmolsanchez/aRNA), so we consider that this variant likely reflects and endogenous mammoth-like origin, also supported by the ∼22 nt average length of mapped aRNA sequences and their elevated damage (Figures S4B and S4C). A cross-mammalian analysis found the G allele segregating in proboscideans (including extant African and Asian elephants and the extinct woolly mammoth), tenrecs, and the Mustela genus, while all other analyzed species harbored the putative ancestral A allele (Figure 4B). All paenungulates (proboscideans and

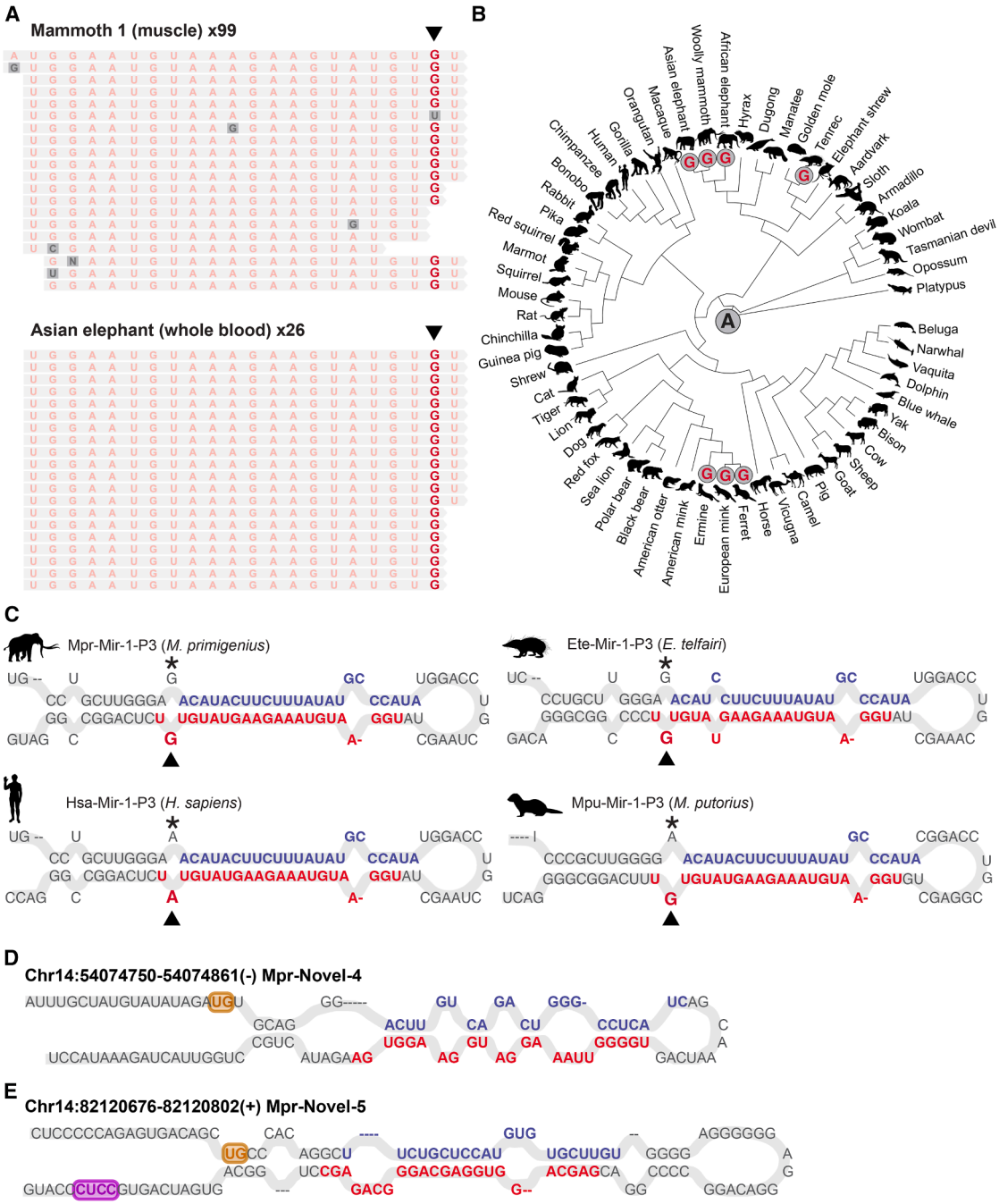


Figure 4. Mammoth-like SNV in Mir-1 sequences and novel microRNA candidates

(A) Schematic of aRNA sequences from mammoth 1 (Yuka) mapped to Mir-1-P3-3p showing a G SNV (arrowhead) at the end of the mature microRNA, also present in Mir-1-P3-3p transcripts from modern Asian elephant blood small RNA-seq data. Nucleotide mismatches are highlighted with gray shading. See Figure S4 for sequence length distribution of aRNA sequences mapped to microRNA genes.

(B) Cross-species analysis of the G variant among 60 mammalian genomes reveals its presence only in proboscideans (African and Asian elephants, and the woolly mammoth), tenrecs, and the Mustela genus.

(C) Mir-1-P3 hairpins in proboscideans and tenrecs show an A>G variant (arrowhead) near the end of the mature 3p arm (red), coupled with an additional G site (*) in the 5p arm (blue), forming a G-G bulge. This contrasts with A-A (humans) and G-A (ferrets) combinations seen in other mammals (Table S4, #7), as well as other configurations in non-mammal Mir-1 loci (see Figure S4).

(D and E) Novel microRNA candidates identified in the woolly mammoth genome from mammoth 1 (Yuka) aRNA sequences. The 5p (blue) and 3p (red) arms, along with UG (orange) and CNNC (violet) processing motifs, are highlighted in the predicted hairpin structure.

See also Figure S4.

hyracoids), sirenids, and tenrecs showed an additional G mutation in the opposite 5′ (5p) arm of Mir-1, forming a G-G unpaired bulge in proboscideans and tenrecs (Figure 4C; Table S4, #7). This bulge was absent in other mammals, where A-A or G-A combinations predominated (Figure 4C; Table S4, #7). Whether these variants affect Mir-1-P3 expression, maturation, or post-transcriptional function remains unclear, though expression changes linked to microRNA variants have been previously reported.[47]

Novel microRNA loci predicted with aRNA data

One practical advantage of RNA-seq data is the ability to discover novel loci by identifying expression evidence in genomic regions theoretically devoid of coding sequences. Here, we conducted a comprehensive annotation of microRNA genes in the Asian elephant (mEleMax1) and woolly mammoth[23] assemblies, identifying 359 loci (Table S4, #3). However, additional undetected microRNAs likely remain. To expand the mammoth microRNA complement, we analyzed Yuka's aRNA data, our best-performing sample in the microRNA fraction (Table S4, #4). Using miRDeep2,[48] we identified two promising novel microRNA candidates (Table S5, #1; Data S1).

The first, termed Mpr-Novel-4, showed aRNA sequences mapped to both arms of a predicted RNA hairpin. After UMI-deduplication, two sequences aligned to each arm, consistent with known microRNA biogenesis dynamics[49] (Data S1). Cytosine deamination in the 5p arm supports the antiquity and damage profile of these sequences. A homology search across 60 mammalian genomes identified the locus in all but marsupials and sloths (Table S5, #2), overlapping the 3′ UTR of the hexokinase 2 (*HK2*) gene. However, only modern elephants and mammoths had sequences fully matching the mapped aRNAs. These, along with other afrotherians and armadillos, showed sequences compatible with a microRNA precursor structure, including a UG processing motif upstream of the predicted 5p arm. Folding prediction resembled that of typical microRNA hairpins[50] (Figure 4D). Other mammals had homologous sequences too divergent or truncated, likely impairing functional microRNA expression (Table S5, #2). The second candidate, Mpr-Novel-5, was only found in elephantids (Table S5, #2), spanning most of the 15th intron of the glycerol-3-phosphate acyltransferase 2 (*GPAT2*) gene. It may represent a proboscidean-specific mirtron,[51] as it was absent in sirenids and hyracoids. Only two aRNA sequences mapped to its predicted 3p arm (Data S1). As with Mpr-Novel-4, the extended sequence formed a hairpin structure and included UG and CNNC motifs typical of bona fide microRNA loci (Figure 4E; Table S5, #2). Neither candidate was expressed in modern elephant microRNA sequencing data from blood extracts.[52]

Although promising, these sequences remain hypothetical novel microRNA loci. Experimental validation is required to confirm their ability to generate mature microRNAs, identify biological targets, and assess their regulatory roles.

aRNA sequences reflect a muscle-specific transcriptome

We analyzed the transcriptional abundance in mammoth tissues by aligning aRNA sequences to the longest annotated protein-coding and noncoding transcripts per each gene in the Asian elephant (mEleMax1) assembly. As before, Yuka (mammoth 1) outperformed other specimens in both the number and quality of confidently identified protein-coding transcripts (Tables S2, #1 and S5, #3). A total of 342 protein-coding mRNAs and 902 noncoding RNAs were detected with breadth of coverage $\geq$ 5% in Yuka's aRNA sequencing data (Table S2, #1), with some of the most abundant transcripts showing significantly enriched functions essential for skeletal muscle metabolism (Figure 5A; Table S5, #3 and #4). Among the most abundant and reliably captured protein-coding mRNAs, to mention a few, are: titin (*TTN*) and obscuring (*OBSCN*), responsible for the sarcomere and sarcoplasmic reticulum structure integrity and Ca^{2+} metabolism[53,54]; nebulin (*NEB*), which regulates thin filament length during sarcomere assembly,[55] and kelch-like protein 40 (*KLHL40*), a striated muscle-specific protein that binds to and stabilizes NEB[56]; filamin-c (*FLNC*) and nebulin-related anchoring protein (*NRAP*), both key elements for muscle fiber development[57,58]; microtubule-actin cross-linking factor 1 (*MACF1*), a regulator of myonuclei positioning and maintenance[59]; xin-α (*XIRP1*) and xin-β (*XIRP2*), responsible for the reorganization of actin filaments under mechanical strain[60]; ankyrin 1 (*ANKRD1*), a transcription factor involved in muscle remodeling under stress[61]; actin α (*ACTA1*) and myosin heavy chains α and β (*MYH6* and *MYH7*), which form the core of the thin and thick sarcomere filaments, respectively[62]; and troponin T1 and C1 (*TNNT1* and *TNNC1*), both predominantly expressed in slow-twitch muscle fibers along with *MYH6* and *MYH7*. These are essential components of the troponin complex attached to tropomyosin,[63,64] whose β isoform (*TPM2*) is also detected in Yuka's skeletal muscle, jointly with an abundant range of protein-coding mitochondrial mRNAs (Figure 5A). The presence of *MYH7*, *TNNT1*, and *TNNC1*, as well as *TPM2* transcripts among the most abundant and reliably detected protein-coding mRNAs is indicative of an enriched presence of slow-twitch skeletal muscle fibers within the mammoth tissue sample used in this study.[65]

With regard to the noncoding fraction of the transcriptome, excluding microRNAs, highly abundant rRNAs and tRNA transcripts were detected, jointly with U1 and U2 spliceosomal snRNAs, which are components of the spliceosome complex responsible for mRNA maturation[66] (Figure 5B; Table S5, #5).

Some of these protein-coding and noncoding loci showed muscle-specific expression patterns when compared with modern human tissues (Figures 5C and 5D). Indeed, projection of mammoth transcript abundance onto a human multi-tissue gene expression atlas (Table S5, #6) showed that Yuka's muscle tissue profiles clustered in agreement with skeletal muscle samples when considering both protein-coding and microRNA genes (Figures 6A and 6B).

Similar results were also detected when using modern elephant RNA expression profiles, as well as with previously reported aRNA data from an end-Pleistocene (Tumat) canid,[18] historical wolf skins, and muscle and skin tissues from a Tasmanian tiger[22] (Figures S5A and S5B; Table S5, #6). However, mammoths 4 and 10 did not exhibit profiles matching their known tissue origin (Figure S5C). Further analysis of the most abundant protein-coding loci found in Yuka compared with human expression profiles confirmed concordant clustering patterns, something that was also reproduced in the microRNA

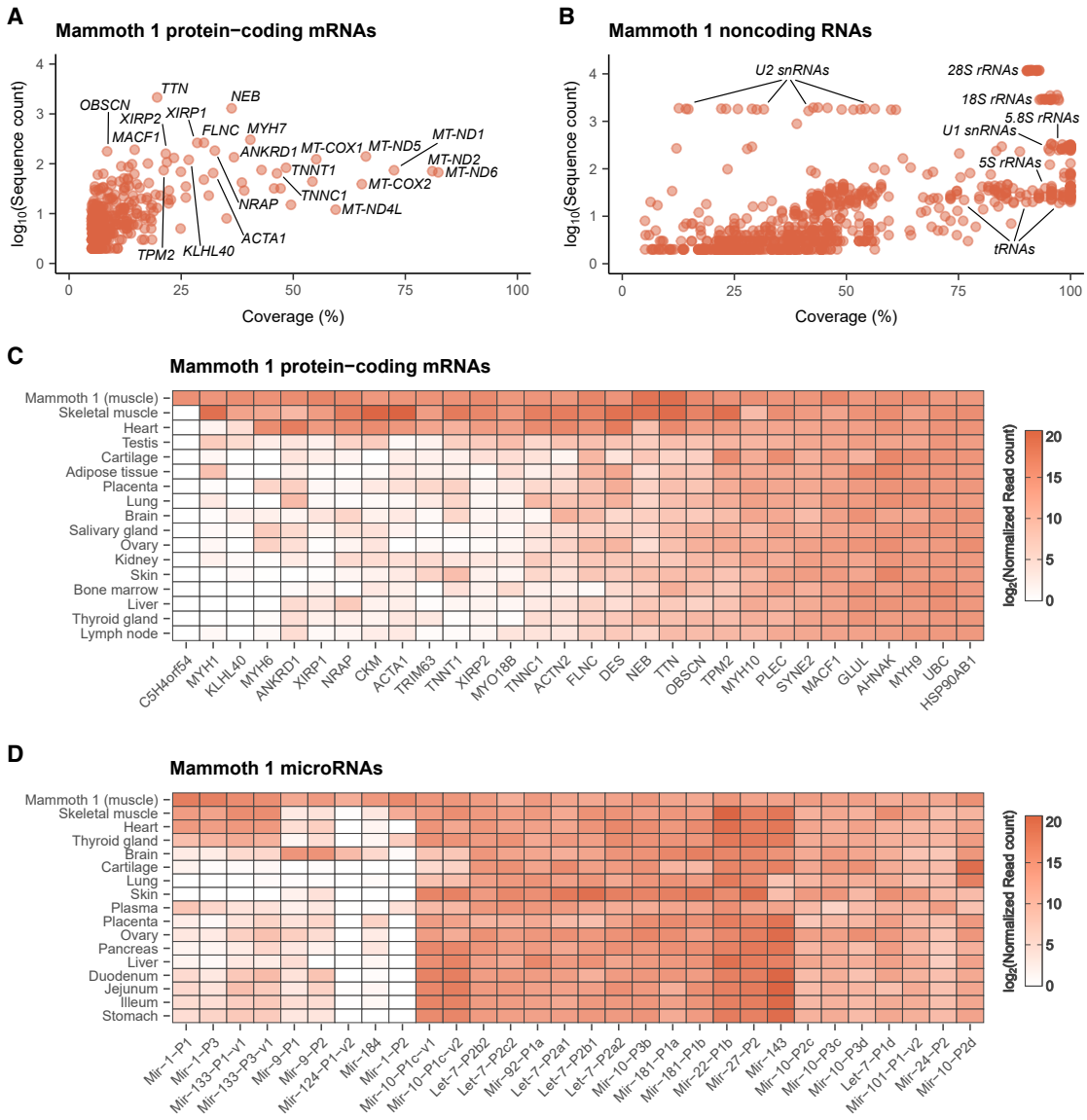


Figure 5. Transcriptional profiles from Yuka show tissue-specific skeletal muscle metabolic functions

(A and B) Transcript abundance and breadth of coverage of (A) protein-coding and (B) noncoding loci supported by mapped aRNA sequences.

(C and D) Heatmap depicting the top-30 most abundant (C) protein-coding and (D) microRNA loci in the muscle tissue of Yuka (mammoth 1) compared with modern extant expression profiles in multiple human tissues. Abundance estimates are normalized across all considered samples with the trimmed mean of M values method (TMM) and $\log_2$ transformed (see STAR Methods).

fraction (Figure S6). Mir-1 and Mir-133 transcripts were highly abundant in Yuka's skeletal muscle. These microRNA families are well known for their muscle-specific regulatory functions[67] and were also found in historical muscle expression profiles of the extinct Tasmanian tiger.[22] Three microRNAs (Mir-124-3p, Mir-184-5p, and Mir-9-5p) found in mammoth 1 (Yuka) have all been identified as regulators of cell phagocytosis[68] and cell proliferation.[69,70] Although they were detected in our aRNA data, thus suggesting their relative abundance in mammoth skeletal muscle, they seem to be particularly lowly expressed or absent in muscle tissue from modern humans but abundant in brain (Figure 5D). In contrast, microRNAs known to be essential for hepatocyte metabolism, such as Mir-122,[71] were absent in mammoth muscle but present in ancient liver from an end-Pleistocene canid[19] (Figure S6). This exemplifies the muscle-specific nature of the microRNAs detected in Yuka's muscle tissue.

DISCUSSION

This study presents evidence for the preservation of RNA molecules in permafrost-preserved woolly mammoth skeletal muscle and skin tissues beyond preconceived time limitations, with relevant insights gained from the Yuka specimen. This is, to our knowledge, the oldest preserved transcriptional signatures

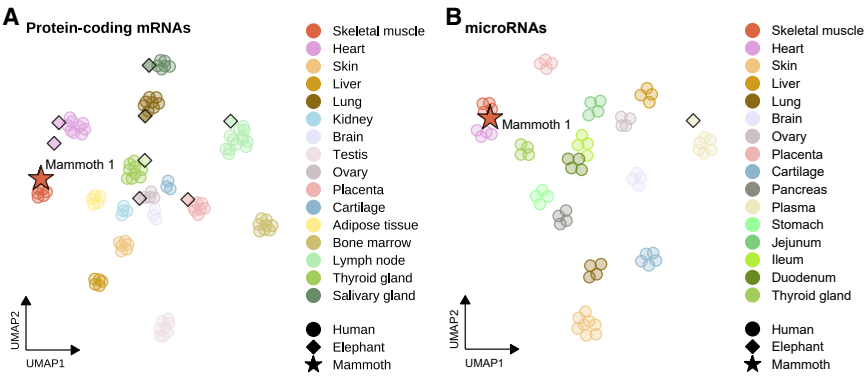


Figure 6. Clustering of modern and ancient tissues shows concordant tissue-specific gene expression patterns

Tissue-specific clustering of (A) protein-coding and (B) microRNA transcriptional profiles from Yuka's skeletal muscle (mammoth 1). A uniform manifold approximation and projection (UMAP) approach was used for clustering a reference human tissue atlas. aRNA abundance profiles were then projected onto the reference embedding. See Figure S5 for UMAP clustering of other historical and aRNA sequencing data, as well as Figure S6 for gene-specific UMAP clustering of Yuka's skeletal muscle expression profiles. See also Figures S5 and S6.

reported so far, more than twice as old as the RNA fragments previously recovered from an end-Pleistocene canid.[18,19] We outline a complement of quality checks and validation techniques to allow a fair comparison between aRNA and aDNA sequencing data obtained from the same specimens, as well as to comprehensively characterize ancient transcriptomes and predict novel loci in extinct species, albeit with limited and fragmented resolution. We also provide a number of recommendations to properly isolate, sequence, filter, align, and quantify ancient transcriptomes, as well as to identify possible sources of DNA contamination and sequence bias that we hope might guide researchers further developing the aRNA field. Metatranscriptomic and metagenomic analyses showed an overall variable endogenous nucleotide preservation across samples, revealing how DNA and RNA can be jointly present in ancient remains, even if partially masked by the presence of pre- and postmortem microbial contamination. Endogenous ribosomal RNAs were highly abundant, as expected for eukaryote cell transcriptomes.[72] Other noncoding fractions like snRNAs and tRNAs were also prevalent. This can be considered as indirect evidence of their abundant expression right before cell death. The presence of aRNA sequences mapped to intronic regions from unspliced nuclear transcripts prior to intron removal, as well as the overall increased proportion of snRNAs, which intervene in mRNA splicing in the nucleus[4] and are not expected in such a high concentration,[72] are all suggestive of differential transcript preservation dependent of cell compartmentalization. In this way, the intricate nuclear chromatin architecture, including scaffolding proteins dedicated to genome maintenance,[73] might have conditioned an increased protection of nuclear elements within, as opposed to the more exposed cytoplasm. Recent findings on the preservation of tridimensional chromatin structures through time in woolly mammoth ancient nuclei[23] support this hypothesis. Overly abundant mitochondrial mRNAs might also conform to these dynamics, despite the high mitochondrial copy number expected in skeletal muscle tissue.[74]

We have also confidently identified fragmentary but reliable evidence of over 300 protein-coding mRNAs and around 60 different microRNAs in woolly mammoth skeletal muscle. Many of these have relevant tissue-specific functions in sarcomere architecture, myofibrillar contraction, and cell development,[54–57,60–63,67] mirroring muscle metabolism and gene expression dynamics in modern extant species. We also gathered enough data to infer a predominance of slow-twitch muscle fibers in the mammoth tissue from Yuka, as revealed by the selective abundance of *MYH7*, *TNNT1*, and *TNNC1* and *TPM2* mRNA transcripts. In this way, we might have glimpsed the final pulses of transcriptional tissue-specific activity from the extinct woolly mammoth.

Besides, we resolved that Yuka was a male XY mammoth using RNA and DNA sequencing evidence, despite previous conflicting sex assignment based on external anatomical examination. Nucleotide variation and time-dependent sequence damage provided additional support to such findings.

Albeit promising, our results define pertinent questions yet to be solved: is aRNA going to become an established broad research field or will it be restricted to scarce exceptional instances involving desiccated and/or chemically fixed specimens and permafrost-preserved samples? Have we reached the limits of detectable transcriptomes through time, or can we expect that older remains harbor yet-to-be-found RNA complements? What additional insights can RNA provide that DNA or proteins fall short to attain? Methodological developments to improve and refine RNA isolation and sequencing from historical/ancient remains are probably needed. In this regard, highly fragmented RNA transcripts might present with damaged blunt ends, thus hampering proper enzymatic reactions during library preparation, something that we did not account for and that warrants further research. A comprehensive modeling of RNA decay through time and across tissues is also lacking, given the vast variability of preservation conditions in museum collections. Unlike DNA, which is a static blueprint virtually common to all healthy cells within a given organism, RNA reflects dynamic transcriptional responses in a tissue-specific manner and depicts the universe of noncoding regulatory elements that do not translate to functional proteins. We also note that highly abundant RNA families such as ribosomal RNAs detected in this study, if present, could be used as hallmarks for species identification when endogenous DNA might be scarce and/or of limited use, since these may concentrate in old remains at several orders of magnitude greater than DNA molecules. Moreover, short-lived transcriptional changes to developmental and environmental conditions might be fixed through time, thus allowing the study

of gene regulation in the cell metabolism beyond DNA sequence analysis.

In this way, our aRNA data provide biological insights into gene regulatory pathways and expression dynamics that occurred in living woolly mammoth cells several dozen millennia ago, not only through the detection of muscle-specific transcriptional profiles but also by identifying putative novel small RNA regulatory elements expressed in mammoth skeletal muscle.

Altogether, our results highlight the potential of aRNA molecules to persist across deep timescales, revealing once-forgotten layers of biological information complementing aDNA and protein sequencing approaches.

Limitations of the study

In this study, we only analyzed permafrost-preserved mummified soft tissues (skeletal muscle and skin) from woolly mammoths. Additional instances of cold-preserved metazoan RNAs come from soft, dry tissues of a 14,300-year-old end-Pleistocene canid,[18] and from a 5,300-year-old cold-mummified human.[20] Although still sparse, our results and the few others previously available highlight the potential for future successful aRNA studies. Alternative taphonomic conditions preserving RNA molecules have mainly been demonstrated in more recent historical samples, such as a medieval human pelvic bone,[21] dry wolf skins,[18] as well as in muscle and skin tissues from a desiccated Tasmanian tiger stored at room temperature for over a century.[22] This suggests that RNA fragments can survive under specific dry and/or cold conditions where enzymatic degradation is halted, offering potential for broader aRNA research. Similarly, the survival of soft tissues in ancient specimens may indicate exceptional biomolecular preservation, supporting previous choices to prioritize mummified tissues over bones.[18,19] Nevertheless, we acknowledge that most ancient remains found in paleo-archaeological fields are subfossil bones or teeth, so it is important to assess whether current aRNA methods can be extended to such mineral-rich tissues. As mentioned above, there seems to be only one instance of successful RNA extraction from a medieval bone so far.[21] However, aDNA methods have been validated and extensively applied in both soft and hard tissues,[75–78] suggesting that with appropriate adjustments in sampling, digestion, and purification methods, similar performance may be achievable if RNA molecules are still preserved. As such, we acknowledge that our RNA isolation and sequencing methods were not fully optimized, and improvements in tissue homogenization, transcript end-repair, or alternative library synthesis approaches may enhance RNA capture, as undetected losses likely occurred due to suboptimal chemistry. Likewise, broader, multi-tissue studies with a more comprehensive sampling strategy are needed to demonstrate the future feasibility and value of the aRNA research field.

DNA contamination traces present in our sequencing data may reflect inefficiencies during RNA isolation, but, given the low RNA yield and minor DNA spillover, we caution against aggressive DNA removal that might compromise RNA retention. The high number of PCR cycles used in our study may also have introduced bias toward certain transcript families. Nevertheless, we note that we did not aim to accurately quantify gene expression or perform any comparison across samples. Instead, we examined relative transcript abundances within each specimen to assess metabolic function and tissue identity, thus validating the endogenous origin of the aRNA profiles.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Emilio Mármol-Sánchez (emilio.marmol.sanchez@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All raw sequence aDNA and aRNA data for samples sequenced in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession numbers BioProject: PRJNA1190340 and PRJNA1256124.
- All code and files needed to reproduce the figures that form part of this work, as well as an integrated pipeline to preprocess, align, and quantify aRNA expression profiles, have been uploaded to GitHub repositories (https://github.com/emarmolsanchez/aRNA).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We acknowledge the support from the National Genomics Infrastructure, funded by the Swedish Research Council and the Uppsala Multidisciplinary Center for Advanced Computational Science, for assistance with massive parallel sequencing and from the UPPMAX and HPC2N computational infrastructure. E.M.-S. acknowledges financial support from the Villum Fonden (Villum Experiment project no. 57875). E.M.-S., L.D., and M.R.F. acknowledge funding from the Strategic Research Area (SFO) program of the Swedish Research Council (VR) through Stockholm University. L.D. acknowledges support from the Swedish Research Council (2021-00625) and the European Union (ERC, PrimiGenomes, 101054984). M.R.F. acknowledges funding from the Swedish Research Council (2022-03953). B.F. acknowledges funding through the Tromsø forskningsstiftelse grant (TFS) (20_SG_BF 'MIRevolution'). N.O. is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. E.B. is financially supported by Austrian Science Funds (FWF, W1225) and European Molecular Biology Organization (EMBO) Scientific Exchange Grant (9587). T.v.d.V. acknowledges support from the SciLifeLab and Wallenberg Data Driven Life Science Program (KAW 2020.0239) and the European Union (ERC, GENOMICIDE - 101160122). J.C.C.-D. acknowledges funding from the European Union's Horizon Europe program under the Marie Skłodowska-Curie Actions Fellowship 101111414. We thank Prof. Tom Gilbert for his insightful discussion on the content of this study.

AUTHOR CONTRIBUTIONS

L.D., M.R.F., and B.F. conceived the project. L.D., M.R.F., B.F., and E.M.-S. designed the study. V.P. and A.P. collected the mammoth tissues. E.M.-S. performed the laboratory methods with help from I.B., M.D., and E.E. B.F. contributed to microRNA analyses. N.O. and Z.P. analyzed metagenomic and metatranscriptomic data. M.A. contributed to data analysis and visualization. M.D. contributed to ancient DNA analyses. E.B. analyzed Y chromosome data. K.B. performed viral RNA discovery. T.v.d.V. performed variant effect analyses. P.K. contributed to clustering analyses. J.C.C.-D. performed age calibration analyses. P.D.H. contributed to sexing analyses. B.A. contributed to viral RNA discovery. B.F., C.F., and K.J.P. contributed to microRNA annotation. E.M.-S. analyzed the data unless specified otherwise. E.M.-S.,

M.R.F., and L.D. wrote the manuscript with input from co-authors. All co-authors read and approved the manuscript.

DECLARATION OF INTERESTS

L.D. is a member of the scientific advisory board at Colossal Biosciences.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Woolly mammoths
  - Modern elephants
  - Tasmanian tiger
  - Canid
  - Humans
- METHOD DETAILS
  - Sample collection and dating
  - Laboratory methods
  - Raw data preprocessing
  - Metagenomics and metatranscriptomics
  - Identification of RNA viruses
  - Mapping
  - Exon-exon and exon-intron junction overlaps
  - Exonic enrichment
  - Genome-wide expression hotspots
  - DNA contamination and amplification bias
  - Sexing
  - Damage analyses
  - Genome-wide variant calling
  - Variant effect prediction
  - MicroRNA annotation
  - MicroRNA variant analysis
  - Novel microRNA prediction
  - Transcript quantification
  - GO term enrichment
  - Tissue clustering
  - Annotation of the woolly mammoth genome
  - Data visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.10.025.

Received: April 29, 2025
Revised: August 25, 2025
Accepted: October 27, 2025
Published: November 14, 2025

REFERENCES

1. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. *21*, 630–644. https://doi.org/10.1038/s41576-020-0258-4.
2. Shang, R., Lee, S., Senavirathne, G., and Lai, E.C. (2023). microRNAs in action: biogenesis, function and regulation. Nat. Rev. Genet. *24*, 816–833. https://doi.org/10.1038/s41576-023-00611-y.
3. Hori, Y., Engel, C., and Kobayashi, T. (2023). Regulation of ribosomal RNA gene copy number, transcription and nucleolus organization in eukaryotes. Nat. Rev. Mol. Cell Biol. *24*, 414–429. https://doi.org/10.1038/s41580-022-00573-9.
4. Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat. Rev. Mol. Cell Biol. *18*, 655–670. https://doi.org/10.1038/nrm.2017.86.
5. Alonso, D., and Mondragón, A. (2021). Mechanisms of catalytic RNA molecules. Biochem. Soc. Trans. *49*, 1529–1535. https://doi.org/10.1042/BST20200465.
6. López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat. Rev. Genet. *9*, 583–593. https://doi.org/10.1038/nrg2398.
7. Henninger, J.E., and Young, R.A. (2024). An RNA-centric view of transcription and genome organization. Mol. Cell *84*, 3627–3643. https://doi.org/10.1016/j.molcel.2024.08.021.
8. Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., et al. (2016). Protein sequences bound to mineral surfaces persist into deep time. eLife *5*, e17092. https://doi.org/10.7554/eLife.17092.
9. van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J.A., Dehasque, M., Sağlıcan, E., et al. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature *591*, 265–269. https://doi.org/10.1038/s41586-021-03224-9.
10. Kjær, K.H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T.S., Michelsen, C.S., Sand, K.K., Jelavić, S., Ruter, A.H., Schmidt, A.M.A., et al. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature *612*, 283–291. https://doi.org/10.1038/s41586-022-05453-y.
11. Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763–776. https://doi.org/10.1016/j.cell.2009.01.019.
12. Venanzi, F.M., and Rollo, F. (1990). Mummy RNA lasts longer. Nature *343*, 25–26. https://doi.org/10.1038/343025b0.
13. Tumpey, T.M., Basler, C.F., Aguilar, P.V., Zeng, H., Solórzano, A., Swayne, D.E., Cox, N.J., Katz, J.M., Taubenberger, J.K., Palese, P., et al. (2005). Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science *310*, 77–80. https://doi.org/10.1126/science.1119392.
14. Xiao, Y.L., Kash, J.C., Beres, S.B., Sheng, Z.M., Musser, J.M., and Taubenberger, J.K. (2013). High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. J. Pathol. *229*, 535–545. https://doi.org/10.1002/path.4145.
15. Patrono, L.V., Vrancken, B., Budt, M., Düx, A., Lequime, S., Boral, S., Gilbert, M.T.P., Gogarten, J.F., Hoffmann, L., Horst, D., et al. (2022). Archival influenza virus genomes from Europe reveal genomic variability during the 1918 pandemic. Nat. Commun. *13*, 2314. https://doi.org/10.1038/s41467-022-29614-9.
16. Worobey, M., Watts, T.D., McKay, R.A., Suchard, M.A., Granade, T., Teuwen, D.E., Koblin, B.A., Heneine, W., Lemey, P., and Jaffe, H.W. (2016). 1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature *539*, 98–101. https://doi.org/10.1038/nature19827.
17. Gryseels, S., Watts, T.D., Kabongo Mpolesha, J.M., Larsen, B.B., Lemey, P., Muyembe-Tamfum, J.J., Teuwen, D.E., and Worobey, M. (2020). A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. Proc. Natl. Acad. Sci. USA *117*, 12222–12229. https://doi.org/10.1073/pnas.1913682117.
18. Smith, O., Dunshea, G., Sinding, M.S., Fedorov, S., Germonpre, M., Bocherens, H., and Gilbert, M.T.P. (2019). Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. PLoS Biol. *17*, e3000166. https://doi.org/10.1371/journal.pbio.3000166.
19. Fromm, B., Tarbier, M., Smith, O., Mármol-Sánchez, E., Dalén, L., Gilbert, M.T.P., and Friedländer, M.R. (2021). Ancient microRNA profiles

of 14,300-yr-old canid samples confirm taxonomic origin and provide glimpses into tissue-specific gene regulation from the Pleistocene. RNA *27*, 324–334. https://doi.org/10.1261/rna.078410.120.

20. Keller, A., Kreis, S., Leidinger, P., Maixner, F., Ludwig, N., Backes, C., Galata, V., Guerriero, G., Fehlmann, T., Franke, A., et al. (2017). miRNAs in ancient tissue specimens of the Tyrolean iceman. Mol. Biol. Evol. *34*, 793–801. https://doi.org/10.1093/molbev/msw291.
21. Shaw, B., Burrell, C.L., Green, D., Navarro-Martinez, A., Scott, D., Daroszewska, A., Van'T Hof, R., Smith, L., Hargrave, F., Mistry,S., et al. (2019). Molecular insights into an ancient form of Paget's disease of bone. Proc. Natl. Acad. Sci. USA *116*, 10463–10472. https://doi.org/10.1073/pnas.1820556116.
22. Mármol-Sánchez, E., Fromm, B., Oskolkov, N., Pochón, Z., Kalogeropoulos, P., Eriksson, E., Biryukova, I., Sekar, V., Ersmark, E., Andersson, B., et al. (2023). Historical RNA expression profiles from the extinct Tasmanian tiger. Genome Res. *33*, 1299–1316. https://doi.org/10.1101/gr.277663.123.
23. Sandoval-Velasco, M., Dudchenko, O., Rodríguez, J.A., Pérez Estrada, C., Dehasque, M., Fontsere, C., Mak, S.S.T., Khan, R., Contessoto, V.G., Oliveira Junior, A.B., et al. (2024). Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample. Cell *187*, 3541–3562.e51. https://doi.org/10.1016/j.cell.2024.06.002.
24. Rudaya, N., Protopopov, A., Trofimova, S., Plotnikov, V., and Zhilich, S. (2015). Landscapes of the 'Yuka' mammoth habitat: A palaeobotanical approach. Rev. Palaeobot. Palynol. *214*, 1–8. https://doi.org/10.1016/j.revpalbo.2014.12.003.
25. Chlachula, J., Dyakonov, V.M., Alekseev, A.N., Protopopov, A.V., Klimovskiy, A.I., and Kolesov, S.D. (2025). The central Indigirka in the Last Ice Age, north-east Arctic Siberia. Quat. Sci. Rev. *349*, 108990. https://doi.org/10.1016/j.quascirev.2024.108990.
26. Maschenko, E.N., Potapova, O.R., Heintzman, P.D., Kapp, J.D., Shapiro, B., Protopopov, A.V., Boeskorov, G.G., Pavlov, I.S., Plotnikov, V.V., Kolesov, S.D., et al. (2021). Morphology, individual age, DNA and sex of the Yuka mammoth (*Mammuthus primigenius*) from Northern Yakutia, Russia. Paleontol. J. *55*, 1230–1259. https://doi.org/10.1134/S003103012111006X.
27. Boeskorov, G.G., Protopopov, A.V., Maschenko, E.N., Potapova, O.R., Plotnikov, V.V., Shchelchkova, M.V., Pavlov, I.S., Klimovsky, A.I., Kolesov, S.D., and Gorokhov, G.V. (2021). History of studies of the female woolly mammoth mummy Yuka (*Mammuthus primigenius* (Blumenbach, 1799)). Paleontol. J. *55*, 1215–1223. https://doi.org/10.1134/S0031030121110022.
28. Boeskorov, G.G., Protopopov, A.V., Mashchenko, E.N., Potapova, O.R., Kuznetsova, T.V., Plotnikov, V.V., Grigoryev, S.E., Belolyubskii, I.N., Tomshin, M.D., Shchelchkova, M.V., et al. (2013). New findings of unique preserved fossil mammals in the permafrost of Yakutia. Dokl. Biol. Sci. *452*, 291–295. https://doi.org/10.1134/S0012496613050116.
29. Yamagata, K., Nagai, K., Miyamoto, H., Anzai, M., Kato, H., Miyamoto, K., Kurosaka, S., Azuma, R., Kolodeznikov, I.I., Protopopov, A.V., et al. (2019). Signs of biological activities of 28,000-year-old mammoth nuclei in mouse oocytes visualized by live-cell imaging. Sci. Rep. *9*, 4050. https://doi.org/10.1038/s41598-019-40546-1.
30. Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., et al. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr. Biol. *25*, 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007.
31. Dolenz, S., van der Valk, T., Jin, C., Oppenheimer, J., Sharif, M.B., Orlando, L., Shapiro, B., Dalén, L., and Heintzman, P.D. (2024). Unravelling reference bias in ancient DNA datasets. Bioinformatics *40*, btae436. https://doi.org/10.1093/bioinformatics/btae436.
32. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
33. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
34. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.
35. Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebersberger, I., Pääbo, S., and Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature *439*, 724–727. https://doi.org/10.1038/nature04432.
36. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628. https://doi.org/10.1038/nmeth.1226.
37. Liu, B., Cao, J., Wang, X., Guo, C., Liu, Y., and Wang, T. (2021). Deciphering the tRNA-derived small RNAs: origin, development, and future. Cell Death Dis. *13*, 24. https://doi.org/10.1038/s41419-021-04472-3.
38. Berkovitz, G.D., Fechner, P.Y., Zacur, H.W., Rock, J.A., Snyder, H.M., Migeon, C.J., and Perlman, E.J. (1991). Clinical and pathologic spectrum of 46,XY gonadal dysgenesis: its relevance to the understanding of sex differentiation. Medicine *70*, 375–383. https://doi.org/10.1097/00005792-199111000-00003.
39. Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. *12*, R18. https://doi.org/10.1186/gb-2011-12-2-r18.
40. Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, e72. https://doi.org/10.1093/nar/gks001.
41. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621. https://doi.org/10.1073/pnas.0704665104.
42. Glocke, I., and Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Res. *27*, 1230–1237. https://doi.org/10.1101/gr.219675.116.
43. Peyrégne, S., and Peter, B.M. (2020). AuthentiCT: A model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome Biol. *21*, 246. https://doi.org/10.1186/s13059-020-02123-y.
44. Parmley, J.L., Chamary, J.V., and Hurst, L.D. (2006). Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol. Biol. Evol. *23*, 301–309. https://doi.org/10.1093/molbev/msj035.
45. Fromm, B. (2024). A renaissance of microRNAs as taxonomic and phylogenetic markers in animals. Zool. Scr. *53*, 754–762. https://doi.org/10.1111/zsc.12684.
46. Cammaerts, S., Strazisar, M., De Rijk, P.D., and Del Favero, J. (2015). Genetic variants in microRNA genes: impact on microRNA expression, function, and disease. Front. Genet. *6*, 186. https://doi.org/10.3389/fgene.2015.00186.
47. Fernandez, N., Cordiner, R.A., Young, R.S., Hug, N., MacIas, S., and Cáceres, J.F. (2017). Genetic variation and RNA structure regulate microRNA biogenesis. Nat. Commun. *8*, 15114. https://doi.org/10.1038/ncomms15114.
48. Friedländer, M.R., MacKowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52. https://doi.org/10.1093/nar/gkr688.

49. Kim, H., Lee, Y.Y., and Kim, V.N. (2025). The biogenesis and regulation of animal microRNAs. Nat. Rev. Mol. Cell Biol. *26*, 276–296. https://doi.org/10.1038/s41580-024-00805-0.

50. Bartel, D.P. (2018). Metazoan microRNAs. Cell *173*, 20–51. https://doi.org/10.1016/j.cell.2018.03.006.

51. Wen, J., Ladewig, E., Shenker, S., Mohammed, J., and Lai, E.C. (2015). Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. PLoS Comput. Biol. *11*, e1004441. https://doi.org/10.1371/journal.pcbi.1004441.

52. Fehlmann, T., Backes, C., Pirritano, M., Laufer, T., Galata, V., Kern, F., Kahraman, M., Gasparoni, G., Ludwig, N., Lenhof, H.P., et al. (2019). The sncRNA Zoo: a repository for circulating small noncoding RNAs in animals. Nucleic Acids Res. *47*, 4431–4441. https://doi.org/10.1093/nar/gkz227.

53. Tskhovrebova, L., and Trinick, J. (2003). Titin: properties and family relationships. Nat. Rev. Mol. Cell Biol. *4*, 679–689. https://doi.org/10.1038/nrm1198.

54. Blondelle, J., Marrocco, V., Clark, M., Desmond, P., Myers, S., Nguyen, J., Wright, M., Bremner, S., Pierantozzi, E., Ward, S., et al. (2019). Murine obscurin and Obsl1 have functionally redundant roles in sarcolemmal integrity, sarcoplasmic reticulum organization, and muscle metabolism. Commun. Biol. *2*, 178. https://doi.org/10.1038/s42003-019-0405-7.

55. Kiss, B., Lee, E.J., Ma, W., Li, F.W., Tonino, P., Mijailovich, S.M., Irving, T.C., and Granzier, H.L. (2018). Nebulin stiffens the thin filament and augments cross-bridge interaction in skeletal muscle. Proc. Natl. Acad. Sci. USA *115*, 10369–10374. https://doi.org/10.1073/pnas.1804726115.

56. Garg, A., O'Rourke, J., Long, C., Doering, J., Ravenscroft, G., Bezprozvannaya, S., Nelson, B.R., Beetz, N., Li, L., Chen, S., et al. (2014). KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. J. Clin. Invest. *124*, 3529–3539. https://doi.org/10.1172/JCI74994.

57. Han, S., Cui, C., Zhao, X., Zhang, Y., Zhang, Y., Zhao, J., Shen, X., He, H., Wang, J., Ma, M., et al. (2022). Filamin C regulates skeletal muscle atrophy by stabilizing dishevelled-2 to inhibit autophagy and mitophagy. Mol. Ther. Nucleic Acids *27*, 147–164. https://doi.org/10.1016/j.omtn.2021.11.022.

58. Lu, S., Borst, D.E., and Horowits, R. (2008). Expression and alternative splicing of N-RAP during mouse skeletal muscle development. Cell Motil. Cytoskeleton *65*, 945–954. https://doi.org/10.1002/cm.20317.

59. Ghasemizadeh, A., Christin, E., Guiraud, A., Couturier, N., Abitbol, M., Risson, V., Girard, E., Jagla, C., Soler, C., Laddada, L., et al. (2021). MACF1 controls skeletal muscle function through the microtubule-dependent localization of extra-synaptic myonuclei and mitochondria biogenesis. eLife *10*, e70490. https://doi.org/10.7554/eLife.70490.

60. Al-Sajee, D., Nissar, A.A., Coleman, S.K., Rebalka, I.A., Chiang, A., Wathra, R., van der Ven, P.F.M., Orfanos, Z., and Hawke, T.J. (2015). Xin-deficient mice display myopathy, impaired contractility, attenuated muscle repair and altered satellite cell functionality. Acta Physiol. (Oxf) *214*, 248–260. https://doi.org/10.1111/apha.12455.

61. Laure, L., Suel, L., Roudaut, C., Bourg, N., Ouali, A., Bartoli, M., Richard, I., and Danièle, N. (2009). Cardiac ankyrin repeat protein is a marker of skeletal muscle pathological remodelling. FEBS J. *276*, 669–684. https://doi.org/10.1111/j.1742-4658.2008.06814.x.

62. Wang, L., Geist, J., Grogan, A., Hu, L.R., and Kontrogianni-Konstantopoulos, A. (2018). Thick filament protein network, functions, and disease association. Compr. Physiol. *8*, 631–709. https://doi.org/10.1002/cphy.c170023.

63. Li, M.X., and Hwang, P.M. (2015). Structure and function of cardiac troponin C (*TNNC1*): Implications for heart failure, cardiomyopathies, and troponin modulating drugs. Gene *571*, 153–166. https://doi.org/10.1016/j.gene.2015.07.074.

64. Wei, B., and Jin, J.P. (2016). TNNT1, TNNT2, and TNNT3: isoform genes, regulation, and structure-function relationships. Gene *582*, 1–13. https://doi.org/10.1016/j.gene.2016.01.006.

65. Bottinelli, R., and Reggiani, C. (2000). Human skeletal muscle fibres: molecular and functional diversity. Prog. Biophys. Mol. Biol. *73*, 195–262. https://doi.org/10.1016/s0079-6107(00)00006-7.

66. Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121. https://doi.org/10.1038/nrm3742.

67. Chen, J.F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., and Wang, D.Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat. Genet. *38*, 228–233. https://doi.org/10.1038/ng1725.

68. Herdoiza Padilla, E., Crauwels, P., Bergner, T., Wiederspohn, N., Förstner, S., Rinas, R., Ruf, A., Kleemann, M., Handrick, R., Tuckermann, J., et al. (2019). mir-124-5p regulates phagocytosis of human macrophages by targeting the actin cytoskeleton via the ARP2/3 complex. Front. Immunol. *10*, 2210. https://doi.org/10.3389/fimmu.2019.02210.

69. Yin, H., He, H., Shen, X., Zhao, J., Cao, X., Han, S., Cui, C., Chen, Y., Wei, Y., Xia, L., et al. (2020). miR-9-5p inhibits skeletal muscle satellite cell proliferation and differentiation by targeting IGF2BP3 through the IGF2-PI3K/Akt signaling pathway. Int. J. Mol. Sci. *21*, 1655. https://doi.org/10.3390/ijms21051655.

70. Fattahi, M., Rezaee, D., Fakhari, F., Najafi, S., Aghaei-Zarch, S.M., Beyranvand, P., Rashidi, M.A., Bagheri-Mohammadi, S., Zamani-Rarani, F., Bakhtiari, M., et al. (2023). microRNA-184 in the landscape of human malignancies: a review to roles and clinical significance. Cell Death Dis. *9*, 423. https://doi.org/10.1038/s41420-023-01718-1.

71. Bandiera, S., Pfeffer, S., Baumert, T.F., and Zeisel, M.B. (2015). miR-122 – A key factor and therapeutic target in liver disease. J. Hepatol. *62*, 448–457. https://doi.org/10.1016/j.jhep.2014.10.004.

72. Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Front. Genet. *6*, 2. https://doi.org/10.3389/fgene.2015.00002.

73. Misteli, T., and Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat. Rev. Mol. Cell Biol. *10*, 243–254. https://doi.org/10.1038/nrm2651.

74. Rath, S.P., Gupta, R., Todres, E., Wang, H., Jourdain, A.A., Ardlie, K.G., Calvo, S.E., and Mootha, V.K. (2024). Mitochondrial genome copy number variation across tissues in mice and humans. Proc. Natl. Acad. Sci. USA *121*, e2402291121. https://doi.org/10.1073/pnas.2402291121.

75. Wales, N., Andersen, K., Cappellini, E., Ávila-Arcos, M.C., and Gilbert, M.T.P. (2014). Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. PLoS One *9*, e86827. https://doi.org/10.1371/journal.pone.0086827.

76. Campos, P.F., and Gilbert, M.T.P. (2019). DNA extraction from keratin and chitin. In Ancient DNA. Methods in Molecular Biology, *1963*, B. Shapiro, A. Barlow, P. Heintzman, M. Hofreiter, J. Paijmans, and A. Soares, eds. (Humana Press), pp. 57–63. https://doi.org/10.1007/978-1-4939-9176-1_7.

77. Zhang, M., Cao, P., Dai, Q.Y., Wang, Y.Q., Feng, X.T., Wang, H.R., Wu, H., Ko, A.M.S., Mao, X.W., Liu, Y.C., et al. (2021). Comparative analysis of DNA extraction protocols for ancient soft tissue museum samples. Zool. Res. *42*, 280–286. https://doi.org/10.24272/j.issn.2095-8137.2020.377.

78. Gilardet, A., Lord, E., García, G.O., Xenikoudakis, G., Douka, K., Wooller, M.J., Rowe, T., Martin, M.D., Le Moullec, M., Anisimov, M., et al. (2025). A high-throughput ancient DNA extraction method for large-scale sample screening. Mol. Ecol. Resour. *25*, e14077. https://doi.org/10.1111/1755-0998.14077.

79. Ramsey, C.B. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360. https://doi.org/10.1017/S0033822200033865.

80. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al.

(2020). The IntCal20 Northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon *62*, 725–757. https://doi.org/10.1017/RDC.2020.41.

81. Díez-del-Molino, D., Dehasque, M., Chacón-Duque, J.C., Pečnerová, P., Tikhonov, A., Protopopov, A., Plotnikov, V., Kanellidou, F., Nikolskiy, P., Mortensen, P., et al. (2023). Genomics of adaptive evolution in the woolly mammoth. Curr. Biol. *33*, 1753–1764.e4. https://doi.org/10.1016/j.cub.2023.03.084.
82. Dehasque, M., Morales, H.E., Díez-del-Molino, D., Pečnerová, P., Chacón-Duque, J.C., Kanellidou, F., Muller, H., Plotnikov, V., Protopopov, A., Tikhonov, A., et al. (2024). Temporal dynamics of woolly mammoth genome erosion prior to extinction. Cell *187*, 3531–3540.e13. https://doi.org/10.1016/j.cell.2024.05.033.
83. van der Valk, T., Dehasque, M., Chacón-Duque, J.C., Oskolkov, N., Vartanyan, S., Heintzman, P.D., Pečnerová, P., Díez-del-Molino, D., and Dalén, L. (2022). Evolutionary consequences of genomic deletions and insertions in the woolly mammoth genome. iScience *25*, 104826. https://doi.org/10.1016/j.isci.2022.104826.
84. Knapp, M., Clarke, A.C., Horsburgh, K.A., and Matisoo-Smith, E.A. (2012). Setting the stage – Building and working in an ancient DNA laboratory. Ann. Anat. *194*, 3–6. https://doi.org/10.1016/j.aanat.2011.03.008.
85. Sinding, M.H.S., Arneborg, J., Nyegaard, G., and Gilbert, M.T.P. (2015). Ancient DNA unravels the truth behind the controversial GUS Greenlandic Norse fur samples: the bison was a horse, and the muskox and bears were goats. J. Archaeol. Sci. *53*, 297–303. https://doi.org/10.1016/j.jas.2014.10.028.
86. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162. https://doi.org/10.1093/nar/gky1141.
87. Dehasque, M., Pečnerová, P., Kempe Lagerholm, V., Ersmark, E., Danilov, G.K., Mortensen, P., Vartanyan, S., and Dalén, L. (2022). Development and optimization of a silica column-based extraction protocol for ancient DNA. Genes *13*, 687. https://doi.org/10.3390/genes13040687.
88. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.
89. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87. https://doi.org/10.1093/nar/gkp1163.
90. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3. https://doi.org/10.1093/nar/gkr771.
91. Pečnerová, P., Palkopoulou, E., Wheat, C.W., Skoglund, P., Vartanyan, S., Tikhonov, A., Nikolskiy, P., van der Plicht, J., Díez-del-Molino, D., and Dalén, L. (2017). Mitogenome evolution in the last surviving woolly mammoth population reveals neutral and functional consequences of small population size. Evol. Lett. *1*, 292–303. https://doi.org/10.1002/evl3.33.
92. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
93. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
94. Kutschera, V.E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D.W.G., Khoonsari, P.E., Nystedt, B., et al. (2022). GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. BMC Bioinformatics *23*, 228. https://doi.org/10.1186/s12859-022-04757-0.
95. Breitwieser, F.P., Baker, D.N., and Salzberg, S.L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. Genome Biol. *19*, 198. https://doi.org/10.1186/s13059-018-1568-0.
96. Pochon, Z., Bergfeldt, N., Kırdök, E., Vicente, M., Naidoo, T., van der Valk, T., Altınışık, N.E., Krzewińska, M., Dalén, L., Götherström, A., et al. (2023). aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. Genome Biol. *24*, 242. https://doi.org/10.1186/s13059-023-03083-9.
97. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
98. Breitwieser, F.P., and Salzberg, S.L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics *36*, 1303–1304. https://doi.org/10.1093/bioinformatics/btz715.
99. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics *70*, e102. https://doi.org/10.1002/cpbi.102.
100. Zerbino, D.R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. Curr. Protoc. Bioinformatics *31*, 11.5.1–11.5.12. https://doi.org/10.1002/0471250953.bi1105s31.
101. Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics *28*, 1086–1092. https://doi.org/10.1093/bioinformatics/bts094.
102. Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience *8*, giz100. https://doi.org/10.1093/gigascience/giz100.
103. Meleshko, D., Hajirasouliha, I., and Korobeynikov, A. (2021). coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. Bioinformatics *38*, 1–8. https://doi.org/10.1093/bioinformatics/btab597.
104. Singh, U., and Wurtele, E.S. (2021). orfipy: a fast and flexible tool for extracting ORFs. Bioinformatics *37*, 3019–3020. https://doi.org/10.1093/bioinformatics/btab090.
105. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics *20*, 473. https://doi.org/10.1186/s12859-019-3019-7.
106. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
107. Buchfink, B., Reuter, K., and Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods *18*, 366–368. https://doi.org/10.1038/s41592-021-01101-x.
108. Olendraite, I., Brown, K., and Firth, A.E. (2023). Identification of RNA Virus–derived RdRp sequences in publicly available transcriptomic data sets. Mol. Biol. Evol. *40*, msad060. https://doi.org/10.1093/molbev/msad060.
109. Charon, J., Buchmann, J.P., Sadiq, S., and Holmes, E.C. (2022). RdRp-scan: A bioinformatic resource to identify and annotate divergent RNA viruses in metagenomic sequence data. Virus Evol. *8*, veac082. https://doi.org/10.1093/ve/veac082.
110. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.
111. Katoh, K., Misawa, K., Kuma, K.I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066. https://doi.org/10.1093/nar/gkf436.
112. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.
113. Tumescheit, C., Firth, A.E., and Brown, K. (2022). CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple

sequence alignments. PeerJ *10*, e12983. https://doi.org/10.7717/peerj.12983.

114. Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Res *9*. ISCB Comm J-304. https://doi.org/10.12688/f1000research.23297.2.

115. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

116. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499. https://doi.org/10.1101/gr.209601.116.

117. Feuerborn, T.R., Palkopoulou, E., van der Valk, T., von Seth, J., Munters, A.R., Pečnerová, P., Dehasque, M., Ureña, I., Ersmark, E., Lagerholm, V.K., et al. (2020). Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets. BMC Genomics *21*, 844. https://doi.org/10.1186/s12864-020-07229-y.

118. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

119. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

120. Pečnerová, P., Díez-del-Molino, D., Dussex, N., Feuerborn, T., von Seth, J., van der Plicht, J., Nikolskiy, P., Tikhonov, A., Vartanyan, S., and Dalén, L. (2017). Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. Curr. Biol. *27*, 3505–3510.e3. https://doi.org/10.1016/j.cub.2017.09.064.

121. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

122. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

123. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234. https://doi.org/10.1073/pnas.1318934111.

124. Smit, A., and Hubley, R. (2015). RepeatModeler Open-1.0. http://www.repeatmasker.org.

125. Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. https://www.repeatmasker.org/.

126. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

127. Clarke, A.W., Høye, E., Hembrom, A.A., Paynter, V.M., Vinther, J., Wyrożemski, Ł., Biryukova, I., Formaggioni, A., Ovchinnikov, V., Herlyn, H., et al. (2025). MirGeneDB 3.0: Improved taxonomic sampling, uniform nomenclature of novel conserved microRNA families and updated covariance models. Nucleic Acids Res. *53*, D116–D128. https://doi.org/10.1093/nar/gkae1094.

128. Kent, W.J. (2002). BLAT–the BLAST-like alignment tool. Genome Res. *12*, 656–664. https://doi.org/10.1101/gr.229202.

129. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36. https://doi.org/10.1111/2041-210X.12628.

130. Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. *49*, D192–D200. https://doi.org/10.1093/nar/gkaa1047.

131. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415. https://doi.org/10.1093/nar/gkg595.

132. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629. https://doi.org/10.1093/bioinformatics/btz931.

133. Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics *13*, 397–406. https://doi.org/10.1074/mcp.M113.035600.

134. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

135. Keays, M. (2025). ExpressionAtlas: Download datasets from EMBL-EBI Expression Atlas. R package version 1.32.0. https://bioconductor.org/packages/ExpressionAtlas.

136. Zhu, G., Mayer-Wagner, S., Schröder, C., Woiczinski, M., Blum, H., Lavagi, I., Krebs, S., Redeker, J.I., Hölzer, A., Jansson, V., et al. (2015). Comparing effects of perfusion and hydrostatic pressure on gene profiles of human chondrocyte. J. Biotechnol. *210*, 59–65. https://doi.org/10.1016/j.jbiotec.2015.06.409.

137. Armstrong, D.L., McGowen, M.R., Weckle, A., Pantham, P., Caravas, J., Agnew, D., Benirschke, K., Savage-Rumbaugh, S., Nevo, E., Kim, C.J., et al. (2017). The core transcriptome of mammalian placentas and the divergence of expression with placental shape. Placenta *57*, 71–78. https://doi.org/10.1016/j.placenta.2017.04.015.

138. Toh, H., Yang, C., Formenti, G., Raja, K., Yan, L., Tracey, A., Chow, W., Howe, K., Bergeron, L.A., Zhang, G., et al. (2022). A haplotype-resolved genome assembly of the Nile rat facilitates exploration of the genetic basis of diabetes. BMC Biol. *20*, 245. https://doi.org/10.1186/s12915-022-01427-8.

139. Lorenzi, L., Chiu, H.S., Avila Cobos, F., Gross, S., Volders, P.J., Cannoodt, R., Nuytens, J., Vanderheyden, K., Anckaert, J., Lefever, S., et al. (2021). The RNA Atlas expands the catalog of human non-coding RNAs. Nat. Biotechnol. *39*, 1453–1465. https://doi.org/10.1038/s41587-021-00936-1.

140. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., and et al.. (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470. https://doi.org/10.1038/nature13182.

141. Srivastava, A., Meisgen, F., Pasquali, L., Munkhammar, S., Xia, P., Ståhle, M., Landén, N.X., Pivarcsi, A., and Sonkoly, E. (2019). Next-generation sequencing identifies the keratinocyte-specific miRNA signature of psoriasis. J. Invest. Dermatol. *139*, 2547–2550.e12. https://doi.org/10.1016/j.jid.2019.05.019.

142. Akat, K.M., Moore-McGriff, D., Morozov, P., Brown, M., Gogakos, T., Correa Da Rosa, J., Mihailovic, A., Sauer, M., Ji, R., Ramarathnam, A., et al. (2014). Comparative RNA-sequencing analysis of myocardial and circulating small RNAs in human heart failure and their utility as biomarkers. Proc. Natl. Acad. Sci. USA *111*, 11151–11156. https://doi.org/10.1073/pnas.1401724111.

143. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25. https://doi.org/10.1186/gb-2010-11-3-r25.

144. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

145. Healy, J., and McInnes, L. (2024). Uniform manifold approximation and projection. Nat. Rev. Methods Primers *4*, 82. https://doi.org/10.1038/s43586-024-00363-x.

146. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics *37*, 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Mammoth 1	This study	Yuka
Mammoth 2	This study	Yukagir
Mammoth 3	This study	Yana-Inidgirka
Mammoth 4	This study	Oymyakon
Mammoth 5	This study	Belaya Gora
Mammoth 6	This study	Gold mammoth
Mammoth 7	This study	Maksunuokha
Mammoth 8	This study	Lyakhovskiy 1
Mammoth 9	This study	Big Lyakhovskiy
Mammoth 10	This study	Chris Waddle
Chemicals, peptides, and recombinant proteins		
USER enzyme	New England Biolabs	NEB #M5508
AccuPrime reaction mix	Life Technologies	Cat #12344040
AccuPrime Pfx DNA polymerase	Life Technologies	Cat #12344024
Agencourt AMPure XP beads	Beckman Coulter	Cat #10136224
Critical commercial assays		
PowerBead tube metal 2.38 mm	Qiagen	Cat #13117-50
mirVana microRNA isolation kit	Thermo Fisher Scientific	Cat #AM1560
Qubit microRNA assay kit	Thermo Fisher Scientific	Cat#Q32881
Qubit RNA HS assay kits	Thermo Fisher Scientific	Cat #Q32852
NEXTFLEX™ small RNA-seq kit v3	Bioo Scientific	Cat #5132-06
Qubit dsDNA broad-range assay kit	Thermo Fisher Scientific	Cat #Q32850
Agilent high-sensitivity DNA kit	Agilent	Cat #5067-4626
NextSeq 500/550 high-output kit v2.5	Illumina	Cat #20024906
NextSeq 1000/2000 P1 XLEAP-SBS kit	Illumina	Cat #20100982
NovaSeq 6000 S4 Reagent Kit v1.5	Illumina	Cat #20028313
Deposited data		
Single-end aRNA sequencing data	This study	PRJNA1190340
Paired-end aDNA sequencing data	This study	PRJNA1190340
Paired-end aRNA sequencing data	This study	PRJNA1256124
KrakenUniq database based on full NCBI NT	Pochon et al.[96]	https://doi.org/10.17044/scilifelab.20205504
Woolly mammoth	Dehasque et al.[82]	E464, Wra4.9Ka
Woolly mammoth	Palkopoulou et al.[30]	E467, Wra4.3K
Woolly mammoth	van der Valk et al.[83]	L163, Chu31.9K
Woolly mammoth	Díez-del-Molino et al.[81]	L164, Chu17.0K
Woolly mammoth	Dehasque et al.[82]	L386, Wra7.4K
Woolly mammoth	Dehasque et al.[82]	L387, Wra5.7K
Woolly mammoth	Dehasque et al.[82]	L414, Wra7.9K
Woolly mammoth	Dehasque et al.[82]	L416, Wra7.1Ka
Woolly mammoth	Dehasque et al.[82]	L422, Wra7.1Kb
Woolly mammoth	Dehasque et al.[82]	L423, Wra6.2K
Woolly mammoth	Dehasque et al.[82]	L456, Wra8.5K

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Woolly mammoth	Dehasque et al.[82]	M17, Wra8.3K
Woolly mammoth	Dehasque et al.[82]	M32, Wra4.9Kb
Woolly mammoth	van der Valk et al.[83]	M6, Wra24.0K
Woolly mammoth	Dehasque et al.[82]	MD076, Wra9.2K
Woolly mammoth	Dehasque et al.[82]	MD080, Wra5.2K
Woolly mammoth	Díez-del-Molino et al.[81]	MD090, Yak52.3K
Woolly mammoth	Palkopoulou et al.[30]	Oymyakon, Oim44.2K
Woolly mammoth	Díez-del-Molino et al.[81]	P004, NSI12.2K
Woolly mammoth	Díez-del-Molino et al.[81]	P005, NSI12.8K
Woolly mammoth	Maschenko et al.[26]	JK294
Woolly mammoth	Maschenko et al.[26]	JK296
Woolly mammoth	Yamagata et al.[29]	Japanese Yuka
Asian elephant	SAMN28571076	PRJNA840935
Human	Uhlén et al.[134]	ERR579130
Human	Uhlén et al.[134]	ERR579141
Human	Uhlén et al.[134]	ERR579142
Human	Uhlén et al.[134]	ERR579143
Human	Uhlén et al.[134]	ERR579149
Human	Uhlén et al.[134]	ERR579152
Human	Fagerberg et al.[133]	ERR315328
Human	Fagerberg et al.[133]	ERR315331
Human	Fagerberg et al.[133]	ERR315356
Human	Fagerberg et al.[133]	ERR315367
Human	Fagerberg et al.[133]	ERR315384
Human	Fagerberg et al.[133]	ERR315389
Human	Fagerberg et al.[133]	ERR315413
Human	Fagerberg et al.[133]	ERR315430
Human	Fagerberg et al.[133]	ERR315435
Human	Fagerberg et al.[133]	ERR315339
Human	Fagerberg et al.[133]	ERR315372
Human	Fagerberg et al.[133]	ERR315376
Human	Fagerberg et al.[133]	ERR315401
Human	Fagerberg et al.[133]	ERR315460
Human	Fagerberg et al.[133]	ERR315464
Human	Fagerberg et al.[133]	ERR315327
Human	Fagerberg et al.[133]	ERR315394
Human	Fagerberg et al.[133]	ERR315414
Human	Fagerberg et al.[133]	ERR315451
Human	Fagerberg et al.[133]	ERR315463
Human	Fagerberg et al.[133]	ERR315326
Human	Fagerberg et al.[133]	ERR315341
Human	Fagerberg et al.[133]	ERR315346
Human	Fagerberg et al.[133]	ERR315353
Human	Fagerberg et al.[133]	ERR315424
Human	Fagerberg et al.[133]	ERR315439
Human	Fagerberg et al.[133]	ERR315444
Human	Fagerberg et al.[133]	ERR315487
Human	Fagerberg et al.[133]	ERR315383
Human	Fagerbeg et al.[133]	ERR315443

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human	Fagerberg et al.[133]	ERR315468
Human	Fagerberg et al.[133]	ERR315494
Human	Fagerberg et al.[133]	ERR315432
Human	Fagerberg et al.[133]	ERR315455
Human	Fagerberg et al.[133]	ERR315477
Human	Fagerberg et al.[133]	ERR315350
Human	Fagerberg et al.[133]	ERR315351
Human	Fagerberg et al.[133]	ERR315352
Human	Fagerberg et al.[133]	ERR315391
Human	Fagerberg et al.[133]	ERR315415
Human	Fagerberg et al.[133]	ERR315446
Human	Fagerberg et al.[133]	ERR315456
Human	Fagerberg et al.[133]	ERR315492
Human	Fagerberg et al.[133]	ERR315380
Human	Fagerberg et al.[133]	ERR315402
Human	Fagerberg et al.[133]	ERR315458
Human	Fagerberg et al.[133]	ERR315482
Human	Fagerberg et al.[133]	ERR315336
Human	Fagerberg et al.[133]	ERR315374
Human	Fagerberg et al.[133]	ERR315375
Human	Fagerberg et al.[133]	ERR315377
Human	Fagerberg et al.[133]	ERR315399
Human	Fagerberg et al.[133]	ERR315476
Human	Fagerberg et al.[133]	ERR315478
Human	Zhu et al.[136]	SRS945156
Human	Zhu et al.[136]	SRS945155
Human	Zhu et al.[136]	SRS945625
Human	Fagerberg et al.[133]	ERR315332
Human	Fagerberg et al.[133]	ERR315342
Human	Fagerberg et al.[133]	ERR315343
Human	Fagerberg et al.[133]	ERR315378
Human	Fagerberg et al.[133]	ERR315431
Human	Fagerberg et al.[133]	ERR315333
Human	Fagerberg et al.[133]	ERR315395
Human	Fagerberg et al.[133]	ERR315396
Human	Fagerberg et al.[133]	ERR315404
Human	Fagerberg et al.[133]	ERR315406
Human	Fagerberg et al.[133]	ERR315425
Human	Fagerberg et al.[133]	ERR315469
Human	Fagerberg et al.[133]	ERR315486
Human	Fagerberg et al.[133]	ERR315329
Human	Fagerberg et al.[133]	ERR315371
Human	Fagerberg et al.[133]	ERR315373
Human	Fagerberg et al.[133]	ERR315387
Human	Fagerberg et al.[133]	ERR315390
Human	Fagerberg et al.[133]	ERR315393
Human	Fagerberg et al.[133]	ERR315426
Human	Fagerberg et al.[133]	ERR315440
Human	Fagerberg et al.[133]	ERR315441

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human	Fagerberg et al.[133]	ERR315471
Human	Fagerberg et al.[133]	ERR315475
Human	Fagerberg et al.[133]	ERR315488
Human	Fagerberg et al.[133]	ERR315493
Human	Fagerberg et al.[133]	ERR315337
Human	Fagerberg et al.[133]	ERR315358
Human	Fagerberg et al.[133]	ERR315363
Human	Fagerberg et al.[133]	ERR315397
Human	Fagerberg et al.[133]	ERR315412
Human	Fagerberg et al.[133]	ERR315422
Human	Fagerberg et al.[133]	ERR315428
Human	Fagerberg et al.[133]	ERR315483
Human	Fagerberg et al.[133]	ERR315491
Human	Fagerberg et al.[133]	ERR315325
Human	Fagerberg et al.[133]	ERR315382
Human	Fagerberg et al.[133]	ERR315418
Human	Fagerberg et al.[133]	ERR315420
Human	Fagerberg et al.[133]	ERR315449
Human	Fagerberg et al.[133]	ERR315459
Elephant	Toh et al.[138]	SRR27318051
Elephant	Toh et al.[138]	SRR27318052
Elephant		SRR19501519
Elephant		SRR19501517
Elephant		SRR19501515
Elephant		SRR19501516
Elephant		SRR19501518
Elephant	Armstrong et al.[137]	SRR3222430
Human	Lorenzi et al.[139]	SRR10264719
Human	Lorenzi et al.[139]	SRR10264720
Human	Lorenzi et al.[139]	SRR10264721
Human	Lorenzi et al.[139]	SRR10264722
Human	Lorenzi et al.[139]	SRR10264489
Human	Lorenzi et al.[139]	SRR10264490
Human	Lorenzi et al.[139]	SRR10264491
Human	Lorenzi et al.[139]	SRR10264492
Human	Srivastava et al.[141]	SRR8858261
Human	Srivastava et al.[141]	SRR8858262
Human	Srivastava et al.[141]	SRR8858263
Human	Srivastava et al.[141]	SRR8858264
Human	Srivastava et al.[141]	SRR8858265
Human	Srivastava et al.[141]	SRR8858266
Human	Srivastava et al.[141]	SRR8858267
Human	Srivastava et al.[141]	SRR8858268
Human	Lorenzi et al.[139]	SRR10264369
Human	Lorenzi et al.[139]	SRR10264370
Human	Lorenzi et al.[139]	SRR10264371
Human	Lorenzi et al.[139]	SRR10264372
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041463

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041468
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041613
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041352
Human	Lorenzi et al.[139]	SRR10264389
Human	Lorenzi et al.[139]	SRR10264390
Human	Lorenzi et al.[139]	SRR10264391
Human	Lorenzi et al.[139]	SRR10264392
Human	Lorenzi et al.[139]	SRR10264628
Human	Lorenzi et al.[139]	SRR10264629
Human	Lorenzi et al.[139]	SRR10264630
Human	Lorenzi et al.[139]	SRR10264631
Human	Lorenzi et al.[139]	SRR10264237
Human	Lorenzi et al.[139]	SRR10264238
Human	Lorenzi et al.[139]	SRR10264239
Human	Lorenzi et al.[139]	SRR10264240
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041531
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041538
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041557
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041563
Human	FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al.[140]	DRR041586
Human	Lorenzi et al.[139]	SRR10264269
Human	Lorenzi et al.[139]	SRR10264270
Human	Lorenzi et al.[139]	SRR10264271
Human	Lorenzi et al.[139]	SRR10264272
Human	Akat et al.[142]	SRR1044425
Human	Akat et al.[142]	SRR1044427
Human	Akat et al.[142]	SRR1044429
Human	Akat et al.[142]	SRR1044431
Human	Akat et al.[142]	SRR1044433
Human	Akat et al.[142]	SRR1044434
Human	Lorenzi et al.[139]	SRR10264980
Human	Lorenzi et al.[139]	SRR10264981
Human	Lorenzi et al.[139]	SRR10264982
Human	Lorenzi et al.[139]	SRR10264983
Human	Lorenzi et al.[139]	SRR10263632
Human	Lorenzi et al.[139]	SRR10263633
Human	Lorenzi et al.[139]	SRR10263634
Human	Lorenzi et al.[139]	SRR10263635
Human	Lorenzi et al.[139]	SRR10263548
Human	Lorenzi et al.[139]	SRR10263549
Human	Lorenzi et al.[139]	SRR10263550
Human	Lorenzi et al.[139]	SRR10263551
Human	Lorenzi et al.[139]	SRR10263656

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human	Lorenzi et al.[139]	SRR10263657
Human	Lorenzi et al.[139]	SRR10263658
Human	Lorenzi et al.[139]	SRR10263659
Human	Lorenzi et al.[139]	SRR10264129
Human	Lorenzi et al.[139]	SRR10264130
Human	Lorenzi et al.[139]	SRR10264131
Human	Lorenzi et al.[139]	SRR10264132
Elephant	Fehlmann et al.[52]	SRR7909615
Tasmanian tiger	Mármol-Sánchez et al.[22]	SRR23147616
Tasmanian tiger	Mármol-Sánchez et al.[22]	SRR23147611
Canid	Smith et al.[18]	SRR8090328
Canid	Smith et al.[18]	SRR8090327
Canid	Smith et al.[18]	SRR8090318
Wolf	Smith et al.[18]	SRR8090324
Wolf	Smith et al.[18]	SRR8090323
Software and algorithms		
OxCal v4.3	Ramsey[79]	https://c14.arch.ox.ac.uk/oxcal.html
bcl2Fastq v1.8.3	Illumina	https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
Cutadapt v3.2	Martin[92]	https://github.com/marcelm/cutadapt/
fastp v0.24	Chen et al.[93]	https://github.com/OpenGene/fastp
FASTX-Toolkit v0.0.14	Debian	https://github.com/Debian/fastx-toolkit
Seqtk v1.3-r106	Lh3	https://github.com/lh3/seqtk
GenErode v0.4	Kutschera et al.[94]	https://github.com/NBISweden/GenErode
KrakenUniq v0.7.3	Breitwieser et al.[95]	https://github.com/fbreitwieser/krakenuniq
Pavian v1.2.0	Breitwieser and Salzberg[98]	https://github.com/fbreitwieser/pavian
Spades v3.15.5	Prjibelski et al.[99]	https://github.com/ablab/spades
Velvet v1.2.10	Zerbino[100]	https://github.com/dzerbino/velvet
Oases v0.2.09	Schulz et al.[101]	https://github.com/dzerbino/oases
orfipy v0.0.4	Singh and Wurtele[104]	https://github.com/urmi-21/orfipy
HHsearch v3.3.0	Steinegger et al.[105]	https://github.com/soedinglab/hh-suite
HMMscan v3.3.2	Eddy[106]	https://github.com/EddyRivasLab/hmmer/tree/master
Diamond BLASTP v2.1.9	Buchfink et al.[107]	https://github.com/bbuchfink/diamond
NCBI BLAST v2.16.0	NCBI	https://blast.ncbi.nlm.nih.gov
MMSeqs2 v15	Steinegger and Söding[110]	https://github.com/soedinglab/MMseqs2
MAFFT v7.520	Katoh et al.[111]	https://mafft.cbrc.jp/alignment/software/
FastTree v2.1.11	Price et al.[112]	https://software.cqls.oregonstate.edu/updates/fasttree-2.1.11/
ClAlign v1.1.4	Tumescheit et al.[113]	https://github.com/KatyBrown/ClAlign
Bowtie v1.3.0	Langmead et al.[34]	https://github.com/BenLangmead/bowtie
Bowtie2 v2.4.2	Langmead and Salzberg[32]	https://github.com/BenLangmead/bowtie2
BWA v0.7.17	Li and Durbin[33]	https://github.com/lh3/bwa
AMBER	Dolenz et al.[31]	https://github.com/tvandervalk/AMBER
gffread v0.12.6	Pertea and Pertea[114]	https://github.com/gpertea/gffread
Sambamba v0.8.0	Tarasov et al.[115]	https://github.com/biod/sambamba
UMI-Tools v1.1.2	Smith et al.[116]	https://github.com/CGATOxford/UMI-tools
SAMtools v.1.8	Li et al.[118]	https://github.com/samtools/samtools
GATK v3.8-0	McKenna et al.[119]	https://github.com/broadinstitute/gatk

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BEDtools v2.31.0	Quinlan and Hall[121]	https://github.com/arq5x/bedtools2
BCFtools v1.8	Danecek et al.[122]	https://github.com/samtools/bcftools
PMDtools v0.60	Skoglund et al.[123]	https://github.com/pontussk/PMDtools
RepeatModeler v2.0.1	Smit and Hubley[124]	https://www.repeatmasker.org/dev/RepeatModeler/
RepeatMasker v4.0.9	Smit et al.[125]	https://www.repeatmasker.org/dev/RepeatMasker/
Variant Effect Predictor (VEP) v113.0	McLaren et al.[126]	https://github.com/Ensembl/ensembl-vep
BLAT	Kent[128]	https://mart.ensembl.org/Homo_sapiens/Tools/Blast
rotl v3.1.0	CRAN	https://CRAN.R-project.org/package=rotl
ggtree v3.12.0	Yu et al.[129]	https://www.bioconductor.org/packages/release/bioc/html/ggtree.html
Rfam v15.0	Kalvari et al.[130]	https://rfam.org/
MiRDeep2	Friedländer et al.[48]	https://github.com/rajewsky-lab/mirdeep2
Mfold	Zuker[131]	https://rothlab.ucdavis.edu/genhelp/mfold.html
ShinyGO v0.81	Ge et al.[132]	https://bioinformatics.sdstate.edu/go/
ExpressionAtlas v1.32.0	Keays[135]	https://bioconductor.org/packages/release/bioc/html/ExpressionAtlas.html
edgeR v4.2.1	Robinson et al.[144]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
dendsort v0.3.4	CRAN	https://cran.r-project.org/web/packages/dendsort/index.html
umap v0.2.10.0	CRAN	https://github.com/tkonopka/umap
Liftoff v1.6.3	Shumate and Salzberg[146]	https://github.com/agshumate/Liftoff
Minimap2 v2.24-r1122	Li[147]	https://github.com/lh3/minimap2
Other		
Annotation file (GTF) of the woolly mammoth genome assembly reported by Sandoval-Velasco et al.[23]	This study	https://github.com/emarmolsanchez/aRNA/tree/main/Annotation
Interactive woolly mammoth metagenomics report generated with Pavian[98]	This study	https://github.com/emarmolsanchez/aRNA/tree/main/Meta
Interactive woolly mammoth metatranscriptomics report generated with Pavian[98]	This study	https://github.com/emarmolsanchez/aRNA/tree/main/Meta

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Woolly mammoths

Permafrost-preserved tissues (muscle and skin) from a total of 10 individual woolly mammoths were used in this study. Given metagenomic, metatranscriptomics and endogenous quantification of mammoth-like endogenous quality data, we decided to focus our analyses on mammoths 1, 4, and 10. Mammoth 1: Muscle tissue from this specimen belongs to Yuka, an exceptionally well-preserved juvenile woolly mammoth discovered on the mainland bluff of Oyogos Yar. Originally identified as female based on external examination of the genital area of the carcass,[26] it has been determined to be a genetic male (XY) by four independent sequencing sources of aDNA data (Table S3, #3, one of them generated for this study), as well as aRNA data reported in this study. Radiocarbon dating places this specimen at approximately 39,000 calibrated YBP (GrA 53289). Mammoth 4: Muscle tissue from this specimen corresponds to the Oymyakon mammoth, a male calf unearthed at the Ol'chan mine site in Yakutia (Oymyakon region) and dated to roughly 44,000 calibrated YBP (GrA-30727). Mammoth 10: Skin tissue used from this specimen belongs to a mammoth informally known as "Chris Waddle", and previously reported as IN18-032.[23] It represents a female individual recovered near Belaya Gora in the Sakha Republic, with an estimated age of >44,900 years (OxA-38763). However, mitochondrial phylogenetic analysis refined its age

to around 52,000 YBP.[23] A summary of additional metadata available for these specimens, along with the remaining mammoths analyzed in this study, is provided in Table S1, #1. Additional publicly available woolly mammoth aDNA sequencing data used for Y chromosome analyses and SNV analyses are reported in Table S3, #2.

Modern elephants

Elephant DNA and small RNA sequencing data are publicly accessible and their identification is reported in Tables S3, #2 and S5, #6.

Tasmanian tiger

aRNA sequencing data from a Tasmanian tiger (NRM-MA590213) previously reported[22] was used for tissue-specific clustering analyses. Additional details are available in Table S5, #6.

Canid

aRNA sequencing data from end-Pleistocene canid previously reported[18] was used for tissue-specific clustering analyses. Additional details are available in Table S5, #6.

Humans

Publicly available RNA-seq data from a collection of human tissues were retrieved and used for tissue-specific clustering analyses. Additional details are available in Table S5, #6.

METHOD DETAILS

Sample collection and dating

A total of ten tissue samples, each of which belonging to a different woolly mammoth specimen, were opportunistically collected from museum collections except for one, which was directly sampled from the field and stored at the Center for Palaeogenetics (CPG) ancient DNA lab facilities (Table S1, #1). All tissues were labeled either as fibrous/muscle-like or skin. Radiocarbon dating was already available for 8 out of the 10 mammoths, but only 4 of them had traceable ^{14}C identification (Table S1, #1). All available radiocarbon dates with corresponding error estimates were calibrated with OxCal v4.3[79] using the IntCal20 Northern Hemisphere calibration curve[80] and are reported in calibrated years before present (YBP, Table S1, #1). For mammoths without enough data to report calibrated YBP, age is determined as recorded in previous studies[81–83] or as stated by sample providers.

Laboratory methods

All laboratory procedures were performed in dedicated ancient DNA laboratory facilities within the Centre for Palaeogenetics (CPG) department at Stockholm University in Sweden. Strict standard guidelines for working with ancient biomolecules and avoiding external contamination were carefully followed.[84]

Ancient RNA

We followed protocols as previously reported for historical RNA extraction.[22] Briefly, approximately 80 mg of tissue was sectioned from each woolly mammoth sample using a sharp scalpel and pulverized using liquid nitrogen using a mortar and pestle. Subsequently, the resulting tissue powder was submerged in 1 mL of digestion buffer[85] suited for highly fibrous and keratin-rich tissues and incubated for 30 minutes at 37°C. Optionally, for samples that showed limited digestion after incubation, an additional homogenization step was implemented by mechanical lysis using 2 mL PowerBead tubes (Qiagen) loaded with 2.38-mm metallic beads in a TissueLyser LT equipment (Qiagen). The resulting lysate for each sample was then used for total RNA extraction with the mirVana microRNA isolation kit (Thermo Fisher Scientific) and Acid-Phenol:Chloroform solution according to the manufacturer's specifications except for the following steps: (1) substituting the initial lysis/binding buffer with the homogenized lysate obtained after tissue digestion, and (2) performing the final elution in 25 μL ultrapure nuclease-free H_2O and repeating the elution flow through the filter cartridge twice. No size selection on the total RNA extract was performed. Total RNA concentration from each eluted extract was then determined in triplicate using Qubit microRNA and Qubit RNA HS assay kits in a Qubit 2.0 fluorometer (Thermo Fisher Scientific). Total RNA extracts concentration using both microRNA and RNA HS Qubit assays are shown in Table S1, #2. Subsequently, we prepared RNA sequencing libraries with the NEXTFLEX™ small RNA-seq kit v3 (Bioo Scientific) using 23× PCR amplification cycles with no size selection. Negative controls were included in duplicate by preparing additional libraries without the addition of RNA extracts in the initial ligation reaction. A positive control spike-in was added by using a 21-nt microRNA-like sequence not matching any known microRNA according to miRBase database[86] and provided by the NEXTFLEX™ small RNA-seq library prep kit. Library concentration per each sample was determined with a Qubit dsDNA broad-range assay kit (Thermo Fisher Scientific), while cDNA fragment size distribution was assessed with the Agilent high-sensitivity DNA kit assay in a Bioanalyzer 2100 equipment (Agilent Technologies). Fragment size profiles for each woolly mammoth tissue cDNA library and additional controls are shown in Figure S1.

Ancient DNA

Approximately 50 mg of tissue was sectioned from each woolly mammoth sample using a sharp scalpel. Subsequently, all samples were submerged in 1 mL of a digestion buffer[85] as done for aRNA and incubated overnight at 56°C. DNA extraction was then carried out as reported by Dehasque et al. protocol from day two.[87] We then prepared double-stranded sequencing libraries using the

protocol by Meyer and Kircher,[88] with modifications as reported by Dehasque et al.,[87] including treatment with either 3 or 6 μL of USER enzyme (New England Biolabs) to excise uracil bases from post-mortem damage except at CpG sites.[89] Clean-up was done with MinElute purification columns (Qiagen). Library indexing was performed in 25 μL reaction volumes containing 1× AccuPrime reaction mix (Life technologies), 0.3 μM of each indexing primer, 1.25U AccuPrime Pfx DNA polymerase (Life technologies), and 3 μL of each DNA library. Unique barcodes were used for dual indexing of the libraries.[90] The following conditions were applied for indexing the PCR reaction: 95°C for 2 minutes, 12 to 16 cycles of 95°C for 15 seconds, 60°C for 30 seconds, and 68°C for 30 seconds. Polymerase chain reaction (PCR) amplification was performed following Pečnerová et al.[91] In order to determine the most appropriate number of PCR cycles for indexing PCR per each sequencing library, 12 PCR cycles were initially set following reaction conditions as described above. We then ran the amplified libraries on an agarose gel and used the intensity of the gel bands as indicative of the optimal numbers of PCR cycles to apply for the final indexing PCR. This is intended to reduce the PCR cycles to the optimal minimum and reduce clonality. Amplified indexed libraries were then pooled in equal volumes per sample and size selected using Agencourt AMPure XP beads (Beckman Coulter). A high-sensitivity DNA chip was then used to assess the library pool concentration in a BioAnalyzer 2100 equipment (Agilent Technologies).

Sequencing

Ancient RNA libraries were sequenced on a NextSeq 500 sequencing system (Illumina) using the Illumina NextSeq500/550 high-output sequencing reagent kit v2.5 (single-end, 75 cycles). After the first run including all ten woolly mammoth samples analyzed in this study, mammoth 1 (Yuka) cDNA library was individually re-sequenced with the same NextSeq 500 protocol to obtain deep sequencing data. Further details on deep sequencing statistics are available at Table S2, #1 and Figure S2C. Additionally, we generated paired-end sequencing data from the same mammoth 1 cDNA library on a NextSeq 2000 system (Illumina) with the NextSeq 1000/2000 P1 XLEAP-SBS Reagent Kit (300 cycles, see Table S2, #4). This was intended to ascertain the longest endogenous mammoth-like aRNA sequences able to be characterized. Ancient DNA libraries were sent to the National Genomics Infrastructure (NGI Stockholm) for sequencing using the Illumina NovaSeq platform (S4 flow cell, 2x100 or 2x150, paired-end).

Raw data preprocessing

Raw sequences from ancient DNA and RNA sequencing products were demultiplexed and converted from Bcl to Fastq using bcl2Fastq v1.8.3 (Illumina).

Ancient RNA

The resulting raw fastq files of each library were processed for adapter removal with the Cutadapt v3.2 tool[92] allowing a minimum of 18 nucleotides per sequence after trimming and a maximum error rate of 10% in adapter sequence detection. Paired-end RNA sequences were merged using the fastp v0.24[93] with the *merge* option. After this step, both single-end and paired-end RNA sequencing data were processed following the same pipeline: Trimmed sequences were collapsed for a first PCR deduplication procedure using the *fastx_collapser* function from FASTX-Toolkit v0.0.14 (https://github.com/Debian/fastx-toolkit). Unique molecular identifiers (UMIs) included during library preparation were then removed from trimmed and collapsed sequences with the *trimfq* tool (*-b 4 -e 4*) from Seqtk v1.3-r106 (https://github.com/lh3/seqtk), while retaining UMI sequence tags for further UMI-based deduplication.

Ancient DNA

The resulting raw fastq files of each library were processed with GenErode v0.4 pipeline[94] using default parameters for ancient/historical DNA. We then merged paired-end sequences and adapters were removed with fastp v0.22.[93] Only merged fragments with a minimum length of 30 nucleotides were retained for subsequent analyses.

Metagenomics and metatranscriptomics

Taxonomic analysis of ancient RNA and DNA sequences was performed to assess the overall species-specific content of each library, targeting both endogenous woolly mammoth sequences, as well as any additional exogenous contamination source from ancient or modern origin. The analysis was implemented on adapter-trimmed sequences before PCR deduplication (DNA), and after UMI removal in the case of RNA sequences. Only sequences with a length of at least 30 nucleotides were considered. Taxonomic classification of the sequences was performed using KrakenUniq v0.7.3[95] with a custom database[96] (https://doi.org/10.17044/scilifelab.20205504) based on the NCBI non-redundant reference nucleotide (NT) database, which is commonly used in the standard *blastn* algorithm.[97] Sequence assignment relied on *k*-mer classification. The resulting taxonomic assignments for DNA and RNA sequences were then assessed for mammoth-like origin using Afrotheria superorder and woolly mammoth (*M. primigenius*) independently as the taxonomic level (Table S1, #3). Comprehensive reports of classification analyses were separately produced for DNA and RNA sequence assignments using the Pavian v1.2.0 R package,[98] and are available at https://github.com/emarmolsanchez/aRNA.

Identification of RNA viruses

Adapter-trimmed RNA sequences after UMI removal without PCR deduplication for each of the ten woolly mammoth samples were individually treated, as well as merged altogether, for contig assembly using Spades v3.15.5,[99] Velvet v1.2.10[100] and Oases v0.2.09[101] tools. For Spades, both *rnaSpAdes*[102] and *rnaviralSPAdes*[103] modes were used on all samples. The *soft_filtered_transcript* fasta output file from *rnaSPAdes*, which includes lower confidence contigs, was used for subsequent analyses in the case of Spades tool. Default settings were implemented for running Velvet, plus the *-short* setting for allowing short single-end sequences and *k*-mer size increments from 13 to 21. Contigs resulting from the first pass Velvet assembly were then combined for a second run

using the *-long* setting with *k*-mer length of 19 and *-conserveLong yes* function. The resulting output was used as input for Oases, merging contigs across all samples. Subsequently, open reading frames (ORFs) were identified using the orfipy v0.0.4 tool[104] with *between_stops* active options and including partial ORFs.

Screening for the presence of RNA viruses was performed using HHsearch v3.3.0,[105] HMMscan v3.3.2[106] and Diamond BLASTP v2.1.9[107] tools. For HHsearch, pHMM alignments from Olendraite et al.[108] were converted into HHsearch profiles and used as targets. For HMMscan, the Olendraite et al.[108] profiles were used directly, along with those from Charon et al.[109] For BLASTP, viral RdRp regions were extracted from sequences provided elsewhere.[108,109] HHsearch and BLASTP results were minimally filtered: HHSearch to keep hits with an E-value <0.05, probability >50 and >20 columns, and BLAST to keep hits with a bit score and percent identity >20, alignment length >30 and E-value <0.05. All putative hits from the three tools implemented were compared to a database of all bacterial peptides available in Ensembl Bacteria release v59 using Diamond BLASTP.[107] Hits initially identified with BLASTP were reassigned as bacterial if they scored higher against a bacterial reference sequence than against any viral sequence. Hits initially identified using pHMMs were reassigned if they had a high-quality BLAST hit (bit score >50, alignment length >50, percent identity >50) against a bacterial reference sequence. Hits which did not match a bacterial reference sequence were compared to the nr database using the *blastp* tool from NCBI BLAST v2.16.0 online webserver (https://blast.ncbi.nlm.nih.gov). Any remaining unidentified hits were clustered at 80% identity using MMSeqs2 v15 tool[110] and the representatives for each cluster where then analyzed using the HHpred online server (https://toolkit.tuebingen.mpg.de/tools/hhpred) against the Pfam-A_v37 and PDB_mmCIF30 databases. Finally, alignments with reference sequences were generated using the *mafft-linsi* option from MAFFT v7.520 tool[111] and viral phylogenies computed with FastTree v2.1.11.[112] Fragment consensus sequences were then generated using the ClAlign v1.1.4[113] tool with the *majority_nongap* consensus option.

Mapping

Ancient RNA

We sought to determine the best genome-wide alignment strategy for woolly mammoth aRNA sequencing data prior to further downstream analyses. To do so, we selected the mammoth 1 sample for mapping. This is based on its outperforming behavior in initial metagenomics and metatranscriptomics analyses, as well as in its high endogenous woolly mammoth DNA content and negligible source of human DNA contamination detected (Figure 1B; Table S1, #3 and #4). PCR-deduplicated and trimmed aRNA sequences from mammoth 1 (Yuka) were then mapped to a concatenation of the mEleMax1 genome assembly plus the woolly mammoth mitogenome (DQ188829) using three different alignment algorithms: i) Bowtie v1.3.0[34] with parameters similar to the ones used for historical RNA sequences[22] (i.e. allowing up to two mismatches within a seed equal to 18 nt, and reporting a maximum of one valid alignment with high sensitivity, *-n 2 -l 18 -k 1 -y --best*). ii) Bowtie2 v2.4.2[32] with *–end-to-end* and *–sensitive* parameters, as used for ancient RNA sequence alignment of an end-Pleistocene canid.[18] This was intended to guarantee a complete mapping of each aRNA sequence from start to end, similar to what is implemented in Bowtie, thus disabling soft clipping or partial alignments as set in the *–local* mode. And iii) BWA v0.7.17[33] with the *aln* algorithm and the same parameters as used for aDNA sequences (see below). Sequence mismatch per each base position and the number of mapped sequences for each of the three alignment algorithms were estimated with the AMBER tool[31] (https://github.com/tvandervalk/AMBER). Comparative alignment results are shown in Figures S2A and S2B. Based on these results, we decided to implement the Bowtie2-based approach for subsequent genome-wide and transcriptome-wide analyses. Genome-wide alignment was performed with a composite genome built by concatenating the mEleMax1 Asian elephant assembly plus the DQ188829 woolly mammoth mitogenome using Bowtie2 as described before for alignment strategies comparison. Transcriptome-wide alignment was performed with the same alignment approach but avoiding reverse-complement mapping with the *norc* option. The Asian elephant transcriptome reference (i.e. intronless transcript sequences with concatenated exonic regions) was built using the gffread v0.12.6 tool within the GFF Utilities software[114] (*-Z -W --force-exons --gene2exon --t-adopt --tlf*). Exonic, intronic, intergenic, exon-exon and exon-intron maps were then deduced from genome-wide and transcriptome-wide gene annotations (GCF_024166365.1, dated 02/26/2023). The following transcript types were included in a merged transcriptome and used as reference for alignment: protein-coding mRNAs, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), long-noncoding RNAs (lncRNAs), and microRNAs. Whenever a given gene locus had more than one transcript isoform annotated, only the longest one per each corresponding locus was considered and included in the transcriptome reference. MicroRNA loci were generated by extending the annotated precursor sequence by 30 nt upstream and downstream. Trimmed aRNA sequences including UMI information were first mapped and the resulting alignment files were sorted and converted to binary files with the Sambamba v0.8.0 tool.[115] UMI-deduplication was then implemented with the *dedup* function from UMI-Tools v1.1.2 software,[116] which by defaults employs a directional methodology based on Hamming distance thresholding for UMI clustering (https://umi-tools.readthedocs.io/en/latest/the_methods.html). aRNA sequences mapped to the transcriptome that were also found aligned to either mitochondrial, intronic or intergenic regions after genome-wide alignment were removed from transcriptome-wide mapping files. aRNA alignment statistics for each woolly mammoth sample are available at Table S2, #1. The same pipeline was applied when mapping aRNA sequences from mammoth 1 (Yuka) and mammoth 10 (Chris Waddle) to the reference-assisted 3D mammoth genome[23] and transcriptome. Comparative alignment efficacy between the mEleMax1 assembly and the mammoth genome was assessed using AMBER[31] with mapping data generated by Bowtie2 applied to aRNA sequencing data from mammoth 1 (Yuka, Figures S2A and S2B; Table S2, #3).

Ancient DNA

We used the GenErode v0.4 pipeline[94] with default parameters for ancient/historical DNA processing. In brief, paired-end sequences were merged, and adapters were trimmed with fastp v0.22.[93] Only merged fragments with a minimum length of 30 nt were retained for alignment against a reference assembly consisting of the Asian elephant genome (GenBank: mEleMax1), the human genome (GenBank: hg19) and the woolly mammoth mitogenome[35] (Krause mammoth, GenBank: DQ188829). This concatenated mixed reference was implemented to allow competitive mapping of DNA sequences among human and elephant genome assemblies and reduce human-derived contaminant DNA aligned to highly conserved genomic regions among both species.[117] Alignment was performed using BWA v0.7.17[33] with the backtrack mode (i.e. *aln*) and parameters adjusted for ancient DNA[30] to disable seeding (*-l 16500*), allowing for additional substitutions (*-n 0.01*), and up to two gaps (*-o 2*). Subsequently, the BWA *samse* algorithm and SAMtools v.1.8[118] were implemented to generate sorted bam files. Finally, the GATK v3.8-0[119] suite was used to realign indels and PCR duplicates were removed using a custom script that marks duplicates based on both the start- and end-coordinates of the aligned DNA fragments.[30] Only DNA sequences with mapping quality higher or equal to 30 (MQ$\geq$30) were considered as genuine hits to the concatenated reference. The fraction of DNA sequences mapped to either the mEleMax1 genomic assembly or DQ188829 mitogenome were identified and used for calculating mapping statistics for each woolly mammoth sample. The human contamination and endogenous woolly mammoth fraction were then assessed based on the number of sequences mapped to each assembly included in the competitive mapping. A summary of DNA sequencing data alignment statistics can be found in Table S1, #4.

Exon-exon and exon-intron junction overlaps

aRNA sequences spanning exon-exon (transcriptome-wide) and exon-intron (genome-wide) junctions were classified as such only if they spanned the annotated junction over at least 4 nucleotides from their 5' and 3' ends. The relationship between the number of exon-exon spanning sequences for each transcript and the total number of mapped aRNAs was assessed using Pearson's linear correlation. Significant differences between the total number of exon-intron spanning sequences using aRNA and aDNA data from mammoths 1, 4, and 10 were estimated with a one-tailed Fisher's exact test using the *fisher.test* function in R v4.4.2 (Table S3, #5).

Exonic enrichment

The number of aRNA sequences ($\geq$23 nucleotides) mapped to exonic regions from the longest transcript isoform for each gene for mammoths 1, 4, and 10 was compared with aDNA sequences mapped overlapping the same exonic regions. To allow a fair comparison, aDNA sequences were randomly subsampled to match the number of aRNA sequences mapped to the mEleMax1 assembly genome-wide after UMI-deduplication. We then computed the difference between exon-overlapping aRNA and aDNA sequences as a fold-change value, using aDNA data as baseline. This means that any increment in the number of exonic aRNA sequences compared to aDNA sequences would be represented by a positive fold-change, and vice versa. The same analysis was performed considering intronic regions.

Genome-wide expression hotspots

We divided the Asian elephant mEleMax1 assembly into non-overlapping consecutive bins of 0.5 Megabase pairs and quantified how many mapped aRNA sequences mapped to each of them. Only chromosome-level scaffolds were considered for these analyses, which were performed for mammoths 1, 4, and 10. In order to compare the genome-wide sequence distribution between aRNA and aDNA data, we retrieved deep aDNA sequencing datasets from the same woolly mammoth specimens as reported elsewhere.[23,29,30] Mapping for aDNA and aRNA was performed as described above. To allow a fair comparison, aDNA sequencing data were randomly downsized to match the sequencing depth of their corresponding aRNA datasets once UMI-deduplicated. Random subsampling of aDNA sequences was performed with the *-s* option of SAMtools v.1.8[118] software and defining the corresponding fraction based on the ratio between aDNA and aRNA sequences mapped once deduplicated. aDNA sequences overlapping each genome bin were then quantified and compared against mapped aRNAs. Upon alignment, aDNA sequences are expected to map in a roughly even fashion across the genome, while aRNA sequences would be concentrated on exonic regions. Any over-representation of aRNAs mapped within a given genomic bin would then be indicative of the presence of a coding loci being expressed. In this way, highly abundant RNA types, such as rRNAs, are easily identified by observing genome-wide bins with highly abundant aRNA sequences mapped.

DNA contamination and amplification bias

In order to determine whether DNA carryover contamination might be present in our aRNA extracts and therefore inadvertently amplified and sequenced, we used a computational approach similar to what was implemented for testing DNA contamination in our historical RNA dataset from a Tasmanian tiger.[22] Briefly, we quantified the aRNA sequences mapped to each non-overlapping consecutive bin of the Asian elephant (mEleMax1) assembly as described before, as well as the corresponding coverage per bin for mammoths 1, 4 and 10. We then used aDNA sequencing data from each mammoth generated in this study (i.e. obtained from the same tissue piece but extracted and sequenced using different methodologies) to compare the number of aDNA sequences mapped and the depth of coverage obtained per bin. To allow a fair comparison, we randomly downsized the aDNA data to match each corresponding mammoth sample as described above. Additionally, we accounted for the difference between sequence length of aDNA

and aRNA data affecting the obtained coverage per bin. In this way, we divided the obtained aDNA breadth of coverage per bin by a factor determined as the ratio between the average length of aDNA and aRNA sequences mapped once harmonized by depth of coverage. aRNA sequences mapped to bins with a difference $\leq$ two-fold in the number of mapped aRNA compared to aDNA sequences (depth of coverage), and/or those with a difference $\leq$ two-fold in their normalized breadth of coverage were classified as a potential source of DNA contamination. The overall putative percentage of DNA contamination in our aRNA data for each analyzed mammoth was then calculated from the number of aRNA sequences mapped to all bins flagged as potentially DNA-contaminated divided by the total number of aRNAs mapped genome-wide.

We also sought to determine if PCR amplification biased sequence selection against GC-rich regions, therefore introducing unwanted artifacts in transcript abundance estimates genome-wide. We compared GC content (%) for each transcript analyzed against the total number of aRNA reads mapped, the corresponding coverage obtained from the intronless longest transcript isoform per gene and the linear correlation (Pearson's *r* coefficient) between exonic regions supported by mapped aRNAs and those devoid of any aRNA sequence aligned (Figures S3D–S3F). To do so, we obtained the sequences of each protein-coding and noncoding transcript (including the longest transcript per each protein-coding, rRNA, tRNA, snRNA, snoRNA, lncRNA and microRNA gene) supported by mapped aRNAs and those where no sequences were mapped, separately. Finally, we calculated the GC content (%) of each mapped and unmapped fraction per transcript by using the *comp* tool from Seqtk v1.3-r106 (https://github.com/lh3/seqtk).

Sexing

Ancient RNA

aRNA sequences concordantly mapped to protein-coding loci located to the Y chromosome of mEleMax1 assembly were retrieved. Only those aligned to the same gene at both genome-wide and transcriptome-wide level were considered as evidence of Y-related gene expression in mammoth 1 (Yuka, Table S3, #1).

Ancient DNA

To determine the genomic sex of each of the ten woolly mammoths analyzed in this study, we used a coverage-based sexing method following Pečnerová et al.[120] Briefly, we compared the depth of coverage from chromosome X to the average autosomal depth of coverage, as computed by the GenErode v0.4 pipeline.[94] In this way, females (XX) will show a similar coverage to the chromosome X compared to autosomes, given a diploid genome such as the Asian elephant assembly. On the contrary, males (XY) will have roughly half the coverage for chromosome X compared to autosomes, since they harbor only one X copy. Finally, we calculated the ratio between chromosome X and autosomes depth of coverage. A ratio between 1 and 0.8 was considered as female (XX), while a ratio between 0.4 and 0.6 was considered as male (XY). These analyses revealed a discordant sexing for mammoth 1 (Yuka), with a profile compatible with being an XY male, while phenotypical sexing reported it to have female genital structures in previous examinations.[26] We then investigated the presence of Y-related genes in the Yuka genome. aDNA sequences mapping to the mEleMax1 chromosome were retrieved, and those overlapping with the genomic coordinates of the sex-determining region Y (*SRY*) gene were further analyzed. DNA sequencing data from two additional male woolly mammoths (L386 and L414)[82] and one male Asian elephant (SAMN28571076, Table S3, #2) were used for comparison. We then computed the sequence coverage for each base pair of the *SRY* gene using the *coverage* function from BEDtools v2.31.0.[121] These values were then normalized by the total number of aDNA sequences mapped to the *SRY* locus, standardized and linearly rescaled such as $Zi \in [0,1]$ using the *rescale* function from the scales v1.3.0 R package. Single-nucleotide variants (SNVs) present in woolly mammoths compared to the Asian elephant genome within the *SRY* gene were also identified with the BCFtools v1.8[122] *mpileup* tool, allowing a minimum mapping quality and base quality of 30 (*-Q 30 -q 30*), and using the *call* function with a consensus model (*-c*). Biallelic homozygous SNVs for the alternative allele supported by at least three aDNA sequences were kept (Table S3, #4).

Damage analyses

The PMDtools v0.60[123] software was used to identify C>U deamination events in both aRNA and aDNA sequencing data. For aRNA, UMI-deduplicated sequences mapped to the mEleMax1 transcriptome were used, allowing a range of 65 base pairs for deamination inference (*–platypus –range 65*). Only C>U deaminated sites, read as C>T by the sequencer, were considered. Damage analyses focused on microRNA loci (Figure S4C) were performed using Bowtie[34] as alignment tool with specifications tailored to small RNA sequences (*-n 2 -l 18 -k 1 -y –best*). For aDNA data, PCR-deduplicated and indel-realigned sequences mapped to the mEleMax1 assembly were used. Since aDNA extracts were USER-treated for damage correction,[89] we restricted the identification of deamination events to CpG islands within the first 30 nucleotides of any given sequence and a base quality of at least 30 (*–platypus –range 30 –CpG –requirebaseseq 30*).

Genome-wide variant calling

We used RepeatModeler v2.0.1[124] and RepeatMasker v4.0.9[125] tools, as specified in the GenErode v0.4 pipeline,[94] in order to identify and mask repetitive regions within the mEleMax1 assembly, since these can negatively impact variant calling due to sequence misalignment. Subsequently, BCFtools v1.8[122] *mpilepup* (*-Q 30 -q 30*) and *call* (*-c -M*) functions were used for variant identification on a set of 20 selected woolly mammoths[81,82] (*M. primigenius*, Table S3, #2) Indels and SNVs within five base pairs around indels were discarded. SNVs within CpG islands were also excluded, as these regions can be protected from USER treatment and thus accumulate post-mortem damage.[89] Only biallelic SNVs supported by at least five aDNA sequences and no missing data within the

20 analyzed mammoths were kept. Variants located within the Y chromosome were also excluded for these analyses. The same pipeline was implemented for variant calling only on deep aDNA sequencing data from the Yuka mammoth[29] but keeping Y-located SNVs. These analyses resulted in a total of 1,939,200 and 16,889,772 sites identified in the 20 woolly mammoths and the Yuka specimen, respectively. aRNA sequences from mammoth 1 (Yuka) mapped to mEleMax1 assembly after UMI-deduplication and overlapping any of the two variant datasets were then identified with the *intersect* function from BEDtools v2.31.0.[121] SNVs matching deamination profiles (C>T) were flagged as putatively originating from nucleotide damage as opposed to true variant sites (Table S4, #1).

Variant effect prediction

The functional impact of SNVs identified in mammoth 1 (Yuka) genome was inferred by using the Variant Effect Predictor (VEP) v113.0 tool[126] jointly with mEleMax1 gene annotation (GCF_024166365.1). The predicted effects were classified as putative high, moderate or low impact. High impact variants included SNVs causing frameshifts, splice acceptor/donor sites, start/stop loss and stop gained sites. Moderate impact SNVs included those causing missense variants and in-frame deletions/insertions. Low impact SNVs included sites predicted as synonymous variants, splice regions, intron variants, UTR variants, and stop retained variants. The highest potential impact was assigned to variants with multiple predicted impact effects. aRNA sequences from the Yuka mammoth mapped to protein-coding loci and harboring SNVs were retrieved with the strand-aware *intersect* function (-s) of BEDtools v2.31.0[121] (Table S4, #2).

MicroRNA annotation

We used the reference elongated (± 30 nt) precursor microRNA complement for African elephants (*Loxodonta africana*; GCA_000001905.1, LoxAfr3) available at MirGeneDB3.0[127] to remap and annotate them toward the mEleMax1 Asian elephant assembly (Table S4, #3). Mapping was performed with the Bowtie2 v2.4.2 tool with *–end-to-end* and *–sensitive* parameters.

MicroRNA variant analysis

We identified the presence of an A>G variant site located toward the end of the 3' (3p) mature arm of Mir-1-P3 of Asian elephants supported by mapped aRNA sequences from the Yuka mammoth. An additional A>G variant was also found in the opposite site of the microRNA hairpin at the beginning of the 5' (5p) mature arm of Mir-1-P3. To investigate its potential as a species-specific mutation across mammals, we retrieved a total of 60 selected genomes for which microRNA complements were annotated at sufficient resolution. The Asian elephant Mir-1-P3 precursor sequence was then interrogated against each of the selected mammalian assemblies (Table S4, #7) with the BLAT algorithm[128] embedded within Ensembl v113 database and the *blastn* tool from NCBI BLAST v2.16.0 webserver (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the *refseq_genomes* database. The genomic region obtaining the highest score and more significant E-value for each mammalian species was then selected as the Asian elephant Mir-1-P3 homologous sequence. A phylogenetic tree depicting the segregation of the G variant site as opposed to the putative ancestral A site was created with the rotl v3.1.0 R package (https://CRAN.R-project.org/package=rotl) and visualized with ggtree v3.12.0[129] (Figure 4B). Additionally, we used the Mir-1-P3-3p sequence as input for the Rfam v15.0 database.[130] This was intended to verify the existence of the G variant site within Mir-1-P3-3p in any annotated RNA sequence across multiple organisms beyond mammalian genomes (Figure S4), since the Rfam database includes a comprehensive collection of all known noncoding RNA families ever sequenced and annotated.

Novel microRNA prediction

We implemented the miRDeep2[48] software in an attempt to identify novel microRNA loci in the reference-assisted 3D woolly mammoth assembly[23] and showing evidence of expression in our aRNA sequencing data. To do so, we used the whole set of aRNA sequences mapped to the woolly mammoth genome from Yuka (mammoth 1), since it was the specimen showing most promising microRNA abundance profiles (Table S4, #4). The mapper.pl script was first run to align aRNA sequences to the mammoth assembly allowing one mismatch in the seed (*-q*). The resulting .arf file was then used as input of the miRDeep2.pl script including the annotated mature and precursor microRNAs in the mammoth assembly as reference, as well as the annotated Asian elephant mature microRNAs (Table S4, #3) as closely-related species comparison. Novel candidates were then manually selected from the miRDeep2 output by visually inspecting the hairpin structure and supporting aRNA sequences mapped (Table S5, #2). We then identified homologous sequences of the selected novel candidates in 60 mammalian species, including the woolly mammoth, using the BLAT algorithm[128] tailored for searching distant homologies within Ensembl v113 database, and the *blastn* tool from NCBI BLAST v2.16.0 webserver (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the *refseq_genomes* database. Folding of the hairpin structure was predicted with the Mfold[131] software using default settings.

Transcript quantification

The breadth of coverage and transcript abundance supported by aRNA sequences mapped to each longest transcript annotated per gene (including protein-coding and noncoding loci) was estimated with the *coverage* function from BEDtools v2.31.0.[121] MicroRNA abundance was calculated with the strand-aware *intersect* function (*-s*) from BEDtools by only retaining mapped aRNA sequences overlapping either the 5p or 3p mature region of each microRNA precursor transcript in at least 75% of the sequence with a reciprocal

mode (*-f 0.75 -r*). Additionally, we computed an abundance index value for each transcript to account for transcript length biases in the alignment of aRNA sequences. In this way, we applied the following formula:

$$X_i = (1 / L + B_i) \times C_i$$

Where *Li* is the transcript length, *Bi* is the breadth of coverage and *Ci* is the aRNA sequence count associated to each *i* transcript considered in either the protein-coding or the noncoding fraction. Then, the obtained values were linearly rescaled such as $Xi \in [0,1]$. This was applied by using the *rescale* function from scales v1.3.0 R package, which computes:

$$Z_i = X_i / X_{max}, \text{where } 0 \leq Z_i \leq 1$$

This normalization process allows all values to be comparable on a standard scale.

GO term enrichment

Gene ontology (GO) enrichment analysis was performed using the set of protein-coding genes detected as confidently supported (breadth of coverage $\geq$5%) by aRNA sequences from mammoth 1 (Yuka) mapped to their longest annotated transcripts. Riboproteins and mitochondrial mRNA genes were excluded from the query gene set. We used the ShinyGO v0.81 webserver tool[132] with default parameters except for a minimum pathway size of 10. The human reference was used for GO annotation, and the background dataset was defined as all protein-coding transcripts with at least one aRNA sequence mapped in the Yuka dataset. The GO Biological Process category was selected as reference for enrichment analysis.

Tissue clustering

After quantification of both the protein-coding and noncoding fraction, we checked whether the transcriptional complement detected in mammoths 1, 4, and 10 resembled modern tissue-specific expression patterns according to their supposed tissue identity (Figure 1A). As modern reference comparison for protein-coding genes, we retrieved a comprehensive RNA-seq atlas from healthy human tissues[133,134] by using the *getAtlasExperiment* function from ExpressionAtlas v1.32.0[135] R package (E-MTAB-1733, E-MTAB-2836), including a compendium of sequenced transcriptomes from a range of 14 tissues, plus the addition of 3 chondrocyte (cartilage) samples as reported by Zhu et al.[136] A list of RNA-seq data samples used in this study is shown in Table S5, #6. RNA-seq data from modern elephant tissues[137,138] were also retrieved as a close-relative to woolly mammoths (Table S5, #6). For the noncoding fraction, we made use of available microRNA-seq datasets from human healthy tissues,[139–142] and elephant blood[52] (Table S5, #6). Additionally, ancient RNA expression profiles of skeletal muscle and skin tissue from the extinct Tasmanian tiger,[22] as well as from an end-Pleistocene canid and historical wolf skins[18] were also included in our analyses. All expression profiles from annotated protein-coding and microRNA genes in the mEleMax1 assembly were included for tissue clustering, irrespective of their abundance in modern and/or ancient samples. Protein-coding mRNA and microRNA expression profiles were analyzed separately. Raw count values were normalized across all included samples with the trimmed mean of M-values (TMM) method[143] using edgeR v4.2.1,[144] and $\log_2$ transformed after adding a pseudo-count of 1. Sample hierarchical clustering was first implemented with the *hclust* function based on Euclidean distance. To do this, we selected the normalized values of the top 30 most abundant protein-coding and microRNA loci detected in the aRNA sequencing data of mammoth 1 (Yuka). Protein-coding mRNAs and microRNAs were analyzed separately. We then merged these with the averaged expression values of the corresponding top 30 protein-coding and microRNA genes from each considered human reference tissue (Table S5, #6). The resulting dendrograms where then reordered with the *dendsort* function from dendsort v0.3.4 R package in order to rearrange branches based on tissue samples and genes so that the clusters with more divergent gene expression profiles are placed more distant in the hierarchical sorting (Figures 5C and 5D). An additional clustering of human reference tissue samples was also implemented by using the uniform manifold approximation and projection (UMAP) algorithm[145] with the *umap* function from umap v0.2.10.0 R package (https://github.com/tkonopka/umap). UMAP initialization and embedding was computed for both protein-coding and microRNA expression analyses independently, and defined with the following parameters: *n_neighbors=10, metric=''pearson'', spread=10, random_state=40*. Modern elephant samples, as well as historical and ancient aRNA sequencing data from mammoths 1, 4 and 10 generated in this study, Tasmanian tiger[22] and end-Pleistocene canid[18] tissues were then projected onto the previously learned UMAP embedding using the *predict* function in R v4.4.2.

Annotation of the woolly mammoth genome

We generated a comprehensive gene annotation for the reference-assisted 3D mammoth assembly reported by Sandoval-Velasco et al.[23] encompassing protein-coding and noncoding loci annotated in the Asian elephant mEleMax1 assembly (GCF_024166365.1). Among the noncoding loci, we included rRNAs, tRNAs, snRNAs, snoRNAs, lncRNAs and microRNAs. Gene lift-over was performed using the Liftoff v1.6.3 tool[146] and allowing a flanking extended region equal to half the length of each annotated locus to lift over (-flank 0.5) while applying the following parameters within the Minimap2 v2.24-r1122[147] aligner module: i) promoting end-to-end alignment between assemblies (*–end-bonus 100*), and ii) allowing a more flexible secondary-to-primary alignment score ratio to retain additional secondary alignments up to 50 (*-N 50 -p 0.5*).

Data visualization

Additional software tools used for visualization not mentioned before and run within R v4.1.3, v4.3.2, and v4.4.2 include the following R packages: ggplot2 v3.5.1, cowplot v1.1.3, gtools v3.9.5, ggpubr v0.6.0, tidyverse v2.0.0, ggh4x v0.2.8, ggstar 1.0.4, and showtext v0.9-7.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis for GO enrichment was performed as reported by the ShinyGO v0.81[132] tool applying a hypergeometrical test followed by multiple testing correction with the false discovery rate (FDR) method. Statistical analysis for testing the presence of a significant enrichment of the total number of exon-intron spanning sequences using aRNA and aDNA data from mammoths 1, 4, and 10 was performed using a one-tailed Fisher's exact test with the *fisher.test* function in R v4.4.2. The resulting odds ratio and *P*-value are reported in Table S3, #5. aDNA sequencing data was randomly subsampled to match the depth of coverage of aRNA data for the same mammoth samples using the *-s* option of SAMtools v.1.8.[118] aRNA and aDNA sequence quantification, aDNA contamination estimates and coverage metrics were individually computed for each of the ten woolly mammoths analyzed in this study whenever specified above in method details using R v4.4.2 and custom scripts.

ADDITIONAL RESOURCES

Gene annotation for the mammoth reference assembly and code to reproduce the needed files to generate accessory files for gene abundance estimation using aRNA sequencing data are available at https://github.com/emarmolsanchez/aRNA/tree/main/Annotation. Metatranscriptomics and metagenomics reports generated using the Pavian v1.2.0 R package[98] are available at https://github.com/emarmolsanchez/aRNA/tree/main/Meta. Scripts and input data needed for reproducing all main and supplemental figures depicted in this study are available at https://github.com/emarmolsanchez/aRNA/tree/main/Figure_RCode and https://github.com/emarmolsanchez/aRNA/tree/main/Figure_Tables, respectively.

Supplemental figures

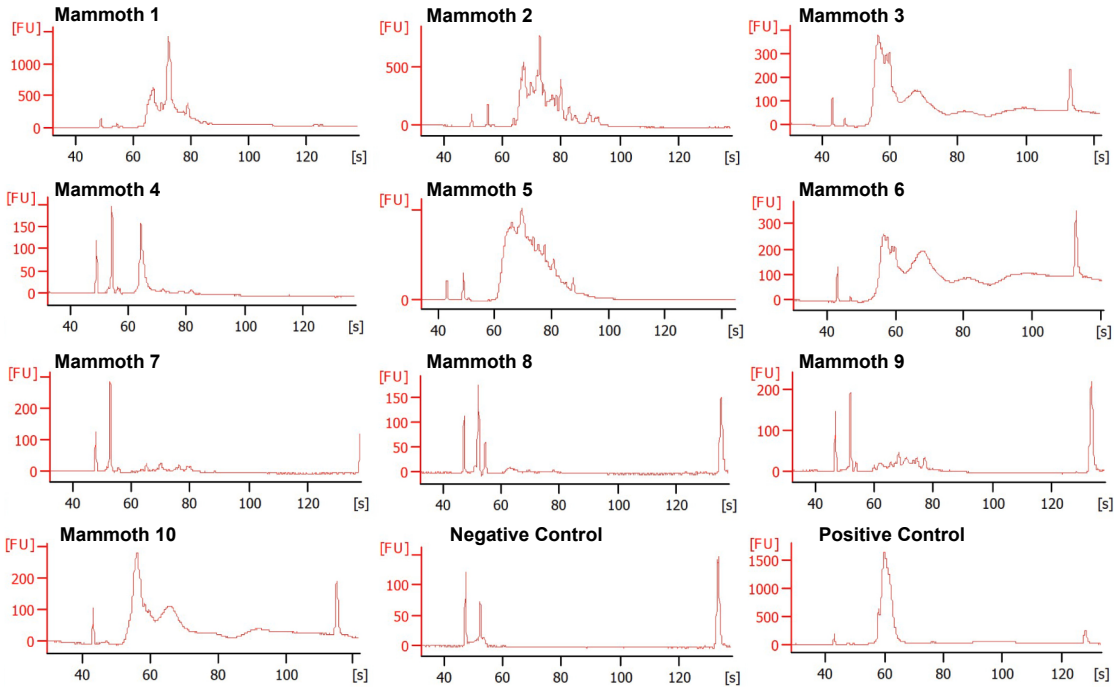


Figure S1. Distribution of cDNA library fragment sizes in bp, related to STAR Methods

Sequencing libraries are prepared using the NEXTFLEX Small RNA-seq kit v.3 protocol.

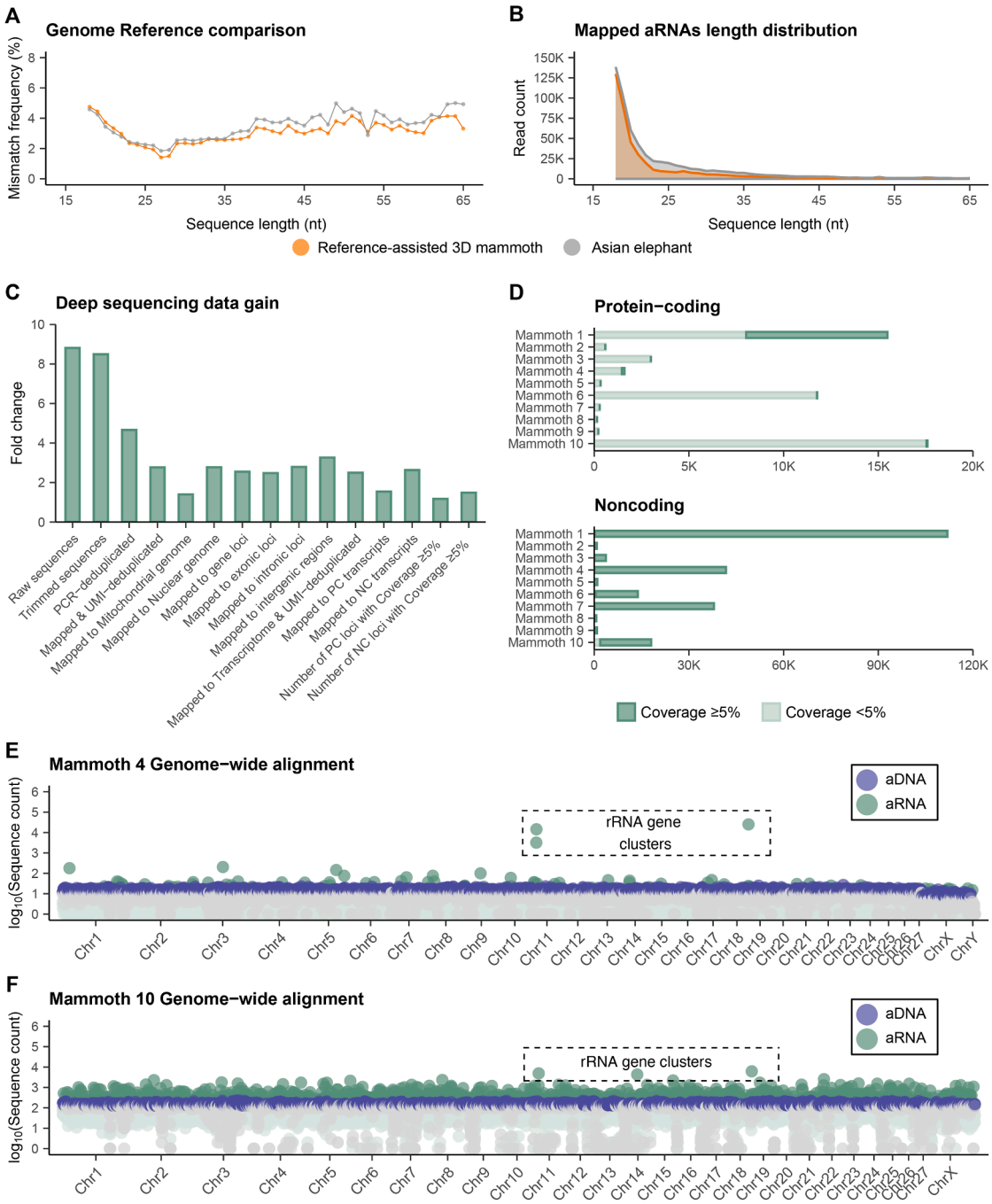


Figure S2. Analysis of mapping and sequencing strategies, and genome-wide mapping patterns, related to Figure 1 and STAR Methods

(A) Comparison of genome assemblies for mapping aRNA sequencing data based on mismatch frequency across sequence length, using aRNAs mapped to the Asian elephant (mEleMax1) and the reference-assisted 3D mammoth assemblies.

(B) Comparison of genome assemblies for mapping aRNA sequencing data based on the number and sequence length of aRNAs mapped to the Asian elephant (mEleMax1) and the reference-assisted 3D mammoth assemblies.

(C) Alignment statistics for mammoth 1 (Yuka) comparing original aRNA sequencing with deep resequencing of the same library, expressed as fold change in raw, processed, and mapped reads.

(D) Number of aRNA reads from all 10 mammoths mapping to annotated genes in the Asian elephant genome, categorized by high (≥5%) and low (<5%) transcript breadth of coverage.

(E and F) Genome-wide mapping patterns of aRNA and aDNA sequences from (E) mammoth 4 and (F) mammoth 10. Each dot represents a non-consecutive 0.5 Mb pair genomic window. aRNA shows uneven distribution, unlike aDNA, highlighting expression hotspots at highly transcribed ribosomal RNAs and other coding and noncoding loci.

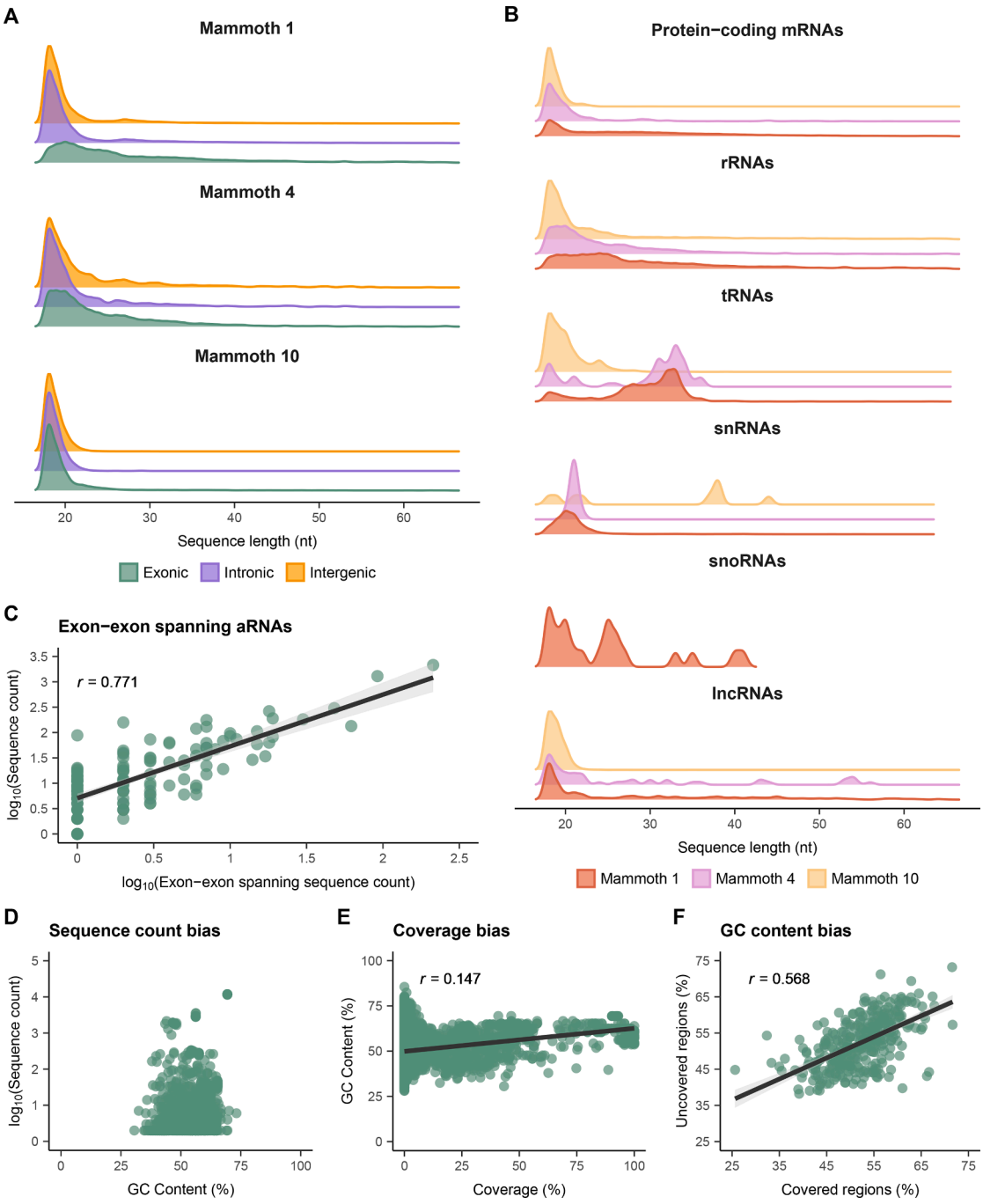


Figure S3. Sequence length distribution and analysis of DNA contamination and amplification bias of aRNA data, related to STAR Methods

(A) Length distribution of sequences mapped to exonic, intronic, and intergenic regions of the Asian elephant assembly (mEleMax1) and the woolly mammoth mitogenome (see STAR Methods).

(B) Length distribution of sequences mapped to protein-coding mRNAs and noncoding RNAs including rRNAs, tRNAs, snRNAs, small nucleolar RNAs (snoRNAs), and long-noncoding RNAs (lncRNAs) annotated in the Asian elephant (mEleMax1).

(C) Scatterplot depicting the correlation between the number of exon-exon-spanning aRNA sequences and the total number of sequences mapped to each corresponding transcript (breadth of coverage $\geq$ 5%) in the Asian elephant transcriptome (mEleMax1). A high linear correlation (Pearson's *r* coefficient = 0.771) is found among the two variables, suggesting that highly abundant transcripts do also concentrate the majority of exon-exon-spanning sequences.

(D) Highly abundant transcripts detected in the skeletal muscle of Yuka (mammoth 1) do not show bias toward higher or lower proportions of GC content (%) in their sequences.

(legend continued on next page)

(E) Highly covered transcripts detected in the skeletal muscle of Yuka (mammoth 1) do not show signs of amplification bias toward higher or lower proportions of GC content (%) in their corresponding sequences (Pearson's *r* coefficient = 0.147).

(F) Regions of reliably detected transcripts (breadth of coverage $\geq$ 5%; *N* = 349, including nuclear and mitochondrial protein-coding loci) covered by any mapped aRNA show similar GC content (%) composition compared with regions devoid of mapped aRNAs (Pearson's *r* coefficient = 0.568). This suggests that preservation of aRNAs through time is not dependent on sequence composition.

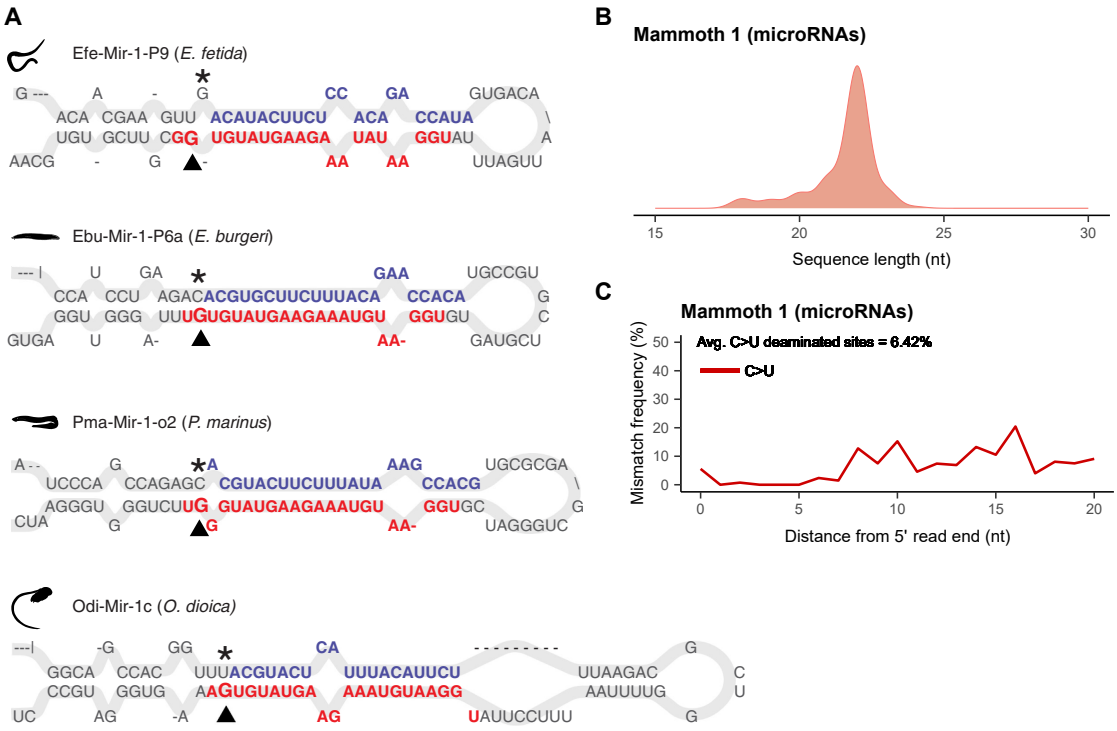


Figure S4. Cross-species SNV analysis and length distribution of aRNAs mapped to microRNA loci, related to Figure 4 and STAR Methods

(A) Redworm (*E. fetida*), hagfish (*E. burger*), sea lamprey (*P. marinus*), and oikopleura (*O. dioica*) Mir-1 hairpins harbor an A>G SNV (arrowhead) located toward the end of the mature 3p arm (red) as seen in proboscideans, tenrecs, and members of the Mustela genus (Figure 4C). The opposite site (*) located toward the beginning of the 5p arm (blue) forms a G-U variant combination in redworm and oikpleura, and G-C in hagfish and sea lamprey Mir-1 sequences, as opposed to the G-G variant combination found in woolly mammoths, extant elephants, and tenrecs (Figure 4C; Table S4, #7).

(B) Sequence length distribution of aRNAs mapped to microRNA loci annotated in the Asian elephant transcriptome (mEleMax1).

(C) Cytosine deamination damage analysis (C>U, read as C>T by the sequencer) based on aRNA sequences from mammoth 1 mapped to microRNA loci. To allow a better representation of deamination profiles in microRNA genes we used Bowtie as the alignment tool (see STAR Methods).

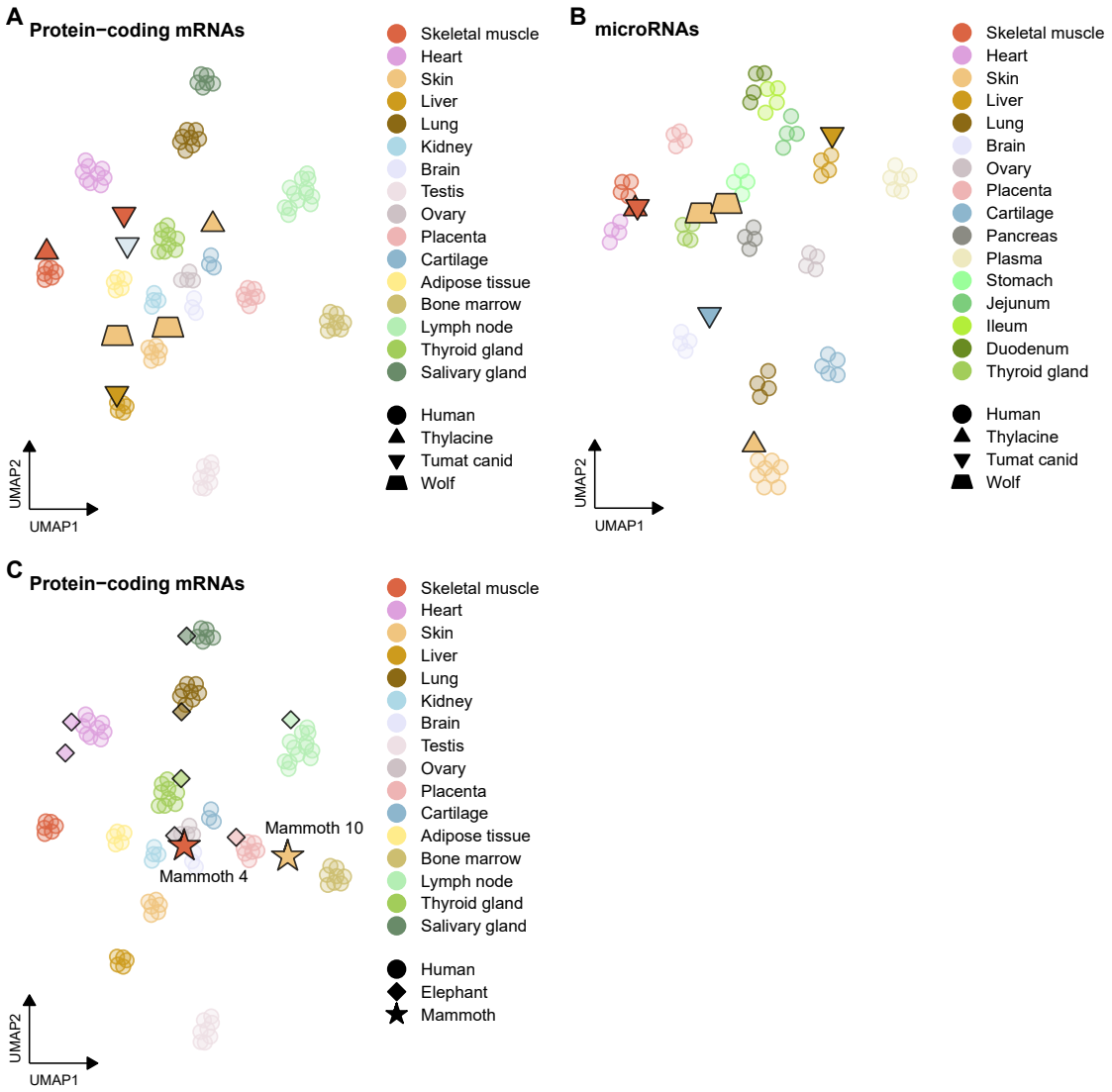


Figure S5. Historical and aRNA sequencing data show partially preserved tissue-specific transcriptomes, related to Figure 6 and STAR Methods

(A and B) Tissue-specific clustering of (A) protein-coding and (B) microRNA transcriptional profiles from historical Tasmanian tiger and ancient end-Pleistocene canid aRNA sequencing data.

(C) Tissue-specific clustering of protein-coding transcriptional profiles from mammoths 4 and 10 aRNA sequencing data. A uniform manifold approximation and projection (UMAP) approach was used for clustering a reference human tissue atlas. Historical and ancient aRNA abundance profiles were then projected onto the reference embedding.

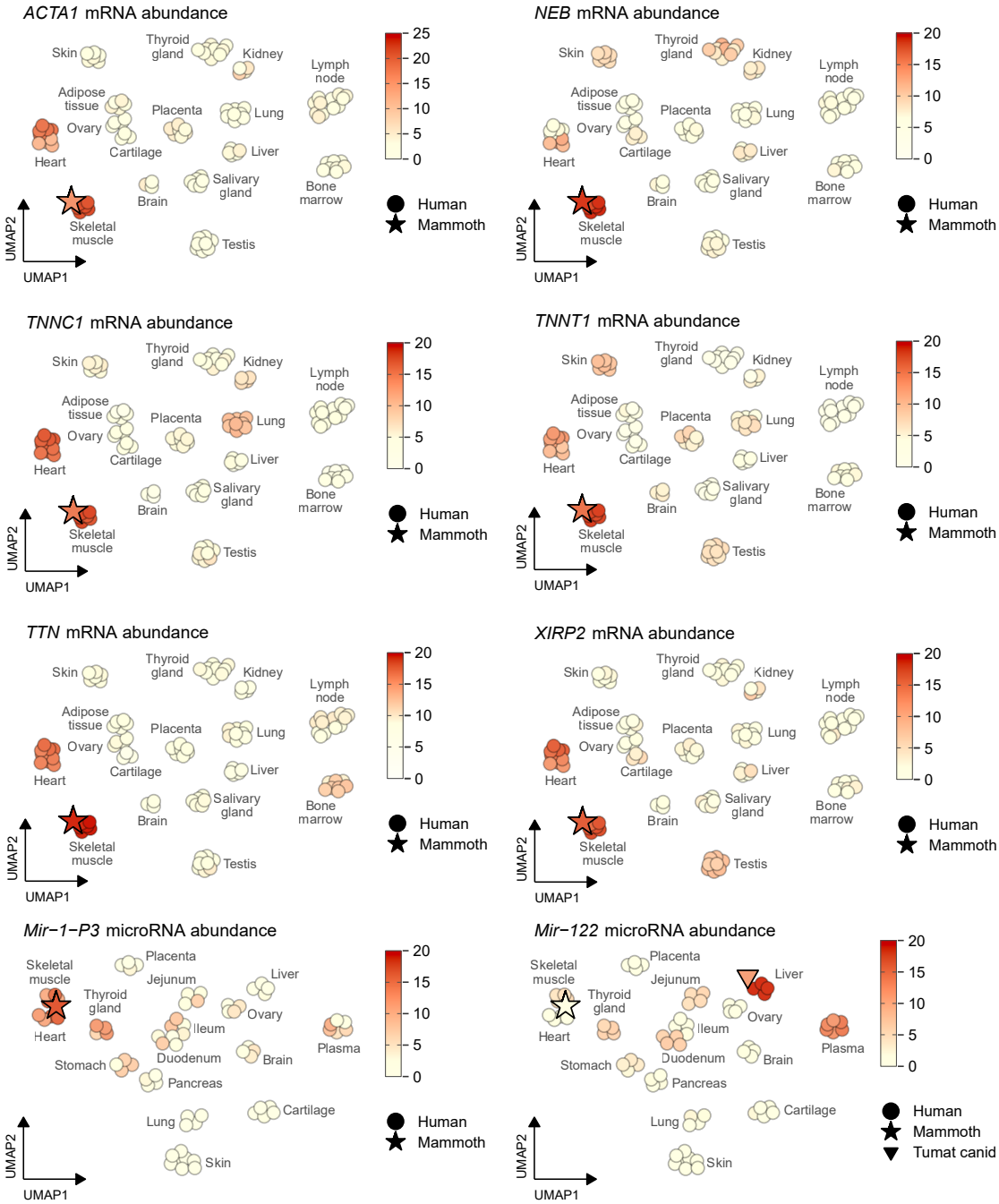


Figure S6. Yuka's muscle tissue shows muscle-specific transcriptional abundance, related to Figure 6 and STAR Methods

Tissue-specific clustering of selected abundant protein-coding mRNAs and microRNAs with breadth of coverage ≥ 5% from mammoth 1 (Yuka) aRNA sequencing data mapped to the Asian elephant transcriptome (mEleMax1). Transcript abundance is shown as a color gradient based on $\log_2$ normalized sequence count values. Transcript abundance normalization was implemented with the trimmed mean of M values approach (TMM, see STAR Methods).