Recovery of genomic and transcriptomic profiles from decades-old FFPE brain tissues

Caroline Robinson Christiansen[1,2,3*], Emily Hansen Firoozfard[1,2*], Nikolay Oskolkov[4], M. Thomas P. Gilbert[2,5], Sarah S. T. Mak[2], Martin Wirenfeldt[6,7,8], Christina Kjær[1†], Emilio Mármol-Sánchez[2†]

[1]Department of Technology, Faculty of Health and Technology, University College Copenhagen, Copenhagen, Denmark. [2]Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen. [3]Department of Pathology, Rigshospitalet, The Capital Region of Denmark, Copenhagen, Denmark. [4]Metabolic Research Group, Latvian Institute of Organic Synthesis, Riga, Latvia. [5]University Museum, NTNU, Trondheim, Norway. [6]Department of Pathology, Esbjerg Hospital, University Hospital of Southern Denmark, Esbjerg, Denmark. [7]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark. [8]BRIDGE (Brain Research - Inter Disciplinary Guided Excellence), Department of Clinical Research, University of Southern Denmark, Odense, Denmark.

[*]Shared first authorship

[†]These authors contributed equally

***Corresponding authors:**

Emilio Mármol-Sánchez (emilio.marmol.sanchez@gmail.com)

Christina Kjær (chkj@kp.dk)

Abstract

Neurological, neurodegenerative, and psychiatric disorders impose substantial morbidity and disability worldwide, yet their molecular basis remains incompletely understood, in part due to limited access to human brain tissue. The Danish Brain Collection, comprising brains from individuals who lived in Danish psychiatric institutions from the 1940s to the 1980s, represents a unique but largely untapped resource for retrospective molecular investigation. Here, we assess the feasibility of extracting and sequencing DNA and RNA from decades-old FFPE brain tissue. We systematically evaluate how extraction and library preparation strategies influence nucleic acid yield and quality, and show that RNA end-repair prior to library preparation substantially enhances transcript diversity, improving data quality from highly degraded samples. Despite extensive fragmentation, we recover biologically informative transcriptomic profiles, including protein-coding and microRNA expression profiles that retain clear tissue specificity.

These results establish the Danish Brain Collection as a viable resource for genomic and transcriptomic analyses and demonstrate the broader potential of archival FFPE tissues for large-scale molecular studies.

Keywords: FFPE, brain, RNA, DNA, Danish Brain Collection, End repair.

Introduction

Neurological, neurodegenerative, and psychiatric disorders affect over three billion individuals worldwide, representing one of the most significant public health challenges of our time [1]. These conditions diminish the quality of life of affected individuals and impose enormous economic burdens on families and healthcare systems [2]. Yet our ability to diagnose, predict disease progression, and effectively treat these disorders remains limited. Next-generation sequencing of brain-derived DNA and RNA has begun illuminating the molecular mechanisms underlying conditions like schizophrenia and major depressive disorder [3], Alzheimer's disease

54 and related dementias [4], or epilepsy [5], revealing disease-associated genetic variants, disrupted
55 pathways, and potential therapeutic targets. Despite these advances, routine clinical protocols still
56 predominantly rely on symptom-based assessment, which carries inherent limitations: symptoms
57 overlap between disorders, early disease presents subtly, and subjective interpretation can lead to
58 diagnostic uncertainty [6]. This introduces the risk of misdiagnosis leading to inappropriate
59 treatment [7]. Moreover, most pathologies remain challenging to treat [8]. Therefore, the current
60 gap between clinical intervention and patients' needs stems from an incomplete molecular
61 understanding of the pathological changes in the brain underlying these diseases.
62 A major issue faced by researchers is that large-scale molecular studies ideally require fresh tissue
63 in which nucleic acids remain intact. Acquiring such samples is exceptionally difficult due to
64 ethical considerations, logistical challenges, and donor scarcity [9], which hinders our ability to
65 capture the full spectrum of disease heterogeneity. Formalin-fixed paraffin-embedded (FFPE)
66 tissue archives represent a potential alternative. Over the decades, health institutions have
67 accumulated vast FFPE repositories encompassing diverse patient populations, disease states, and
68 historical periods.
69 One such resource is the Danish Brain Collection (DBC) (https://brain-collection.rsyd.dk/).
70 Assembled between 1945 and 1982, this repository comprises 9,479 brains from Danish
71 psychiatric patients, including an estimated 50,000 FFPE tissue blocks representing multiple brain
72 regions, and spanning widely different treatment eras dominated by divergent infectious disease
73 landscapes and environmental exposures. It ranks among the world's largest brain collections and
74 encompasses an exceptionally broad diagnostic spectrum: dementia, major depression,
75 schizophrenia, epilepsy, Huntington's and Parkinson's disease, among others. Examining
76 molecular signatures across this nearly four-decade window offers unique opportunities to
77 investigate whether brain disorder biology has remained constant or evolved through time,
78 something that contemporary biobanked archives alone cannot easily reproduce. Despite its
79 potential, the Danish Brain Collection remains largely unexplored using modern molecular
80 techniques. No systematic efforts have applied high-throughput sequencing to these samples,
81 likely due to ethics restrictions and concerns about their age and degradation status. Indeed,

formaldehyde fixation is known to create extensive cross-links and fragmentation in biomolecules present in the tissues [10], severely compromising genetic material integrity. For sequencing, these issues translate into low yield, poor library complexity, and high error rates [11]. Even after optimization efforts [12–15], the molecular viability of decades-old FFPE brain tissues archived in pathology collections has remained largely uncharacterized, particularly where fixation protocols were not standardized.

Nevertheless, despite the extensive induced cross-linking and chemical damage created in tissues upon FFPE treatment [16], valuable molecular information can still be recovered [17–20], even from spatially localized areas of a given FFPE block [21–23]. Indeed, compelling evidence suggests that the main limiting factor affecting nucleic acid integrity is not the age of the FFPE sample per se, but rather the chemical conditions with which the original tissues were treated [16,24–28]. Moreover, recent methodological developments have enabled the recovery of genomic and transcriptomic data from specimens up to thousands of years old using specialized protocols handling extreme fragmentation and chemical modifications [29–34]. Therefore, the main question is not whether genetic material may survive over time, but whether molecular information remains sufficiently intact to yield robust biological insights.

In this pilot study, we aimed to determine whether informative high-throughput DNA and RNA sequencing data from a subset of decades-old FFPE brain samples archived in the DBC can be isolated, sequenced and analyzed, so as to reconstruct tissue-specific gene expression patterns of the original human brain tissues. To do so, we first worked on optimizing laboratory methods for DNA and RNA extraction and library preparation and then applied custom bioinformatic pipelines inspired by previous reports on historical archived tissues from museum collections [32].

Methods

Study design

We used human post-mortem brain tissue in the form of formalin-fixed paraffin-embedded (FFPE) tissue archived at The Danish Brain Collection, which is located at Brain Research - Inter Disciplinary Guided Excellence (BRIDGE), University of Southern Denmark, OUH, the Region Psychiatry of Southern Denmark, Odense, Denmark. The Danish Brain Collection is a biobank comprising 9,479 brains, which were obtained during autopsies conducted at Danish state mental hospitals from 1945 to 1982 (https://brain-collection.rsyd.dk/). A total of 10 samples from 4 patients were included, representing the following anatomical brain regions: pons, thalamus, cerebellar vermis, and corpus striatum. An overview of the 10 FFPE blocks used and their tissue age and identity is shown in **Table 1**. The included patients were identified by searching for "epilepsy" in a database of the original patient diagnoses, and then choosing patients based on information from discharge letters stored on a GDPR-secured server, both provided by the Open Patient data Explorative Network (OPEN), OUH and Department of Clinical Research, University of Southern Denmark, Odense, Denmark. The specific pathological cohort is not the focus of the current study, which is instead intended to develop a methodological framework for the molecular study of aged FFPE blocks stored at the Danish Brain Collection.

Table 1: FFPE brain samples included in this study.

Patient	Sex	Age at death (years)	Year of death	Sample age (years)	ID	Brain tissue	Fixative agent
A	F	68	1952	73	A1	Pons	10% FF
					A2	Corpus striatum	
B	M	52	1964	61	B1	Thalamus	10% FF
					B2	Cerebellar vermis	
					B3	Pons	
					B4	Corpus striatum	
C	F	38	1967	58	C1	Pons	N/A
					C2	Corpus striatum	
D	M	63	1973	52	D1	Cerebellar vermis	Cajal's fixative
					D2	Corpus striatum	

M = male; F = female; FF = formalin fixation; N/A: No information available.

Ethics declaration

129 The study was approved by The Danish National Committee on Health Research Ethics (ethical
130 permission journal and supplement no. CVK-2308632 and CVK-112142) in accordance with the
131 Danish Act on Research Ethics Review of Health Research Projects (Komitéloven) and all
132 methods were performed in accordance with relevant guidelines and regulations.
133 The study is based on archival human biological material collected between 1945 and 1982, prior
134 to the introduction of requirements for obtaining informed consent for the samples in question.
135 Consequently, individual informed consent has not been obtained. This general issue concerning
136 The Brain Collection has been addressed previously in 2006 by The Danish National Committee
137 on Biomedical Research Ethics journal number 2005-7049-6.
138 Under Danish law, research use of such archival material without consent may be permitted
139 subject to approval by the competent research ethics committee, cf. Komitéloven. The present
140 study has been reviewed and approved on this basis.
141
142 **Preprocessing**
143 FFPE blocks from each patient were selected based on the registered time of death of the
144 individual, to ensure that a range of different preservation ages were included in the pilot study
145 to assess any potential effects of storage time on the quality of molecular data. Additionally, FFPE
146 blocks containing tissue from four different neuron-rich brain areas were included, representing
147 tissue-specific variation across different brain regions. Tissue identity of each block is provided
148 in **Table 1.** First, each FFPE block was sectioned into three separate pieces to ensure comparable
149 cell-rich grey matter content. To remove paraffin from the tissue, two pieces were incubated
150 overnight at 50°C, then for 1 hour at 60°C. The first piece was left in its melted state (referred to
151 as "Melted") and stored at -20°C. The second piece was re-embedded in paraffin and not used
152 any further in this study. The third piece remained unprocessed in its original state (referred to as
153 "original FFPE") and stored at room temperature in a dry, dark environment before being
154 transferred to -20°C.
155
156 **Isolation of DNA and RNA**

157 All laboratory procedures were carried out in clean laboratory facilities specifically dedicated to
158 the processing of historical and/or ancient samples within the GLOBE Institute (University of
159 Copenhagen, Denmark). DNA was extracted from FFPE blocks using three approaches: (i) the
160 DNeasy Blood & Tissue Kit (Qiagen) with heat-mediated decrosslinking at 90°C [16], (ii) the
161 same kit with an additional deparaffinization step using Deparaffinization Solution (Qiagen), and
162 (iii) the AllPrep® DNA/RNA FFPE Kit (Qiagen) for simultaneous DNA and RNA recovery. RNA
163 was independently extracted using the mirVana™ miRNA Isolation Kit (Thermo Fisher
164 Scientific), selected based on its prior successful application for RNA extraction from ancient
165 tissue samples [32,34], with additional pretreatment steps including heat-mediated
166 deparaffinization and decrosslinking. Detailed protocols for each extraction method, including
167 reagent volumes, incubation conditions, and all modifications to the manufacturers'
168 recommendations, are provided in **Supplementary Methods**. Additional metadata for DNA and
169 RNA extracts is also available in **Supplementary Tables 1** and **2**. Nucleic acid concentrations
170 were measured using Qubit™ High Sensitivity dsDNA and RNA kits (Thermo Fisher Scientific)
171 with a Qubit v3.0 fluorometer (Thermo Fisher Scientific). Size distribution and integrity of the
172 DNA and RNA extracts for each sample were assessed using the High Sensitivity RNA
173 ScreenTape, and High Sensitivity D1000 DNA ScreenTape in an Agilent 2200 TapeStation
174 System (Agilent Technologies).
175
176 **Library preparation**
177 DNA libraries were prepared following the Blunt-End-Single-Tube (BEST) protocol [35,36],
178 comprising end-repair, adapter ligation, and fill-in reaction steps. Libraries were quantified by
179 qPCR (Mx3005P, Agilent Technologies), indexed with unique dual 8-bp indices, and purified
180 using HighPrep PCR beads (MagBio Genomics). cDNA libraries from RNA extracts (n = 20,
181 from mirVana and AllPrep extractions) were prepared using the NEXTFLEX Small RNA-Seq
182 Kit v4 with Unique Dual Indexes (Revvity), without size selection. For a subset of 10 RNA
183 samples (from A1, A2, B2, C2, D1 pairs), end-repair was performed prior to additional library
184 preparation using T4 Polynucleotide Kinase (T4 PNK, NEB) to repair damaged 5' and 3' termini

185 on fragmented RNA molecules. A detailed protocol for all library preparation steps, including
186 reagent compositions, thermocycler programs, and clean-up procedures, is provided in
187 **Supplementary Methods**. Library concentrations were quantified using the Qubit™ dsDNA
188 Quantification Assay Kit (Thermo Fisher Scientific) with a Qubit 2.0 Fluorometer (Thermo Fisher
189 Scientific). The size distribution of the DNA and cDNA libraries was measured using the High
190 Sensitivity DNA assay (Agilent Technologies) with a Bioanalyzer 2100 system (Agilent
191 Technologies).
192
193 **Sequencing**
194 The DNA (n = 33) and cDNA (n = 30) libraries were separately submitted to Novogene (Planegg,
195 Germany), and independently sequenced on an Illumina NovaSeq X Plus system, with a full 10B
196 lane in 150 bp paired-end configuration. For cDNA libraries, a homogeneous distribution of
197 sequencing effort across all samples was applied. For DNA libraries, we balanced the sequencing
198 effort so that each of the four original individuals obtained equivalent amounts of sequences at a
199 target genome-wide coverage of at least 10×.
200
201 **Statistical analysis**
202 Statistical comparison of DNA and RNA yield between extraction methods and pretreatment
203 conditions was performed using two-sided paired t-tests in R v4.4.2. *P*-values are reported without
204 correction for multiple testing.
205
206 **Sequence analysis**
207 Raw FASTQ files were quality-filtered using fastp v0.24.1 [37] with the following parameters:
208 automatic adapter detection and trimming, polyG tail removal, paired-end read merging, a
209 minimum sequence length of 30 nt for DNA and 18 nt for RNA data, and a minimum overlap of
210 30 and 15 matching nucleotides for paired-end merging of DNA and RNA sequences,
211 respectively. Only successfully merged paired sequences were retained for subsequent analyses.
212 The human RefSeq assembly (GRCh38.p14, GCF_000001405.40) and associated gene

213 annotations were used as reference. A reference transcriptome was generated using gffread
214 v0.12.6 [38], retaining only the longest transcript isoform per gene across protein-coding
215 messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), long noncoding
216 RNAs (lncRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs). The
217 microRNA (miRNA) annotation was retrieved from miRGeneDB v3.0 [39].
218 Alignment was performed using Bowtie2 v2.4.2 [40] with end-to-end sensitive mode. DNA
219 sequences were first sorted with the sambamba v0.8.0 tool [41], then deduplicated using the
220 markdup function from SAMtools v.1.10 [42]. RNA sequences were not deduplicated, as the
221 absence of unique molecular identifiers (UMIs) in the library preparation protocol prevents
222 reliable distinction between PCR duplicates and genuinely identical RNA fragments derived from
223 highly expressed loci. RNA library complexity was instead estimated as 1 minus the duplication
224 rate reported by fastp. Genic and intergenic mapping was resolved using BEDtools v2.31.0 [43].
225 RNA sequences mapped to the reference transcriptome that were also identified as either
226 overlapping intergenic or intronic regions genome-wide were flagged and removed. Taxonomic
227 classification of DNA and RNA sequences was performed using KrakenUniq v0.7.3 [44] with a
228 custom database [45] to characterize endogenous human content and exogenous contamination.
229 Additional details are described in **Supplementary Methods**.
230
231 **Gene expression quantification**
232 RNA sequences mapped to the human reference transcriptome were assigned to each
233 corresponding gene with the intersect function from BEDtools v2.31.0 [43], and the percentage
234 of the transcript body supported by any mapped read (breadth of coverage) was determined with
235 the BEDtools coverage function. For miRNA genes, RNA sequences mapped to mature miRNA
236 loci were considered as successful alignments if they overlapped the corresponding region by at
237 least 75% of their length (bedtools intersect -f 0.75). Genes were considered successfully captured
238 if they had a transcript-wide coverage >10%.
239
240 **Damage analysis**

241 The PMDtools v0.60 [46] software was used to identify C>U deamination events in both RNA
242 and DNA sequencing data. For DNA data, PCR-deduplicated sequences mapped to the GRCh38
243 assembly were analyzed over the first 50 nucleotides of each mapped sequence. For RNA,
244 sequences mapped to the custom transcriptome were assessed over 30 nucleotides with single-
245 stranded mode enabled. Only C>U deaminated sites, read as C>T by the sequencer, were
246 considered.
247
248 **Clustering and tissue-specific gene expression analysis**
249 Protein-coding mRNA and miRNA quantification for each RNA sample were used as input for
250 Principal Component Analysis (PCA) after variance-stabilizing transformation with DESeq2
251 v1.44.0 [47]. For tissue-specific clustering, UMAP dimensionality reduction was applied using
252 the umap v0.2.10.0 R package (https://github.com/tkonopka/umap). As modern reference, we
253 used a comprehensive human RNA-seq atlas retrieved with the getAtlasExperiment function (E-
254 MTAB-2836) from ExpressionAtlas v1.32.0 R package [48] including 11 healthy tissues [49,50],
255 supplemented with chondrocyte samples [51] and four brain regions [52]. For the microRNA
256 fraction, small RNA-seq datasets from human healthy tissues were used [52–54]. A
257 comprehensive list of all reference samples is provided in **Supplementary Table 3**.
258 Quantification values were normalized with the trimmed mean of M-values (TMM) method [55]
259 using edgeR v4.2.1 [56] and $\log_2$ transformed. FFPE RNA data were then projected onto the
260 learned UMAP embedding. Further details on reference data processing and UMAP parameters
261 are provided in **Supplementary Methods**.
262 Hierarchical clustering was performed separately on protein-coding mRNA and miRNA profiles
263 from end-repaired libraries (n = 10) using the top 100 most variable genes. TMM-normalized
264 $\log_2$-transformed CPM values were first converted to z-scores per gene and clustered with the
265 Ward's minimum variance method (ward.D2) and Euclidean distance, then visualized using
266 ComplexHeatmap v2.20.0 [57]. Tissue specificity was quantified using the tau (τ) index [58],
267 computed from the reference multi-tissue atlas. The tau index was defined as:
268

269 $$\tau = \frac{\sum_{i=1}^{n}(1-\hat{x}_i)}{n-1}; \quad \hat{x}_i = \frac{x_i}{\max(x)}$$

270

271 where n is the number of tissues, $\hat{x}_i$ is the expression normalized to the tissue with maximum
272 expression, and x_i is the average expression in tissue i. Tau values range from 0 (uniform
273 expression) to 1 (tissue-restricted). Validated tissue markers were defined as genes with: (i) $\tau \geq$
274 0.5; (ii) peak expression in cerebellar vermis or corpus striatum in the reference atlas; (iii)
275 detectable FFPE abundance (average $\log_2$ CPM $\geq$ 3); (iv) transcript coverage > 10% in the
276 corresponding FFPE tissue (protein-coding genes only); and (v) concordant directionality
277 between the reference atlas and FFPE data. Full details on clustering and filtering criteria, as well
278 as on gene selection are provided in **Supplementary Methods.**

279

280

281 # Results

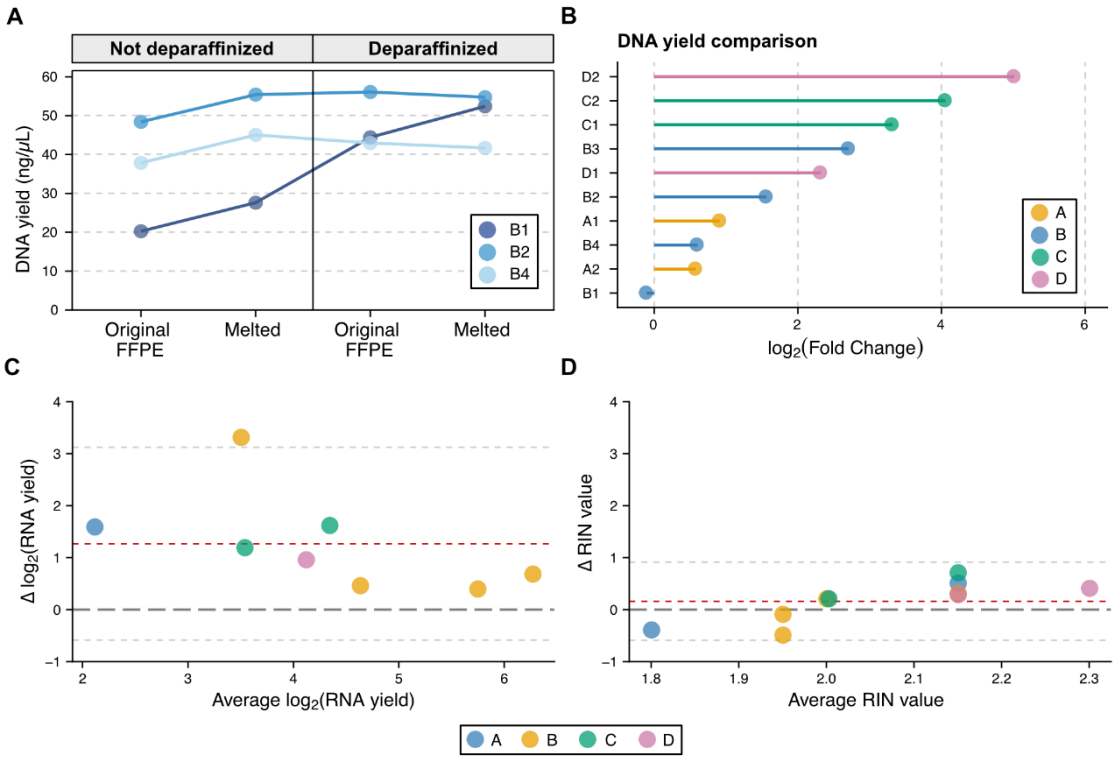
282 We have analyzed post-mortem formalin-fixed paraffin-embedded (FFPE) brain tissue from the
283 Danish Brain Collection, obtained from autopsies carried out at Danish state hospitals between
284 1945 and 1982 (**Table 1**). Both DNA and RNA were extracted in bulk, sequenced and analyzed
285 comparing different laboratory approaches for pretreatment, nucleic acid isolation and
286 transcriptional damage repair. Complete metadata information including pretreatment, extraction
287 batch, sample identity, extraction method and nucleic acid yield is available in **Supplementary**
288 **Table 1** for DNA and **Supplementary Table 2** for RNA extracts.

289

290 **Pretreatment of FFPE blocks and nucleic acid extraction method affect DNA yield**

291 We first sought to compare the effectiveness of different methods for removing paraffin from
292 FFPE brain tissues, and how this impacted the DNA concentration obtained after nucleic acid
293 extraction. Overnight paraffin melting significantly increased DNA yield compared to original
294 untreated FFPE samples without paraffin removal (*P*-value = 1.8×10^{-4}), and this was also

295 observed when applying a deparaffinization solution (**Fig. 1A**), although it did not reach statistical
296 significance (*P*-value = 0.1752). All samples tested for pretreatment conditions and their DNA
297 yield are in **Supplementary Table 4**.



298

299 **Figure 1: Pretreatment and extraction protocol impact on DNA and RNA yield.** (**A**) DNA
300 yield obtained from FFPE samples B1, B2, and B4 using two pretreatment approaches: physical
301 removal of excess paraffin (original vs. melted samples) and enzymatic deparaffinization (see
302 Methods). (**B**) Fold change of the difference in DNA yield between the AllPrep and Blood &
303 Tissue methods for DNA extraction. A positive fold change value means a higher DNA yield was
304 obtained with the AllPrep protocol, and vice versa. (**C**) Bland-Altman plot on the agreement
305 between AllPrep and mirVana methods for RNA extraction according to the yield obtained with
306 both protocols. The dark red dotted line represents the average difference in RNA yield in $\log_2$
307 scale. Light grey dotted lines represent the upper and lower confidence intervals at ±1.96 standard
308 deviations from the average. (**D**) Bland-Altman plot on the agreement between AllPrep and
309 mirVana methods for RNA extraction according to the RNA integrity (RIN) values obtained with
310 both protocols. The dark red dotted line represents the average difference in RIN values. Light

311 grey dotted lines represent the upper and lower confidence intervals at ±1.96 standard deviations
312 from the average. Additional data from patients A, B, C, and D can be found in **Table 1**.
313
314 Next, we compared two different DNA extraction approaches on FFPE blocks. The DNeasy
315 Blood & Tissue method from Qiagen is designed for simple and efficient recovery of DNA from
316 different types of tissues. Incidentally, it can also recover RNA molecules during the extraction
317 protocol [16], but we did not use this fraction from this particular kit. Moreover, it is not
318 specifically tailored for FFPE samples, so we adapted it to include an initial decrosslinking step
319 following prior reports demonstrating the efficacy of heat-mediated decrosslinking for FFPE
320 DNA recovery [16]. In contrast, the other method we tested (AllPrep® DNA/RNA FFPE also from
321 Qiagen) is specifically designed to work with FFPE tissues and allows simultaneous isolation of
322 DNA and RNA molecules from the same sample. Despite this, it is a more time-consuming and
323 involved protocol requiring more than double the time for its manual implementation. We again
324 extracted DNA in an independent batch from all ten FFPE brain samples using both approaches.
325 Upon comparison of both extraction kits, we found differences in both DNA yield and fragment
326 length distribution, with an overall higher but variable DNA yield obtained with the AllPrep
327 protocol compared to the Blood & Tissue protocol (**Fig. 1B**). Numeric values for total DNA
328 concentration in all analyzed samples of this batch are available in **Supplementary Table 5**.
329
330 **Extraction protocol affects RNA yield but not RNA integrity**
331 Next, to determine the efficacy of RNA isolation from decades-old FFPE tissues, we tested two
332 different kits and evaluated both RNA yield and transcript integrity (as expressed by RIN values).
333 We again compared the AllPrep kit, which allows RNA extraction jointly with DNA, and the
334 mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific), which is optimized for short-length
335 RNA recovery. The extraction method significantly influenced RNA yield, but not the RNA
336 integrity. The use of the AllPrep kit resulted in extracts with significantly higher RNA
337 concentration than the mirVana protocol (P-value = 3.79×10^{-3}). Interestingly, the AllPrep
338 method showed a higher tendency to improve the RNA yield over the mirVana kit at lower

339 concentrations (**Fig. 1C**). Conversely, there were no significant differences in RIN values between
340 the mirVana and AllPrep techniques (**Fig. 1D**, *P*-value = 0.2197), and the AllPrep kit only
341 marginally improved the RNA integrity obtained at RIN values >2. Numeric values for total RNA
342 concentration and RIN values in all analyzed samples are available in **Supplementary Table 6**.
343

344 **Decades-old FFPE brain samples yield mappable nucleic acids**

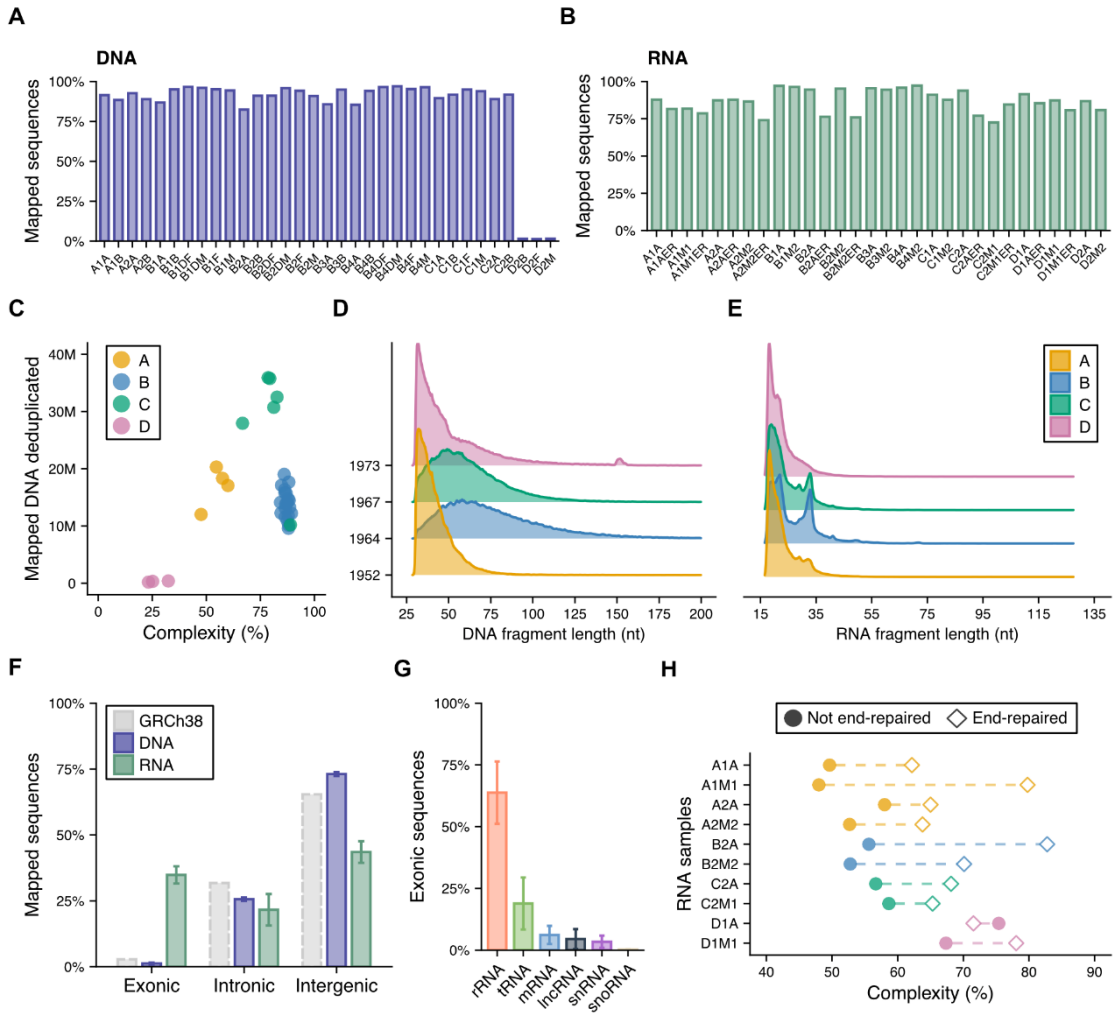
345 After sequencing and mapping to the human (GRCh38) genome assembly, both DNA and RNA
346 from FFPE brain tissues in the DBC showed a high alignment rate, with overall above 75% of
347 sequences successfully mapped (**Supplementary Table 7**, **Figs. 2A** and **2B**). However, DNA
348 sequences belonging to patient D had ultra-low successful mapping, with only about 1.5% of them
349 mapped to the human assembly (**Supplementary Table 7**, **Fig. 2A**). The remaining unmapped
350 sequences in patient D libraries were attributable to a combination of microbial DNA consistent
351 with environmental contamination (**Supplementary Table 8**, see Methods), library artifacts, and
352 unmappable ultrashort fragments resulting from severe DNA degradation. These observations,
353 combined with high duplication rates (**Supplementary Table 7**) indicate severe DNA
354 degradation in these FFPE blocks, likely reflecting variation in historical fixation conditions
355 within the collection as described in **Table 1**.

356 Consistent with this observation, samples from patients A, B, and C clearly outperformed those
357 from patient D not only in the number of mapped DNA sequences, but also in library complexity
358 (**Fig. 2C**), i.e. the number of uniquely aligned sequences after deduplication. Moreover, the
359 fragment length distribution of successfully mapped DNA sequences was markedly shifted
360 toward shorter fragments in samples from patients A and D compared to patients B and C, which
361 showed longer and more dispersed fragment lengths with an average of ~60 nucleotides (**Fig.**
362 **2D**). Interestingly, patient D, despite being the most recently deceased of the four individuals
363 (date of death recorded in 1973, **Table 1**), showed shortened DNA fragments mapped to the
364 human assembly, similar to the profiles observed for patient A, the oldest of the four (**Fig. 2D**).
365 This, together with the very low number of mapped DNA sequences and extensive degradation

366 of DNA in the FFPE blocks from this individual, suggests that chemical treatment during fixation,

367 rather than age, is the main driver of DNA quality observed in our data.

368



369

370 **Figure 2: Sequencing and mapping statistics of DNA and RNA libraries from decades-old**

371 **FFPE brain samples.** (**A**) Percentage of DNA sequences successfully mapped to the human

372 genome assembly (GRCh38) per library (see **Table 1** and **Supplementary Table 1** for additional

373 data). (**B**) Percentage of RNA sequences successfully mapped to the human genome assembly

374 (GRCh38) per library (see **Table 1** and **Supplementary Table 2** for additional data). (**C**) Library

375 complexity of DNA samples, defined as the percentage of uniquely mapped sequences after

376 deduplication. Fragment length density distribution of (**D**) DNA and (**E**) RNA sequences mapped

377 to the human genome, grouped by patient and the recorded date of death highlighting the age of

378 the FFPE blocks. (**F**) Proportion of mapped sequences overlapping exonic, intronic, and

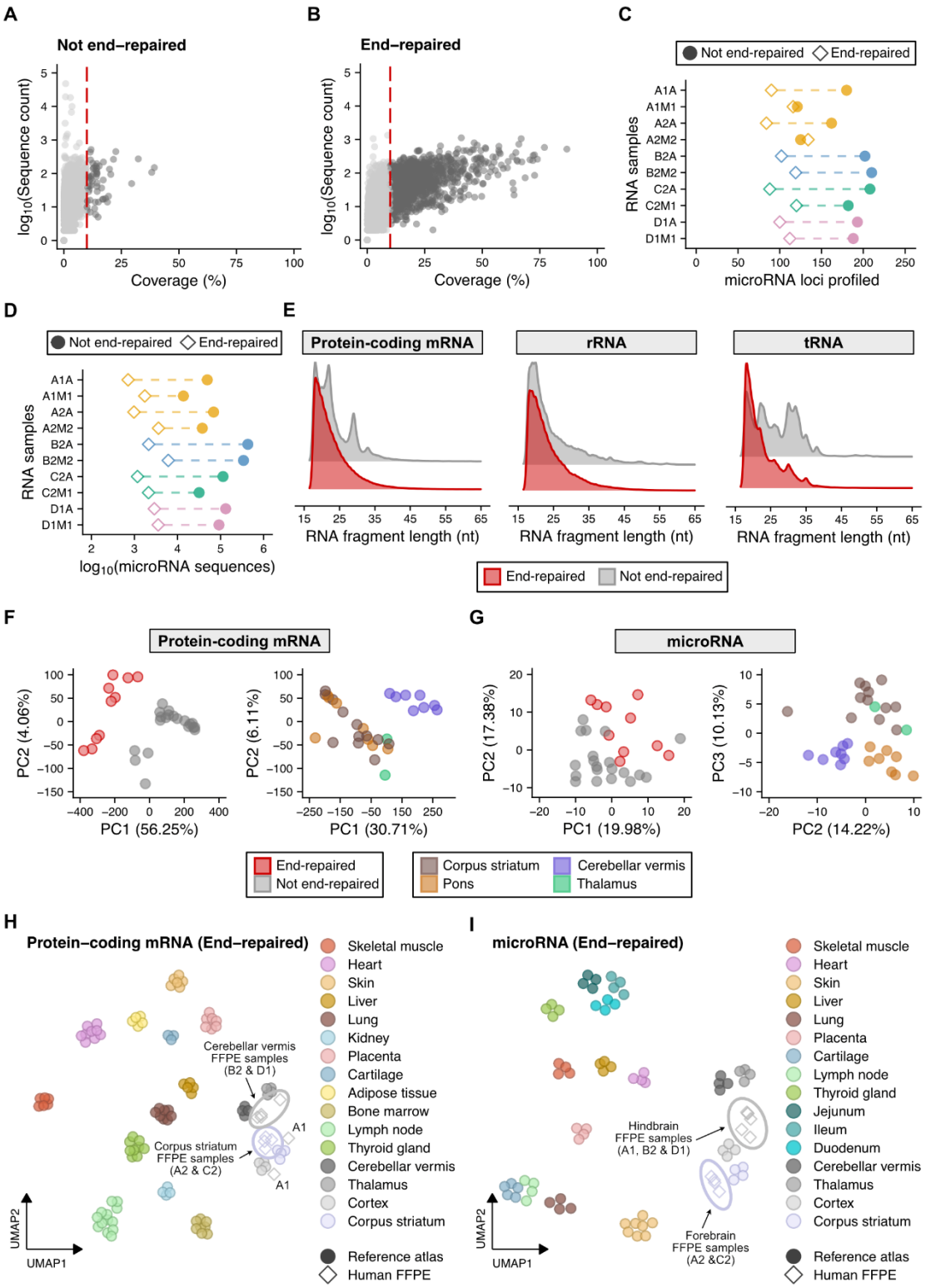
379 intergenic regions for DNA and RNA samples. Grey dashed values represent the proportion of
380 exonic, intronic and intergenic regions annotated genome-wide in the human assembly
381 (GRCh38). See Methods for more details on the loci considered. (**G**) RNA biotype composition
382 of sequences mapped to exonic regions, showing the relative proportion of ribosomal RNA
383 (rRNA), transfer RNA (tRNA), protein-coding messenger RNA (mRNA), long-noncoding RNA
384 (lncRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA) annotated loci. (**H**)
385 Library complexity of RNA samples processed with and without T4 PNK end-repair prior to
386 library preparation.
387
388 These patterns contrast with those observed for the mapped RNA sequences, where most
389 fragments were of ultrashort length, between 18 and 30 nucleotides (**Fig. 2E**). As with DNA
390 samples, no clear age-related differences in RNA fragment length distribution were observed.
391 Interestingly, contrary to what was observed in DNA, FFPE RNA from patient D samples showed
392 profiles similar to the rest (**Supplementary Table 7**), including scarce microbial RNA
393 contamination (**Supplementary Table 9**). Nevertheless, despite their short overall fragment
394 length, RNA sequences showed clear signs of transcriptome-like mapping patterns, with nearly
395 35% of mapped fragments falling within annotated exonic regions of the human genome assembly
396 (GRCh38, **Fig. 2F, Supplementary Table 7**). Intronic and intergenic mapping RNAs accounted
397 for approximately 22% and 44% of the sequences, respectively. Conversely, DNA sequences
398 showed markedly different proportions, more closely reflecting the overall genomic distribution
399 of exonic, intronic, and intergenic regions, where DNA fragments mapped to exonic regions
400 accounted for less than 2% of the total mapped sequences (**Fig. 2F**). Among exonic RNA
401 sequences, the vast majority overlapped with rRNA loci, followed by tRNA, and protein-coding
402 loci (**Fig. 2G**). Other less abundant noncoding loci such as lncRNAs, snRNAs and snoRNAs were
403 also detected (**Fig. 2G**). Comprehensive abundance and transcript coverage profiles for all
404 transcript types analyzed (see Methods) are available in **Supplementary Tables 10-15.**
405
406 **End-repair treatment increases library complexity but shortens RNA fragments**

407 We next evaluated the effect of end-repairing damaged RNA fragment termini from decades-old
408 FFPE samples on adapter ligation efficiency during library preparation, and its impact on the
409 resulting transcriptomic sequencing data. To do so, we processed 10 samples (**Supplementary**
410 **Table 2**) with and without applying T4 PNK end-repair prior to library preparation (see Methods).
411 All end-repaired samples showed a relative increase in library complexity (**Fig. 2H**), defined as
412 the fraction of non-redundant (i.e. non-PCR-duplicate) RNA fragments. Samples without end-
413 repair treatment had an average 57% complexity, meaning that over 40% of sequencing effort
414 likely consisted of redundant PCR duplicates, while end-repaired samples showed an average
415 71% complexity (**Fig. 2H**).
416 Consistent with the higher library complexity observed for RNA samples, end-repair also
417 significantly increased the number of protein-coding mRNA loci profiled at >10% transcript
418 coverage (**Figs. 3A** and **3B**), a threshold we used as a heuristic for reliably detected genes [32].
419 This effect was observed across all 10 samples evaluated with and without end-repair treatment
420 (**Supplementary Figs. 1-2**). Furthermore, for some of the more degraded and poorly profiled
421 samples, particularly those from patients A and C, end-repair was essential to recover any protein-
422 coding mRNA genes above this coverage threshold (**Supplementary Figs. 1-2**). Interestingly,
423 several highly abundant mRNA loci detected in samples without end-repair (**Fig. 3A**), although
424 at ultra-low coverage (<10%), showed reduced abundance but increased transcript coverage after
425 end-repair (**Fig. 3B**). This pattern suggests that end-repair enables a more even sampling of
426 mRNA fragments across transcripts, rather than libraries being dominated by PCR duplicates of
427 a small number of highly abundant loci. However, the increased resolution obtained for the
428 protein-coding fraction in end-repaired RNA libraries did not extend to microRNA loci. Libraries
429 without end-repair treatment yielded a higher number of profiled microRNA genes (**Fig. 3C**) and
430 more mapped RNA fragments overlapping miRNA loci compared to end-repaired samples (**Fig.**
431 **3D**), suggesting that end-repair is detrimental for microRNA recovery. A comprehensive
432 quantification of all microRNAs detected is available in **Supplementary Table 16**.
433 End-repaired samples also showed a reduced average RNA fragment length compared to samples
434 without end-repair, an effect that was consistent across all gene types analyzed but particularly

435 pronounced for the most abundant transcript classes: rRNAs, tRNAs, and protein-coding mRNAs

436 (**Fig. 3E**).

437



438

439 **Figure 3: End-repair effects on RNA library composition and tissue-specific gene expression**
440 **profiling from FFPE brain samples.** (**A**) Sequence count vs. transcript coverage for protein-
441 coding mRNAs in the 2A library from patient B without end-repair treatment (B2A sample). The
442 red dashed line marks a 10% coverage threshold, defined as the lower limit to consider a given
443 protein-coding mRNA gene as successfully profiled. (**B**) Sequence count vs. transcript coverage
444 for protein-coding mRNAs in the 2A library from patient B with end-repair treatment (B2AER
445 sample). (**C**) Number of microRNA loci profiled per sample, comparing libraries with and without
446 end-repair treatment. (**D**) Number of mapped microRNA sequences per sample, comparing
447 libraries with and without end-repair treatment. (**E**) RNA fragment length density distribution for
448 protein-coding mRNA, rRNA, and tRNA genes, comparing end-repaired (red) and non-end-
449 repaired (grey) libraries. (**F**) PCA of protein-coding mRNA expression profiles across all 30 FFPE
450 brain samples. The left panel shows all samples colored by end-repair status. The right panel
451 shows the same data after regressing out the end-repair effect, with samples colored by brain
452 region. (**G**) PCA of microRNA expression profiles across all 30 FFPE brain samples. The left
453 panel shows all samples colored by end-repair status. The right panel shows the same data with
454 samples colored by brain region. (**H**) UMAP embedding of protein-coding mRNA expression
455 from end-repaired FFPE brain samples (n = 10) projected onto a human multi-tissue RNAseq
456 reference atlas. (**I**) UMAP embedding of microRNA expression from end-repaired FFPE brain
457 samples projected onto a human multi-tissue small RNAseq reference atlas. FFPE sample pairs
458 per tissue belong to patients A (n = 4, A1: Pons & A2: Corpus striatum), B (n = 2, B2: Cerebellar
459 vermis), C (n = 2, C2: Corpus striatum), and D (n = 2, D1: Cerebellar vermis). The hindbrain
460 region comprises pons and cerebellar vermis tissues, while the forebrain region comprises corpus
461 striatum tissue.
462
463 **Damage profiles indicate moderate cytosine deamination in FFPE-derived DNA and RNA**
464 To assess the extent of chemical damage in the sequenced nucleic acids, we examined nucleotide
465 misincorporation patterns along mapped DNA and RNA fragments. For DNA, all samples
466 showed a characteristic cytosine-to-uracil (C>U) mismatch enrichment at the 5' end of fragments

467 and the complementary guanine-to-adenine (G>A) enrichment at the 3' end (**Supplementary Fig.**
468 **3**), consistent with cytosine deamination induced by formalin fixation and post-mortem hydrolytic
469 damage, as typically observed in degraded and archival DNA [10,59,60]. The magnitude of these
470 terminal misincorporation frequencies was moderate, reaching approximately 2-4% at the first
471 position across all samples, with a gradual decay toward background levels within the first 5-10
472 nucleotides. This pattern was consistent across all four patients regardless of sample age. For
473 RNA samples, cytosine deamination frequencies were assessed along end-repaired fragments
474 (**Supplementary Fig. 4**). Average deamination rates ranged from less than 1% to ~3.5%, with a
475 slight enrichment at the 5' terminal positions relative to the fragment body. Compared to the DNA
476 profiles, RNA deamination rates were of similar or slightly higher magnitude but showed a more
477 uniform distribution along the fragment length rather than a steep terminal enrichment, consistent
478 with the combined effects of formalin-induced and hydrolytic cytosine deamination acting across
479 the full length of single-stranded RNA molecules. Samples from patients A and D showed higher
480 deamination rates compared to those from patients B and C (**Supplementary Fig. 4**), broadly
481 consistent with the increased DNA fragmentation observed for these individuals (**Fig. 2D**).
482

483 **Decades-old FFPE brain samples retain tissue-specific gene expression signatures**

484 To determine whether the transcriptomic data recovered from FFPE brain samples captured
485 biologically meaningful gene expression patterns, we performed dimensionality reduction
486 analyses on both protein-coding mRNA and microRNA quantification profiles. PCA on protein-
487 coding mRNA expression confirmed that end-repair status was the dominant source of variance,
488 accounting for 56.25% of the total variation along PC1 (**Fig. 3F**, left panel). After regressing out
489 the end-repair effect, the corrected PCA (**Fig. 3F**, right panel) revealed a clear separation by brain
490 region along the first two components, with cerebellar vermis samples clustering distinctly from
491 pons, corpus striatum, and thalamus samples. For microRNAs, PCA showed a less structured
492 pattern (**Fig. 3G**, left panel) when examining the end-repair status, but separation by brain region
493 became apparent when examining PC2 vs. PC3 (**Fig. 3G**, right panel), again with cerebellar

494 vermis as the most distinct tissue. No clear influence of sex was observed for either protein-coding
495 mRNAs or microRNAs (**Supplementary Figs. 5A** and **5B**).
496 To place these expression profiles in a broader biological context, we first projected the FFPE
497 brain protein-coding mRNA profiles from end-repaired samples onto a human multi-tissue
498 reference atlas (**Fig. 3H**). The FFPE brain samples consistently clustered adjacent to reference
499 brain tissue samples, distinct from all other tissues included in the reference atlas, including
500 skeletal muscle, heart, liver, lung, kidney, skin, and gut-derived tissues, among others. Moreover,
501 we found a clear divergence between end-repaired FFPE samples obtained from cerebellar vermis
502 tissue compared to those from the corpus striatum (**Fig. 3H**). Samples from patient A (Pons tissue)
503 did not cluster with other hindbrain-derived tissues such as the cerebellar vermis, nor did they
504 form a separate cluster. This result confirms that despite decades of formalin fixation and paraffin
505 embedding, these archival brain samples retain gene expression signatures that are recognizably
506 brain-specific, reflecting differentiated protein-coding transcriptional programs between
507 forebrain and hindbrain regions in the central nervous system. A similar pattern was found when
508 projecting FFPE samples without end-repair treatment onto the same multi-tissue protein-coding
509 human reference atlas, with all samples clustering nearby or adjacent to brain-derived tissues
510 (**Supplementary Fig. 5C**). However, only those obtained from cerebellar vermis tissue showed
511 a clear tissue-specific clustering pattern.
512 For microRNA profiles, projection onto a multi-tissue human microRNA expression atlas also
513 showed a clearly defined clustering pattern, with end-repaired FFPE brain samples dividing into
514 those derived from the hindbrain (pons and cerebellar vermis), and those derived from the
515 forebrain (corpus striatum), again reflecting a shared post-transcriptional regulatory program
516 among brain tissues of common developmental origin (**Fig. 3I**). When considering FFPE samples
517 without end-repair treatment, however, such division was not apparent, although all of them
518 clustered adjacent to brain-derived tissues included in the microRNA expression reference
519 (**Supplementary Fig. 5D**).
520 Unsupervised hierarchical clustering of the top 100 most variable protein-coding mRNA genes
521 separated the 10 FFPE samples primarily by patient and then by brain region, with cerebellar

522 vermis and corpus striatum samples forming two distinct clusters (**Supplementary Fig. 6**). The
523 two pons samples, both derived from patient A, clustered separately and were not further
524 evaluated as region-specific. A similar tissue-driven separation was observed for the top 100 most
525 variable miRNAs (**Supplementary Fig. 7**).
526 Furthermore, to determine whether this clustering reflects known regional gene expression
527 patterns, we computed tissue specificity values (τ index) for all protein-coding mRNA and
528 microRNA genes included in the multi-tissue gene expression collections used as reference
529 (**Supplementary Tables 17-18**). Filtering criteria included tissue-restricted expression ($\tau \geq 0.5$),
530 concordant directionality between the reference atlas and FFPE end-repaired samples, detectable
531 abundance in FFPE samples, and adequate transcriptional coverage (see Methods). Across the
532 full transcriptome, 321 protein-coding genes were validated as tissue markers: 177 in cerebellar
533 vermis and 144 in corpus striatum. Of these, 28 fell within the top 100 most variable genes in the
534 hierarchical clustering (18 cerebellar vermis, and 10 corpus striatum; **Supplementary Table 19**).
535 Among the validated cerebellar vermis markers, several canonical cerebellar genes were
536 identified, including the granule cell-associated zinc finger transcription factors *ZIC1* and *ZIC4*,
537 the Purkinje cell synapse organizers *CBLN1* and *CBLN3*, the metabotropic glutamate receptor
538 *GRM1,* the ETS transcription factor *ETV1*, and the voltage-gated potassium channel *KCNA1*,
539 among others. Additional well-established cerebellar markers such as *GABRA6*, *PVALB* and
540 *PAX6* were validated outside the top 100 variable gene set (**Supplementary Table 19**). Corpus
541 striatum-enriched markers included the neuropeptide *CCK*, the postsynaptic calmodulin-binding
542 protein *NRGN*, the calcium/calmodulin-dependent kinase *CAMK2A*, the synaptic protein *DDN*,
543 and the cortexin *CTXN1*. The neuropeptides *SST* and *NPY* were among the strongest striatal
544 markers validated outside the top 100 (**Supplementary Table 19**).
545 For miRNAs, 14 were validated as tissue-specific markers in end-repaired FFPE samples, of
546 which 13 were among the 100 most variable according to hierarchical clustering (**Supplementary**
547 **Table 20**). Four were detected as predominant in cerebellar vermis and ten in corpus striatum.
548 The strongest cerebellar miRNA markers were miR-124 paralogues (miR-124-P1-v1 and miR-
549 124-P2-v2), along with miR-376-P2 and miR-154-P14. Corpus striatum-related miRNAs

550 included miR-137 (τ = 0.97), as well as miR-138-P2, miR-132, and miR-433 (**Supplementary**
551 **Table 20**).
552
553
554 # Discussion
555 In this study, we demonstrate that decades-old FFPE brain tissue from the Danish Brain Collection
556 retains sufficient molecular integrity for informative genomic and transcriptomic analyses.
557 Systematic evaluation of extraction, library preparation, and bioinformatic approaches shows that
558 archival samples yield recoverable DNA and RNA with sequencing data capturing biologically
559 meaningful, tissue-specific expression patterns.
560
561 **Practical considerations for nucleic acid recovery from aged FFPE tissues**
562 Comparison of extraction methods identified the AllPrep protocol as the most suitable for DBC
563 material, yielding higher DNA and RNA concentrations than Blood & Tissue and mirVana while
564 requiring less input tissue. Its ability to simultaneously recover DNA and RNA from the same
565 specimen is particularly advantageous for limited archival material and enables parallel genomic
566 and transcriptomic analyses.
567 We note, however, that the Blood & Tissue protocol for DNA extraction yielded longer fragments,
568 indicating a trade-off between yield and fragment length. While longer DNA sequences may
569 benefit applications such as haplotype phasing or structural variant detection, the higher yield
570 from AllPrep was more advantageous for the short-read sequencing used here.
571
572 **Fixation chemistry drives variability in nucleic acid preservation**
573 Not all samples performed comparably. Tissues from patient D exhibited markedly reduced DNA
574 mapping rates and library complexity relative to the rest of the cohort. Notably, these samples
575 were preserved using Cajal's fixative rather than neutral buffered formalin. Cajal's fixative is
576 highly acidic (pH ~1.5) [61], a condition known to accelerate depurination and phosphodiester

577 bond cleavage in nucleic acids [25]. Consistent with prior reports, acidic or unbuffered fixation
578 leads to increased fragmentation and reduced amplifiability compared to neutral conditions
579 [16,28]. Therefore, the poor performance of patient D samples might be explained by the fixation
580 chemistry rather than by tissue age, given that patient D is paradoxically the most recently
581 deceased individual in the cohort (date of death 1973).
582 Patient A, the oldest individual in the cohort (1952), exhibited the shortest DNA fragment lengths
583 and reduced RNA library complexity. The two available pons samples clustered with other brain
584 regions but were insufficient (n = 2, single individual) for robust assessment of region-specific
585 gene expression. Whether this reflects intrinsic properties of the pons, patient-specific
586 degradation, or both cannot be resolved with the current data. Together with the findings from
587 patient D, these results indicate that fixation chemistry, rather than specimen age, is the dominant
588 determinant of nucleic acid preservation in archival material. Accordingly, detailed metadata on
589 fixation conditions (i.e. fixative type, concentration, pH, duration, and temperature) should be
590 systematically recorded and incorporated into study design, sample selection, and downstream
591 analyses when working with FFPE collections.
592
593 **End-repair improves FFPE transcriptomics**
594 The application of T4 PNK end-repair prior to library preparation emerged as a key step for
595 recovering informative transcriptomic data from highly degraded FFPE RNA. End-repaired
596 libraries showed substantially higher complexity, with a greater proportion of sequences
597 representing independent RNA molecules rather than PCR duplicates. This improvement
598 translated into enhanced gene detection sensitivity, enabling the tissue-specific gene expression
599 profiling reported here.
600 The effect of T4 PNK end-repair likely reflects restoration of damaged 5' and 3' termini on
601 fragmented RNA, which in their unrepaired state cannot efficiently ligate to sequencing adapters
602 during library preparation. This is consistent with reports demonstrating improved mRNA capture
603 from extracellular vesicles and plasma RNA [62,63], and extends these findings to the more
604 severely degraded context of decades-old FFPE tissue. A trade-off is a relative reduction in

605 microRNA representation, as the increased capture of longer repairable RNA fragments likely
606 dilutes the small RNA fraction in the final library. This is an important consideration for study
607 design, as omitting end-repair may be preferable when small RNAs are the primary target.
608 Nevertheless, end-repaired libraries retained biologically informative miRNA profiles that
609 recapitulated tissue-specific patterns observed in untreated samples, and more clearly resolved
610 the hindbrain-forebrain distinction, consistent with differences in embryological origin and post-
611 transcriptional regulation.
612 Furthermore, the shorter fragment length distribution observed in end-repaired libraries likely
613 reflects either preferential restoration of termini on shorter damaged fragments, rendering them
614 ligatable and thereby shifting the overall distribution, additional fragmentation introduced during
615 enzymatic treatment and heat inactivation, or a combination of both. Future optimization could
616 explore alternative inactivation strategies, such as bead-based cleanup without heat exposure, to
617 preserve ultrashort fragments while retaining the gains in library complexity conferred by end-
618 repair.
619
620 **Biological validity of FFPE-derived transcript abundance profiles**
621 A central finding of this study is that archival FFPE brain tissue retains recognizably tissue-
622 specific transcriptomic signatures despite decades of formalin fixation and paraffin embedding.
623 When projected onto independent modern multi-tissue reference atlases, FFPE brain samples
624 clustered with brain-derived tissues and were clearly separated from non-neural tissues, across
625 both protein-coding mRNA and microRNA profiles. Within the brain compartment, cerebellar
626 vermis and corpus striatum were readily distinguished, recapitulating the well-documented
627 transcriptomic divergence between hindbrain and forebrain regions [64–66].
628 Tissue specificity analysis using the tau index provided quantitative support for these
629 observations. Of the 321 protein-coding genes and 14 microRNAs validated as tissue markers,
630 many are canonical regional markers with well-characterized roles in cerebellar or striatal
631 biology. Cerebellar vermis markers included the zinc finger transcription factors *ZIC1* and *ZIC4*,
632 which define granule cell identity [67,68], the synaptic organizers *CBLN1* and *CBLN3*, essential

633 for Purkinje cell synapse formation [69–71], the GABA-A receptor subunit *GABRA6*, exclusively
634 expressed in cerebellar granule cells [72–74], and *PVALB*, a calcium-binding protein enriched in
635 Purkinje cells [75]. In the corpus striatum, validated markers included the interneuron-associated
636 neuropeptides *SST* and *NPY* [76,77], the neuron-specific kinase *CAMK2A* [78,79], and the
637 calcium/calmodulin-dependent kinase *CAMKV* [80].
638 MicroRNA markers were similarly consistent with known regional biology. MiR-137, a robustly
639 forebrain-enriched microRNA and a GWAS locus for schizophrenia [81,82], was among the
640 strongest striatal markers, while the miR-124 family showed the expected cerebellar
641 predominance [83–86]. The recovery of these well-established markers from tissue dating to the
642 1940s and 1950s demonstrates that regional molecular identity can survive the combined effects
643 of post-mortem interval, formalin fixation, and prolonged storage.
644 Not all established markers met our validation criteria. Several canonical striatal genes, including
645 *HTR2C* [87], *DRD2* [88], and *GNAL* [89], were detectable in FFPE corpus striatum samples but
646 did not achieve sufficient transcriptional coverage for robust analysis. This highlights a key
647 limitation of FFPE-based transcriptomics: while transcript presence and relative abundance can
648 be captured, a comprehensive transcriptional reconstruction remains challenging for a subset of
649 loci. Broader sampling across FFPE blocks and brain regions may improve transcript recovery,
650 but further methodological advances will be needed to achieve more complete transcriptome
651 profiling from highly degraded archival material.
652
653 **Implications for the Danish Brain Collection**
654 Our results establish a practical framework for molecular profiling of the Danish Brain Collection
655 and, by extension, similar historical FFPE archives. The combination of AllPrep extraction, T4
656 PNK end-repair, and single-stranded library preparation enabled recovery of tissue-specific
657 expression profiles from samples up to 73 years old at the time of analysis. In parallel, a
658 bioinformatic pipeline adapted from ancient and museum specimen studies [32,34] effectively
659 accommodated the extensive fragmentation and chemical damage characteristic of these
660 materials.

661 The Danish Brain Collection, comprising approximately 9,500 brains spanning four decades of
662 psychiatric history, represents a unique resource for questions inaccessible to contemporary
663 biobanks. These include the molecular correlates of historical diagnostic categories, the impact
664 of changing environmental exposures and treatment regimes on brain molecular profiles, and
665 retrospective investigation of infectious contributors to psychiatric disease in the pre-antibiotic
666 and early antibiotic eras. The metagenomic and metatranscriptomic classification framework
667 applied here, used primarily to assess contamination and library quality, provides a foundation
668 for such investigations and could be extended to systematically screen for pathogen-derived
669 nucleic acids across larger sample sets within the collection. The feasibility of recovering
670 informative genomic and, critically, transcriptomic data from this pilot study opens the door to
671 larger-scale investigations targeting defined diagnostic cohorts within the archive.
672 Nevertheless, several limitations warrant consideration. First, the sample size reported here is
673 modest (four patients, 10 tissue blocks). Second, tissue representation is uneven across brain
674 regions and individuals due to archival constraints, and the lack of matched fresh-frozen controls
675 prevents direct quantification of information loss associated with FFPE processing. Third, DNA
676 sequencing depth, while sufficient to assess mapping efficiency and damage patterns, would need
677 to be increased to support robust variant calling or comprehensive genomic analyses. These
678 limitations can be addressed in future studies through expanded cohorts, deeper sequencing, and
679 inclusion of matched contemporary controls where available.
680
681

682 ## Conclusions

683 In summary, we demonstrate that decades-old FFPE brain tissue from the Danish Brain Collection
684 is amenable to high-throughput DNA and RNA sequencing, retaining biologically meaningful,
685 tissue-specific molecular signatures. These results establish the DBC as a viable resource for
686 retrospective genomic and transcriptomic studies of neurological, neurodegenerative, and
687 psychiatric disorders.

Declarations

Consent for publication

Not applicable

Availability of data and materials

The sequence data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession numbers PRJNA1442830 and PRJNA1442847.

Competing interests

The authors declare that they have no competing interests.

Funding

EM-S acknowledges financial support from the Villum Fonden (Villum Experiment project no. 57875). NO is supported by the European Regional Development Fund project "Functional Omics Analysis of Metabolic Diseases to Advance Drug Discovery Research Excellence in LIOS - TARGETWISE" (no. 1.1.1.5/2/24/A/003). CK acknowledges financial support to this project through The Danish Association of Biomedical Laboratory Scientists.

Authors' contributions

EM-S and CK conceived this study and devised the experimental design. EM-S and CK secured funding. NO performed metagenomic and metatranscriptomic analyses. CRC and EHF performed laboratory work with support and supervision from SSTM and EM-S. CRC, EHF, EM-S, and NO analyzed the data, with input from CK and MTPG. CRC, EHF and EM-S visualized the data. MW

714 facilitated ethics approval and sample acquisition. CRC, EHF, CK and EM-S wrote the
715 manuscript with contributions and agreement from all authors.
716
717 **Acknowledgements**
718 Not applicable
719
720

721 References

722 **1.** World Health Organization. Global status report on neurology. *World Health Organization*, 1–
723 285 (2025).
724 **2.** Mitchell, A. J. *et al.* Estimating global direct health-care spending on neurological and mental
725 health between 2000 and 2019: a modelling study. *Lancet Public Health*. **10**, e401–11 (2025).
726 **3.** Owen, M. J., Bray, N. J., Walters, J. T. R. & O'Donovan, M. C. Genomics of schizophrenia,
727 bipolar disorder and major depressive disorder. *Nat. Rev. Genet*. **26**, 862–77 (2025).
728 **4.** Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related
729 dementias. *Nat. Genet.* **54**, 412–36 (2022).
730 **5.** Perucca, P., Bahlo, M. & Berkovic, S. F. The Genetics of Epilepsy. *Annu. Rev. Genomics Hum.*
731 *Genet*. **21**, 205–30 (2020).
732 **6.** Armstrong, R. A., Lantos, P. L. & Cairns, N. J. Overlap between neurodegenerative disorders.
733 *Neuropathology* **25**, 111–24 (2005).
734 **7.** Martinez, B. & Peplow, P. V. Misdiagnoses of psychiatric disorders and the utility of diagnostic
735 microRNA biomarkers. *NeuroMarkers* **1**, 100002 (2024).
736 **8.** Howes, O. D., Thase, M. E. & Pillinger, T. Treatment resistance in psychiatry: state of the art
737 and new directions. *Mol. Psychiatry.* **27**, 58–72 (2022).
738 **9.** Danner, B. *et al.* Brain banking in the United States and Europe: Importance, challenges, and
739 future trends. *J. Neuropathol. Exp. Neurol.* **83**, 219–29 (2024).

740 **10.** Steiert, T. A. *et al.* A critical spotlight on the paradigms of FFPE-DNA sequencing. *Nucleic*
741 *Acids Res.* **51**, 7143–62 (2023).
742 **11.** Do, H. & Dobrovic, A. Sequence artifacts in DNA from formalin-fixed tissues: causes and
743 strategies for minimization. *Clin Chem.* **61**, 64–71 (2015).
744 **12.** Patel, P. G. *et al.* Reliability and performance of commercial RNA and DNA extraction kits
745 for FFPE tissue cores. *PLoS One* **12**, e0179732 (2017).
746 **13.** Wehmas, L. C. *et al.* Demodifying RNA for Transcriptomic Analyses of Archival Formalin-
747 Fixed Paraffin-Embedded Samples. *Toxicol. Sci.* **162**, 535–47 (2018).
748 **14.** Haynes, H. R. *et al.* Evaluation of the quality of RNA extracted from archival FFPE
749 glioblastoma and epilepsy surgical samples for gene expression assays. *J. Clin. Pathol.* **71**, 695–
750 701 (2018).
751 **15.** Jacobsen, S. B., Tfelt-Hansen, J., Smerup, M. H., Andersen, J.D. & Morling, N. Comparison
752 of whole transcriptome sequencing of fresh, frozen, and formalin-fixed, paraffin-embedded
753 cardiac tissue. *PLoS One*. **18**, e0283159 (2023).
754 **16.** Gilbert, M. T. P. *et al.* The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded
755 Tissues–Which Methods Are Useful When? *PLoS One*. **2**, e537 (2007).
756 **17.** Gilbert, M. T. P. *et al.* Multiplex PCR with minisequencing as an effective high-throughput
757 SNP typing method for formalin-fixed tissue. *Electrophoresis.* **28**, 2361–7 (2007).
758 **18.** Ludyga, N. *et al.* Nucleic acids from long-term preserved FFPE tissues are suitable for
759 downstream analyses. *Virchows Arch.* **460,** 131–40 (2012).
760 **19.** Hedegaard, J. *et al.* Next-generation sequencing of RNA and DNA isolated from paired fresh-
761 frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS*
762 *One*. **9**, e98187 (2014).
763 **20.** Zhao, Y. I. *et al.* Robustness of RNA sequencing on older formalin-fixed paraffin-embedded
764 tissue from high-grade ovarian serous adenocarcinomas. *PLoS One.* **14**, e0216050 (2019).
765 **21.** Gracia Villacampa, E. *et al.* Genome-wide spatial expression profiling in formalin-fixed
766 tissues. *Cell Genom.* **1**, 100065 (2021).

767 **22.** Bai, Z. *et al.* Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded
768 tissues. *Cell.* **187**, 6760 (2024).
769 **23.** Wang, T. *et al.* snPATHO-seq, a versatile FFPE single-nucleus RNA sequencing method to
770 unlock pathology archives. *Commun. Biol.* **7**, 1340 (2024).
771 **24.** Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M. & Okubo, K. Analysis of chemical
772 modification of RNA from formalin-fixed samples and optimization of molecular biology
773 applications for such samples. *Nucleic Acids Res.* **27**, 4436–43 (1999).
774 **25.** Srinivasan, M., Sedmak, D. & Jewell, S. Effect of Fixatives and Tissue Processing on the
775 Content and Integrity of Nucleic Acids. *Am. J. Pathol.* **161**, 1961–71 (2002).
776 **26.** Shi, S.R. *et al.* DNA extraction from archival formalin-fixed, paraffin-embedded tissue
777 sections based on the antigen retrieval principle: heating under the influence of pH. *J. Histochem.*
778 *Cytochem.* **50**, 1005–11 (2002).
779 **27.** Shi, S.R. *et al.* DNA extraction from archival formalin-fixed, paraffin-embedded tissues: heat-
780 induced retrieval in alkaline solution. *Histochem. Cell Biol.* **122**, 211–8 (2004).
781 **28.** Groelz, D. *et al.* Non-formalin fixative versus formalin-fixed tissue: A comparison of
782 histology and RNA quality. *Exp. Mol. Pathol.* **94**, 188–94 (2013).
783 **29.** Smith, O. *et al.* Ancient RNA from Late Pleistocene permafrost and historical canids shows
784 tissue-specific transcriptome survival. *PLoS Biol.* **17**, e3000166 (2019).
785 **30.** van der Valk, T. *et al.* Million-year-old DNA sheds light on the genomic history of mammoths.
786 *Nature* **591**, 265–9 (2021).
787 **31.** Fromm, B. *et al.* Ancient microRNA profiles of 14,300-yr-old canid samples confirm
788 taxonomic origin and provide glimpses into tissue-specific gene regulation from the Pleistocene.
789 *RNA* **27**, 324–34 (2021).
790 **32.** Mármol-Sánchez, E. *et al.* Historical RNA expression profiles from the extinct Tasmanian
791 tiger. *Genome Res.* **33**, 1299–316 (2023).
792 **33.** Sandoval-Velasco, M. *et al.* Three-dimensional genome architecture persists in a 52,000-year-
793 old woolly mammoth skin sample. *Cell* **187**, 3541–3562.e51 (2024).

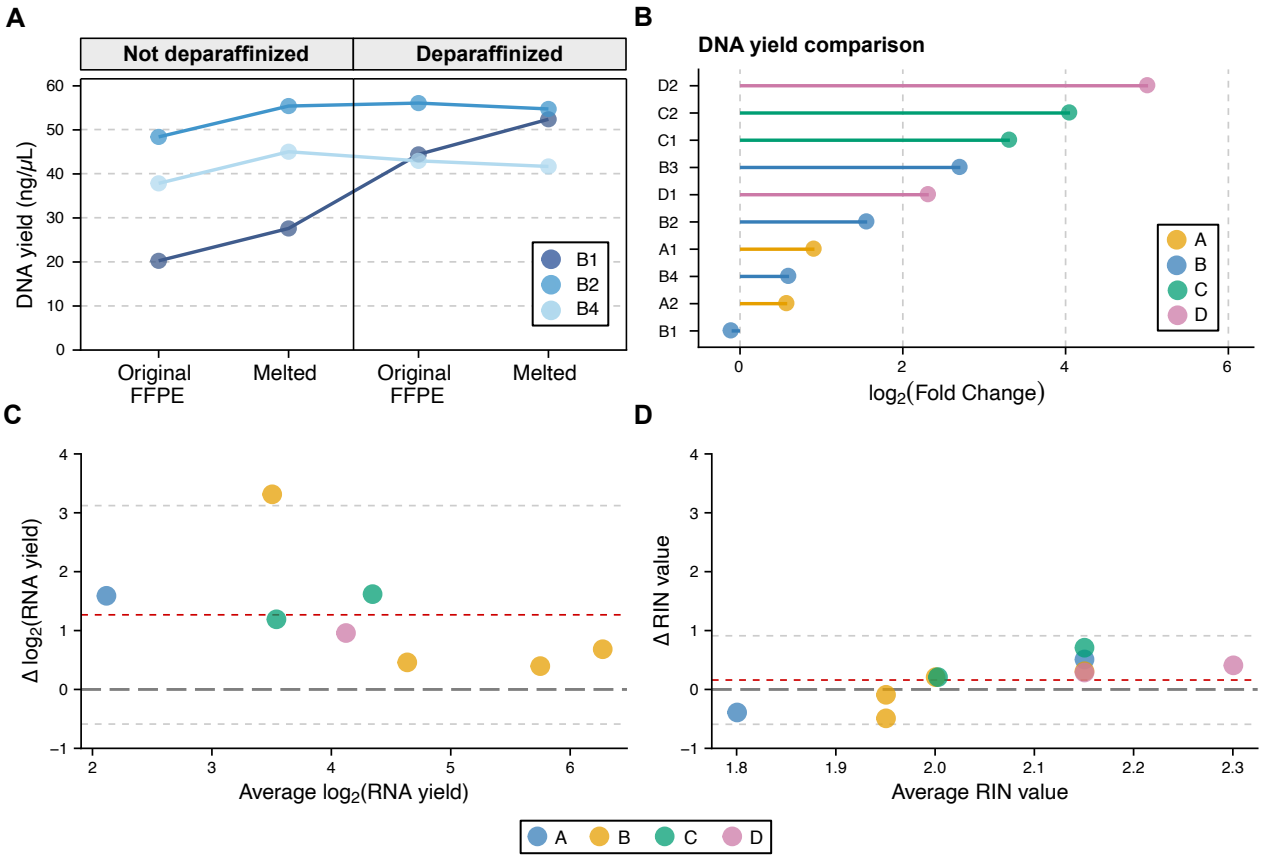
794 **34.** Mármol-Sánchez, E. *et al.* Ancient RNA expression profiles from the extinct woolly
795 mammoth. *Cell* **189**, 52–69.e22 (2026).
796 **35.** Mak, S. S. T. *et al.* Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500
797 sequencing platforms for palaeogenomic sequencing. *Gigascience* **6**, 1–13 (2017).
798 **36.** Carøe, C. *et al.* Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**,
799 410–9 (2018).
800 **37.** Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
801 *Bioinformatics* **34**, i884–90 (2018).
802 **38.** Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Res.* **9**, 304 (2020).
803 **39.** Clarke, A. W. *et al.* MirGeneDB 3.0: improved taxonomic sampling, uniform nomenclature
804 of novel conserved microRNA families and updated covariance models. *Nucleic Acids Res.* **53**,
805 D116–28 (2025).
806 **40.** Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
807 357–9 (2012).
808 **41.** Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing
809 of NGS alignment formats. *Bioinformatics* **31**, 2032–4 (2015).
810 **42.** Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9
811 (2009).
812 **43.** Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
813 features. *Bioinformatics* **26**, 841–2 (2010).
814 **44.** Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: Confident and fast
815 metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 1–10 (2018).
816 **45.** Pochon, Z. *et al.* aMeta: an accurate and memory-efficient ancient metagenomic profiling
817 workflow. *Genome Biol.* **24**, 1–30 (2023).
818 **46.** Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a
819 Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–34 (2014).
820 **47.** Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
821 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

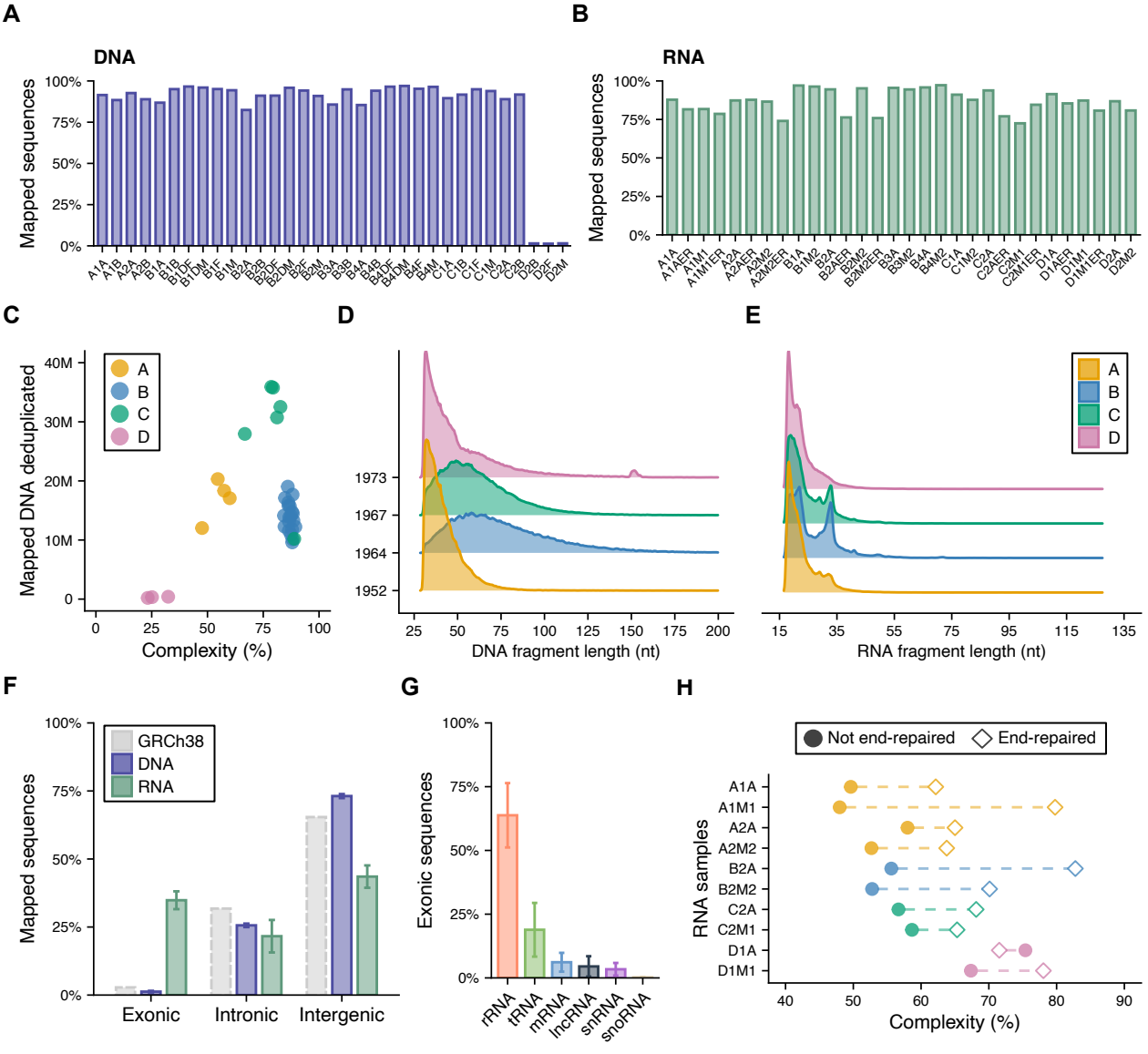
822 **48.** Madrigal, P. *et al.* Expression Atlas in 2026: enabling FAIR and open expression data through
823 community collaboration and integration. *Nucleic Acids Res.* **54**, D147–57 (2026).
824 **49.** Fagerberg, L. *et al.* Analysis of the Human Tissue-specific Expression by Genome-wide
825 Integration of Transcriptomics and Antibody-based Proteomics. *Mol. Cell. Proteomics* **13**, 397–
826 406 (2014).
827 **50.** Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
828 **51.** Zhu, G. *et al.* Comparing effects of perfusion and hydrostatic pressure on gene profiles of
829 human chondrocyte. *J. Biotechnol.* **210**, 59–65 (2015).
830 **52.** Lorenzi, L. *et al.* The RNA Atlas expands the catalog of human non-coding RNAs. *Nat.*
831 *Biotechnol.* **39**, 1453–65 (2021).
832 **53.** Forrest, A. R. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–70
833 (2014).
834 **54.** Srivastava, A. *et al.* Next-Generation Sequencing Identifies the Keratinocyte-Specific miRNA
835 Signature of Psoriasis. *J. Invest. Dermatol.* **139**, 2547–2550.e12 (2019).
836 **55.** Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression
837 analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
838 **56.** Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
839 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).
840 **57.** Gu, Z. & Hübschmann, D. Make Interactive Complex Heatmaps in R. *Bioinformatics* **38**,
841 1460–2 (2022).
842 **58.** Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
843 relationships in human tissue specification. *Bioinformatics* **21**, 650–9 (2005).
844 **59.** Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc.*
845 *Natl Acad. Sci. USA* **104**, 14616–21 (2007).
846 **60.** Guo, Q. *et al.* The mutational signatures of formalin fixation on the human genome. *Nat.*
847 *Commun.* **13**, 4487 (2022).
848 **61.** Kiernan, J. A. *Histological and Histochemical Methods: Theory and Practice* 5th edn (Scion
849 Publishing, 2015).

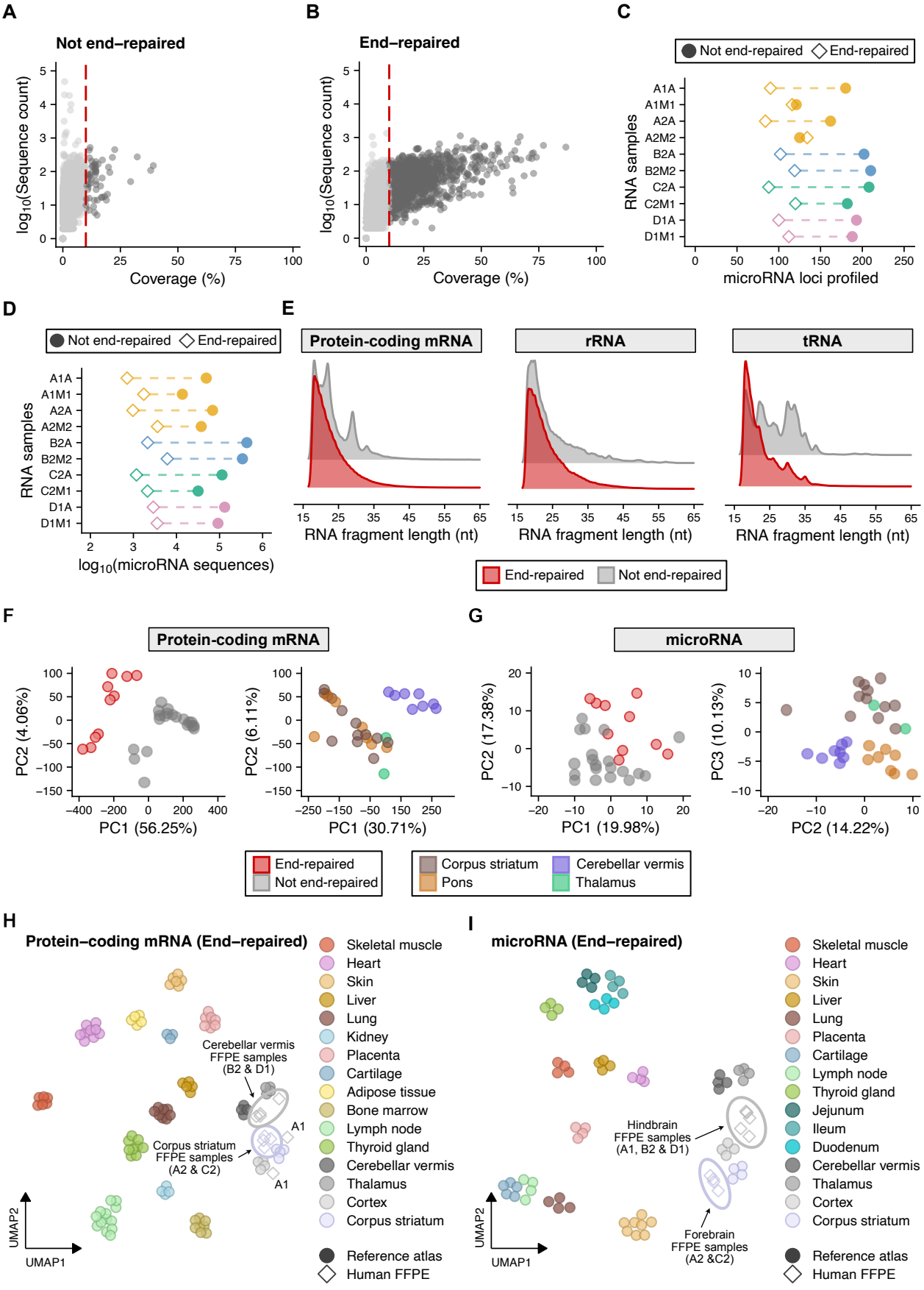
850 **62.** Akat, K. M. *et al.* Detection of circulating extracellular mRNAs by modified small-RNA-
851 sequencing analysis. *JCI Insight* **5**, e127317 (2019).
852 **63.** Solaguren-Beascoa, M. *et al.* Phospho-RNA-Seq Highlights Specific Small RNA Profiles in
853 Plasma Extracellular Vesicles. *Int. J. Mol. Sci.* **24**, 11653 (2023).
854 **64.** Hawrylycz, M.J. *et al.* An anatomically comprehensive atlas of the adult human brain
855 transcriptome. *Nature* **489**, 391–9 (2012).
856 **65.** Hawrylycz, M. *et al.* Canonical Genetic Signatures of the Adult Human Brain. *Nat. Neurosci.*
857 **18**, 1832–44 (2015).
858 **66.** Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues.
859 *Science* **369**, 1318–30 (2020).
860 **67.** Blank, M. C. *et al.* Multiple developmental programs are altered by loss of Zic1 and Zic4 to
861 cause Dandy-Walker malformation cerebellar pathogenesis. *Development* **138**, 1207–16 (2011).
862 **68.** Aruga, J. & Millen, K. J. ZIC1 Function in Normal Cerebellar Development and Human
863 Developmental Pathology. *Adv. Exp. Med. Biol.* **1046**, 249–68 (2018).
864 **69.** Hirai, H. *et al.* Cbln1 is essential for synaptic integrity and plasticity in the cerebellum. *Nat.*
865 *Neurosci.* **8**, 1534–41 (2005).
866 **70.** Yuzaki, M. Cbln1 and its family proteins in synapse formation and maintenance. *Curr. Opin.*
867 *Neurobiol.* **21**, 215–20 (2011).
868 **71.** Takeo, Y. H. *et al.* GluD2- and Cbln1-mediated competitive interactions shape the dendritic
869 arbors of cerebellar Purkinje cells. *Neuron* **109**, 629–644.e8 (2021).
870 **72.** Laurie, D. J., Seeburg, P.H. & Wisden, W. The distribution of 13 GABAA receptor subunit
871 mRNAs in the rat brain. II. Olfactory bulb and cerebellum. *J. Neurosci.* **12**, 1063–76 (1992).
872 **73.** Sathyanesan, A. *et al.* Emerging connections between cerebellar development, behaviour and
873 complex brain disorders. *Nat. Rev. Neurosci.* **20**, 298–313 (2019).
874 **74.** Sieghart, W. *et al.* α6-Containing GABAA Receptors: Functional Roles and Therapeutic
875 Potentials. *Pharmacol. Rev.* **74**, 238–70 (2022).
876 **75.** Brandenburg, C. *et al.* Parvalbumin subtypes of cerebellar Purkinje cells contribute to
877 differential intrinsic firing properties. *Mol. Cell. Neurosci.* **115**, 103650 (2021).

878 **76.** Kumar, U. Somatostatin in Medium-Sized Aspiny Interneurons of Striatum is Responsible for
879 Their Preservation in Quinolinic Acid and N-Methyl-d-Asparate-Induced Neurotoxicity. *J. Mol.*
880 *Neurosci.* **35**, 345–54 (2008).
881 **77.** Garma, L. D. *et al.* Interneuron diversity in the human dorsal striatum. *Nat. Commun.* **15**,
882 6164 (2024).
883 **78.** Klug, J. R. *et al.* Genetic inhibition of CaMKII in dorsal striatal medium spiny neurons reduces
884 functional excitatory synapses and enhances intrinsic excitability. *PLoS One* **7**, e45323 (2012).
885 **79.** Pasek, J. G., Wang, X. & Colbran, R. J. Differential CaMKII regulation by voltage-gated
886 calcium channels in the striatum. *Mol. Cell. Neurosci.* **68**, 234–43 (2015).
887 **80.** Liang, Z. *et al.* The pseudokinase CaMKv is required for the activity-dependent maintenance
888 of dendritic spines. *Nat. Commun.* **7**, 13282 (2016).
889 **81.** Ripke, S. *et al.* Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature*
890 **511**, 421–7 (2014).
891 **82.** Cheng, Y. *et al.* Partial loss of psychiatric risk gene Mir137 in mice causes repetitive behavior
892 and impairs sociability and learning via increased Pde10a. *Nat. Neurosci.* **21**, 1689–703 (2018).
893 **83.** Ludwig, N. *et al.* Distribution of miRNA expression across human tissues. *Nucleic Acids Res.*
894 **44**, 3865–77 (2016).
895 **84.** Xu, J. *et al.* miR-124: A Promising Therapeutic Target for Central Nervous System Injuries
896 and Diseases. *Cell Mol. Neurobiol.* **42**, 2031–53 (2021).
897 **85.** Papadimitriou, E. *et al.* A miR-124-mediated post-transcriptional mechanism controlling the
898 cell fate switch of astrocytes to induced neurons. *Stem Cell Reports* **18**, 915–35 (2023).
899 **86.** Son, G. *et al.* miR-124 coordinates metabolic regulators acting at early stages of human
900 neurogenesis. *Commun. Biol.* **7**, 1393 (2024).
901 **87.** Mickey, B. J. *et al.* Striatal Dopamine Release and Genetic Variation of the Serotonin 2C
902 Receptor in Humans. *J. Neurosci.* **32**, 9344–50 (2012).
903 **88.** Bonnavion, P. *et al.* Striatal projection neurons coexpressing dopamine D1 and D2 receptors
904 modulate the motor function of D1- and D2-SPNs. *Nat. Neurosci.* **27**, 1783–93 (2024).

905 **89.** El Atiallah, I. *et al.* Loss-of-function of GNAL dystonia gene impairs striatal dopamine
906 receptors-mediated adenylyl cyclase/cyclic AMP signaling pathway. *Neurobiol. Dis.* **191**, 106403
907 (2024).


A
Not deparaffinized
Deparaffinized
DNA yield (ng/μL)
60
50
40
30
20
10
0
Original FFPE
Melted
Original FFPE
Melted
B1
B2
B4
B
DNA yield comparison
D2
C2
C1
B3
D1
B2
A1
B4
A2
B1
0
2
4
6
log2(Fold Change)
A
B
C
D
C
Δ log2(RNA yield)
4
3
2
1
0
−1
2
3
4
5
6
Average log2(RNA yield)
D
Δ RIN value
4
3
2
1
0
−1
1.8
1.9
2.0
2.1
2.2
2.3
Average RIN value
A
B
C
D


A
DNA
Mapped sequences
100%
75%
50%
25%
0%
A1A A1B A2A A2B B1A B1B B1DF B1DM B1F B1M B2A B2B B2DF B2DM B2F B2M B3A B3B B4A B4B B4DF B4DM B4F B4M C1A C1B C1F C1M C2A C2B D2B D2F D2M
B
RNA
Mapped sequences
100%
75%
50%
25%
0%
A1A A1AER A1M1 A1M1ER A2A A2AER A2M2 A2M2ER B1A B1M2 B2A B2AER B2M2 B2M2ER B3A B3M2 B4A B4M2 C1A C1M2 C2A C2AER C2M1 C2M1ER D1A D1AER D1M1 D1M1ER D2A D2M2
C
Mapped DNA deduplicated
40M
30M
20M
10M
0
0 25 50 75 100
Complexity (%)
A
B
C
D
D
1973
1967
1964
1952
25 50 75 100 125 150 175 200
DNA fragment length (nt)
E
A
B
C
D
15 35 55 75 95 115 135
RNA fragment length (nt)
F
Mapped sequences
100%
75%
50%
25%
0%
GRCh38
DNA
RNA
Exonic
Intronic
Intergenic
G
Exonic sequences
100%
75%
50%
25%
0%
rRNA
tRNA
mRNA
lncRNA
snRNA
snoRNA
H
Not end-repaired
End-repaired
RNA samples
A1A
A1M1
A2A
A2M2
B2A
B2M2
C2A
C2M1
D1A
D1M1
40 50 60 70 80 90
Complexity (%)


A
Not end–repaired
log10(Sequence count)
Coverage (%)
B
End–repaired
log10(Sequence count)
Coverage (%)
C
Not end-repaired
End-repaired
RNA samples
A1A
A1M1
A2A
A2M2
B2A
B2M2
C2A
C2M1
D1A
D1M1
microRNA loci profiled
D
Not end-repaired
End-repaired
RNA samples
log10(microRNA sequences)
E
Protein-coding mRNA
rRNA
tRNA
RNA fragment length (nt)
End-repaired
Not end-repaired
F
Protein-coding mRNA
PC2 (4.06%)
PC1 (56.25%)
PC2 (6.11%)
PC1 (30.71%)
G
microRNA
PC2 (17.38%)
PC1 (19.98%)
PC3 (10.13%)
PC2 (14.22%)
End-repaired
Not end-repaired
Corpus striatum
Pons
Cerebellar vermis
Thalamus
H
Protein–coding mRNA (End–repaired)
Cerebellar vermis FFPE samples (B2 & D1)
Corpus striatum FFPE samples (A2 & C2)
A1
A1
UMAP2
UMAP1
Skeletal muscle
Heart
Skin
Liver
Lung
Kidney
Placenta
Cartilage
Adipose tissue
Bone marrow
Lymph node
Thyroid gland
Cerebellar vermis
Thalamus
Cortex
Corpus striatum
Reference atlas
Human FFPE
I
microRNA (End–repaired)
Hindbrain FFPE samples (A1, B2 & D1)
Forebrain FFPE samples (A2 &C2)
UMAP2
UMAP1
Skeletal muscle
Heart
Skin
Liver
Lung
Placenta
Cartilage
Lymph node
Thyroid gland
Jejunum
Ileum
Duodenum
Cerebellar vermis
Thalamus
Cortex
Corpus striatum
Reference atlas
Human FFPE