A: Project title

Ancient RNA: uncovering the missing link to reveal a new era in Paleogenetics.

B: In a nutshell

- I propose there are plausible yet currently overlooked grounds supporting the extensive survival of RNA in natural history collections.
- I will explore this by applying recently developed methods for extracting and characterizing highly fragmented nucleic acids, thus facilitating the dawn of paleotranscriptomics.
- If RNA survival is as prevalent as I anticipate, the findings stand to initiate a new era in paleogenetics, with a holistic approach to the central dogma of molecular biology (from DNA to RNA and proteins). This has the potential to both transform our understanding of animal and associated pathogens evolution and contribute towards future de-extinction efforts.

C: Research idea and context

The analysis of **ancient DNA (aDNA) has gained considerable momentum over the past decade**, allowing the study of extinct or extant genomes from organisms that lived up to 2.5 million years ago [1]. This has enabled the reconstruction of genomes and historical ancestry of multiple extinct renowned species such as Neanderthals [2], woolly mammoths [3] or Tasmanian tigers [4].

Paleogenomics (aDNA) is now an established and active research field, and paleoproteomics is rapidly emerging [5]. **However, the characterization of animal historic/ancient RNA (aRNA) has essentially been ignored** (Fig. 1). I postulate the reason behind aRNA's oblivion is a widely accepted preconception that all molecular biologists learn in their basic training: that **RNA is an extremely fragile molecule**. This is based on the reasoning that RNA is rapidly degraded by RNAses, making it virtually inexistent shortly after cell death if not immediately fixed, and certainly not durable through long periods of time.

I believe this to be fairly reductionist (Fig. 1). A handful of studies have determined flash-frozen historic/ancient animal remains [6–8] and formalin-fixed paraffin-embedded (FFPE) tissues [9,10] as a promising source of aRNA. Other common preservation methods in museum collections – e.g. desiccation or liquid ethanol fixation – are still mainly unexplored, and hold similar ability to inactivate RNAses, thus prolonging RNA survival.

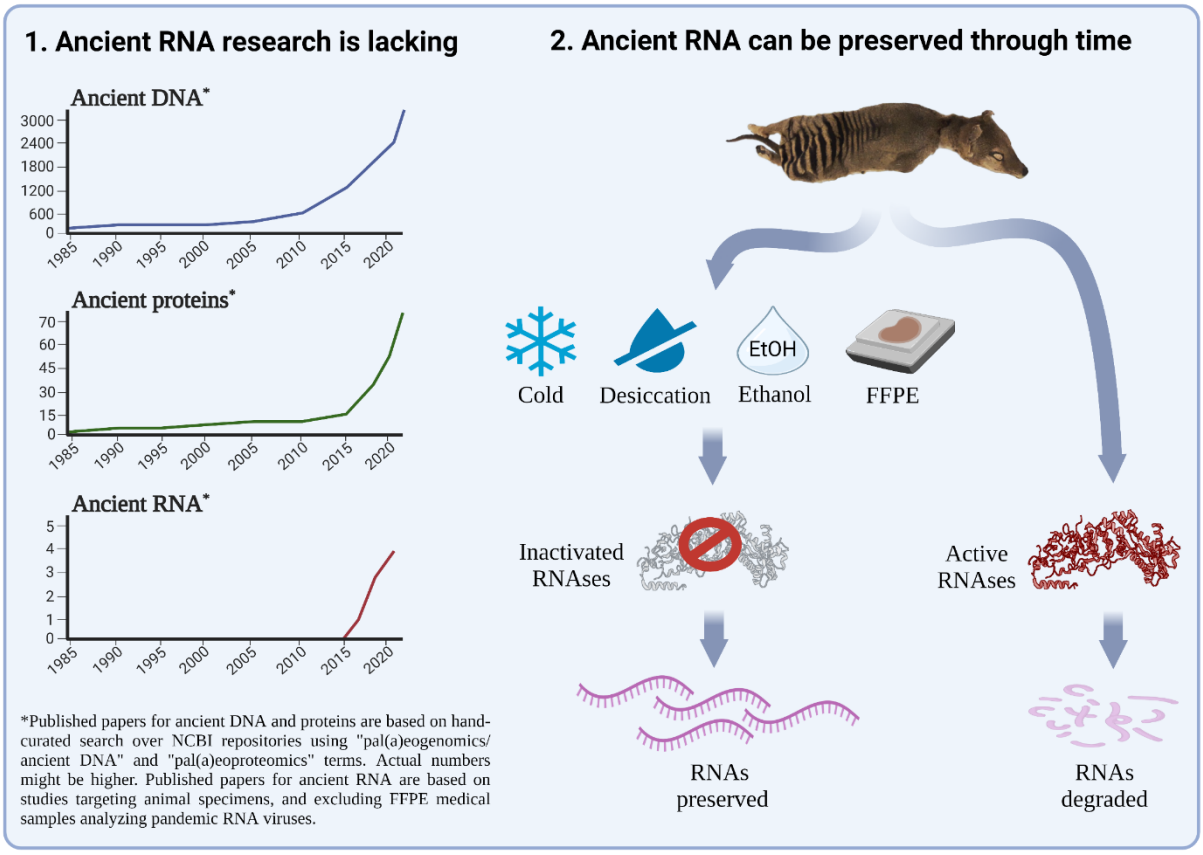

Figure 1: Rationale of the project

The key question is, **are these preservation methods generalizable conditions in which an animal's RNA molecule could survive?** I aim to answer this question by both a systematic investigation targeting differentially preserved animal tissues, combined with applying specific extraction chemistry tailored for isolating and sequencing likely fragmented and extremely low concentrated RNA molecules [1]. This has the potential to uncover numerous layers of information widely overlooked in paleo-studies so far, such as tissue-specific gene expression dynamics, non-coding regulatory mechanisms or transcriptional evidence of known and novel loci.

D: Proposed method of solution or concept

I propose to **develop the paleotranscriptomics field** by recovering, sequencing, and analyzing the preservation of aRNA molecules over time (Fig. 2), thus gaining novel methodological, biological, and evolutionary perspectives from long forgotten ancient transcriptomes.

First, I will determine a selection of different target tissues – i.e. skin, bones, muscle, and possibly other preserved soft organs such as brain, liver, intestinal tract, etc. – commonly found in dry and wet museum collections (at least 50 from each), prioritizing extinct and/or extant specimens with sequenced genomes. In a first phase (months 1-6), I will retrieve, sample, and extract total RNA isolates. As methodological

approaches for aRNA recovery are still underexplored, I propose to evaluate and customize diverse protocols adapted to each tissue and preservation variables.

Second, since I expect a high rate of fragmentation for the aRNA molecules recovered, I will implement a small RNA-targeted extraction procedure using microRNA-focused commercial kits for RNA isolation [7]. Sequencing libraries will be prepared with both total RNA extracts and after rRNA depletion for a better resolution, introducing unique molecular identifiers (UMIs) to enable accurate quantification of RNA profiles. Additionally, ancient DNA for each sample will be extracted with conventional methodologies to cross-compare results at the genomic and transcriptomic level.

Third, during months 7-24 of the 2-year project, I propose a two-phase framework:

1) Endogenous aRNA transcriptomes will be analyzed genome-wide at the coding (mRNA) and non-coding regions (rRNA, tRNA, microRNA, etc.), provided a reference genome is available, and compared with aDNA profiles for each tissue and specimen successfully sequenced (months 7-16). This enables a comprehensive characterization of transcriptional differences across tissues, something unachievable using aDNA.

2) Other sources of sequenced aRNA and aDNA from bacterial, viral and/or other eukaryotic contamination will be characterized using dedicated metagenomics and metatranscriptomics approaches [11] (months 17-24). Additionally, I aim to explore the temporal origin of the contaminant fraction based on nucleic acids damage patterns.

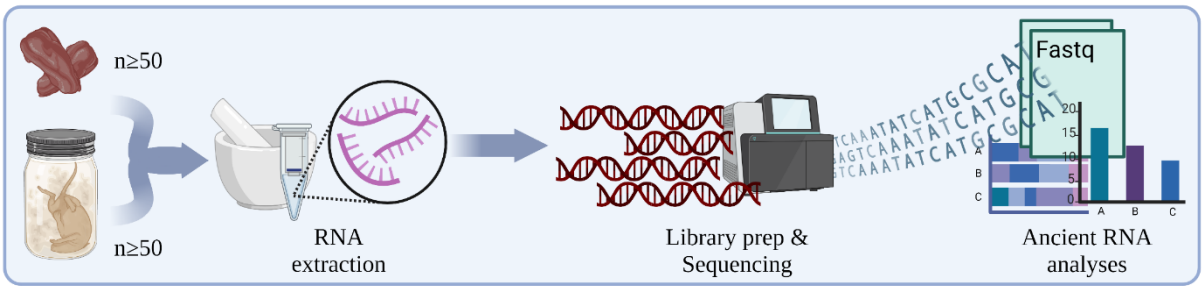

Figure 2: Methodological approach.

E: Major gains and obstacles

The expected gains are to successfully recover aRNA sequences **to study ancient transcriptomes from animal museum specimens**. Ancient RNA from prokaryotic or viral origin is also expected to be found, which will foster a highly overlooked field with scarce yet prominent instances of success [9,10,12]. This will allow the refinement of methodological techniques for aRNA analysis that are very much needed in the field. Besides, recovering transcriptional dynamics could provide novel insights into gene

regulatory mechanisms in extinct species, something that is currently missing, potentially boosting de-extinction efforts [13]. Failure to successfully recover animal aRNA would be a significant obstacle. However, endogenous aDNA retrieval is more probable, as well as the sequencing of modern or ancient metagenomes. Hence, this leaves sufficient room for reorienting analyses if required. Future developments could explore sedimentary aRNA sources without known origin, as already proven for aDNA [1].

F: Appropriateness

From a methodological perspective, a multi-species multi-tissue approach to the study of endogenous aRNA molecules in animals is completely lacking, with only few recent studies having demonstrated that ancient transcriptomes are able to survive in flash-frozen materials [6–8]. I here propose to explore the potential of museum collections to provide a comprehensive understanding of the perks and possibilities within the paleotranscriptomics research field. While risky, our approach has the potential to uncover the still missing link in Paleogenetics, ancient RNA, leading to integrative paleo-studies covering genomics, proteomics, and transcriptomics.

G: Probable objections

Despite recent promising yet limited results [6–8], I predict the main objection to the proposed study would be that this is doomed to failure, RNA does not last long enough to be found in old tissues. To this, I reply that since RNAses are the main reason why RNA molecules are quickly degraded after cell death, inactivating them with cold, desiccation or denaturing chemicals – e.g. ethanol solution – renders RNA a similar durability to that of DNA. Additional objections might arise, such as: why is there the need to analyze aRNA as opposed to the well-established aDNA? Is aRNA going to become a common finding or will it be restricted to scarce instances of extremely well-preserved specimens? I counter these with 1) aRNA can provide information about cell identity, non-coding regulatory mechanisms and transcriptional dynamics that aDNA, as being common for all cells, is not able to provide, thus it is of great interest to uncover this missing piece in paleogenetic studies. And 2) I believe that with appropriate methods still to be refined, a vast yet unexplored compendium of tissues preserved in dry and wet museum collections awaits further analysis to unravel novel insights for a new era in paleotranscriptomics.

H: References

1. Kjær KH et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature. 2022;612:283–91.

2. Green RE et al. A draft sequence of the Neandertal genome. Science. 2010;328:710–22.

3. Palkopoulou E et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr Biol. 2015;25:1395–400.

4. Feigin C et al. A Chromosome-Scale Hybrid Genome Assembly of the Extinct Tasmanian Tiger (*Thylacinus cynocephalus*). Genome Biol Evol. 2022;14:evac048.

5. Hendy J et al. A guide to ancient protein studies. Nat Ecol Evol. 2018;2:791–9.

6. Keller A et al. miRNAs in ancient tissue specimens of the Tyrolean Iceman. Mol Biol Evol. 2017;34:793-801.

7. Smith O et al. Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. PLoS Biol. 2019;17:e3000166.

8. Fromm B et al. 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. 2021;27:324–34.

9. Gryseels S et al. A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. Proc Natl Acad Sci U S A. 2020;117:12222-12229.

10. Patrono LV et al. Archival influenza virus genomes from Europe reveal genomic variability during the 1918 pandemic. Nat Commun. 2022;13:2314.

11. Breitwieser FP et al. KrakenUniq: confident and fast metagenomics classification using unique *k*-mer counts. Genome Biol. 2018;19:198.

12. Worobey M et al. 1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature. 2016;539:98–101.

13. Shapiro B. Pathways to de-extinction: how close can we get to resurrection of an extinct species? Funct Ecol. 2017;31:996–1002.