



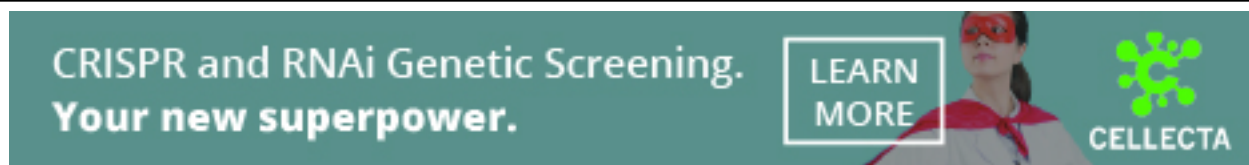
Profiling chromatin accessibility in formalin-fixed paraffin-embedded samples

Vamsi Krishna Polavarapu, Pengwei Xing, hua Zhang, et al.

Genome Res. published online July 14, 2021

Access the most recent version at doi:[10.1101/gr.275269.121](https://doi.org/10.1101/gr.275269.121)

P<P	Published online July 14, 2021 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

1 **TITLE PAGE:**

2 **TITLE:**

3 **Profiling chromatin accessibility in formalin-fixed paraffin-**
4 **embedded samples**

5

6 **AUTHORS:**

7 Vamsi Krishna Polavarapu^{1,3}, Pengwei Xing^{1,3}, Hua Zhang^{1,3}, Miao Zhao¹, Lucy Mathot¹,
8 Linxuan Zhao¹, Gabriela Rosen¹, Fredrik J. Swartling¹, Tobias Sjöblom¹, Xingqi Chen^{1,2*}

9

10 **AFFILIATIONS:**

11 ¹Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala
12 University, 75108, Uppsala, Sweden

13 ²Beijer Laboratories, Uppsala University, Uppsala, Sweden

14

15 ³Contributed equally

16

17 **CORRESPONDING AUTHORS:**

18 * xingqi.chen@igp.uu.se

19

20 **KEY WORDS: FFPE, chromatin accessibility**

21 **ABSTRACT:**

22 Archived formalin-fixed paraffin-embedded (FFPE) samples are the global standard format
23 for preservation of the majority of biopsies in both basic research and translational cancer
24 studies, and profiling chromatin accessibility in the archived FFPE tissues is fundamental to
25 understanding gene regulation. Accurate mapping of chromatin accessibility from FFPE
26 specimens is challenging because of the high degree of DNA damage. Here, we first
27 showed that standard ATAC-seq can be applied to purified FFPE nuclei but yields lower
28 library complexity and a smaller proportion of long DNA fragments. We then present FFPE-
29 ATAC, the first highly sensitive method for decoding chromatin accessibility in FFPE tissues
30 that combines Tn5-mediated transposition and T7 *in vitro* transcription. The FFPE-ATAC
31 generates high-quality chromatin accessibility profiles with 500 nuclei from a single FFPE
32 tissue section, enables the dissection of chromatin profiles from the regions of interest with
33 the aid of hematoxylin and eosin (H&E) staining, and reveals disease-associated chromatin
34 regulation from the human colorectal cancer FFPE tissue archived for more than 10 years.
35 In summary, the approach allows decoding of the chromatin states that regulate gene
36 expression in archival FFPE tissues, thereby permitting investigators, to better understand
37 epigenetic regulation in cancer and precision medicine.

38

39 INTRODUCTION:

40 Decoding the landscapes of chromatin regulatory elements in human disease, specifically
41 cancer, is of critical importance in preclinical diagnosis and treatment(Qu et al. 2017).
42 Recently developed technologies, such as the assay for transposase-accessible chromatin
43 by sequencing (ATAC-seq) (Buenrostro et al. 2013)and DNase I hypersensitivity sequencing
44 (DNase-seq)(Jin et al. 2015), allow profiling of chromatin accessibility in cells and frozen
45 tissues. Archived formalin-fixed paraffin-embedded (FFPE) tissues are the global standard
46 format for preservation of the majority of biopsies in basic research and translational cancer
47 studies(Fox CH 1985), and it has been reported that more than 20 million FFPE specimens
48 are newly archived every year in the United States alone(Waldron et al. 2012). Accordingly,
49 profiling gene regulation in the archived FFPE tissue can be invaluable for translational
50 cancer research. Chromatin structure is still preserved during FFPE sample preparation and
51 long-term storage(Fanelli et al. 2010; Jin et al. 2015; Cejas et al. 2016). However, it has
52 proven difficult to apply the currently available highly sensitive chromatin accessibility
53 decoding technologies to FFPE tissue samples because of the high degree of DNA damage
54 that occurs during sequencing library preparation of these samples (Chin et al. 2020).
55 Moreover, it is desirable that a minimum number of FFPE tissue sections be used in the
56 analysis, as the tissues of interest are limited. The currently required input for chromatin
57 structure studies from FFPE samples is either couples of tissue sections or whole tissue
58 block (Fanelli et al. 2010; Jin et al. 2015; Cejas et al. 2016), and this precludes conducting
59 analyses at high resolution. To this end, we developed FFPE-ATAC, the first highly sensitive
60 method for decoding the chromatin accessibility in FFPE tissues, by combining the Tn5-
61 mediated transposition and T7 *in vitro* transcription.

62 RESULTS:

63 Standard ATAC-seq on FFPE samples

64 During formalin fixation, the formaldehyde in the formalin reacts with primary amines to form
65 Schiff bases, and with the amides to form hydroxymethyl compounds, resulting in the

66 formation of large chromatin complexes(Fox CH 1985). To decode the chromatin states in
67 the FFPE samples, it is essential to disrupt these chromatin complexes using reverse cross-
68 linking(Fanelli et al. 2010; Cejas et al. 2016). In standard ATAC-seq for live cells or frozen
69 tissues, accessible genomic sites are amplified and enriched through the polymerase chain
70 reaction (PCR) by using primers that hybridize with the universal Tn5 adaptors(Buenrostro et
71 al. 2013). In our previously established ATAC-seq (Chen et al. 2016) and Pi-ATAC(Chen et
72 al. 2018) technologies, we used a reverse cross-linking step to remove mild formaldehyde
73 cross-linking and performed ATAC-seq in the mildly fixed cells at the bulk and single-cell
74 levels. However, we learned that the reverse cross-linking step can cause a high degree of
75 DNA damage and introduce DNA breaks in extensively fixed cells and the FFPE tissues (**Fig.**
76 **1A**) (Martelotto et al. 2017). Furthermore, we assumed that if such DNA breaks occur at
77 accessible chromatin sites in FFPE tissues, and, if so, that this might hamper PCR
78 amplification of those accessible chromatin sites with the standard ATAC library preparation
79 strategy (**Fig. 1A**). To test our hypothesis, we developed an optimized protocol for the
80 isolation of high-quality nuclei from mouse liver and kidney FFPE tissue sections with 20 μ m
81 in thickness (**Supplemental Fig. S1A, S1B**, see **Methods**). Following the reverse cross-
82 linking strategy, we indeed observed many DNA breaks in the genomic DNA purified from
83 isolated FFPE nuclei (**Supplemental Fig. S1C**). In addition, only short fragments were
84 obtained when the standard ATAC-seq procedure was used on 50 000 nuclei isolated from
85 mouse FFPE liver and kidney tissues (**Supplemental Fig. S1D, S1E**; see **Methods**),
86 suggesting that DNA breaks indeed occur at accessible chromatin sites with long DNA
87 lengths and that this further hampers PCR amplification of those regions (**Fig. 1A**). We then
88 sequenced the libraries obtained through standard ATAC-seq on isolated FFPE nuclei
89 (**Supplemental Fig. S2A**), and prepared standard ATAC-seq libraries on frozen samples
90 collected from the same mouse liver and kidney samples as FFPE samples (**Supplemental**
91 **Fig. S2B**). Then, we compared the sequencing libraries obtained by standard ATAC-seq on
92 FFPE samples with those obtained by standard ATAC-seq on frozen samples (**Fig. 1B-I**,
93 **Supplemental Fig. S2, Supplemental Fig. S3, Supplemental Code**). This resulted in

94 several findings. **i)** The proportion of long DNA fragments (longer than 146 bp) obtained from
95 standard ATAC-seq on FFPE samples (30.76% +/- 1.38% for liver and 43.15% +/- 0.5% for
96 kidney) was lower than that obtained from standard ATAC-seq on frozen samples (50.02%
97 +/- 4.7% for liver and 59.17% +/- 2.57% for kidney) (**Fig. 1B, 1C**). Furthermore, the
98 proportion of mononucleosome fragments enriched at transcription start sites (TSS) was
99 also lower from the standard ATAC-seq on FFPE samples (**Supplemental Fig. S2C, S2D**).

100 **ii)** The library complexity obtained from standard ATAC-seq on FFPE samples was much
101 lower than that obtained through standard ATAC on frozen samples (**Fig. 1D, 1E**). **iii)** The
102 proportion of mitochondrial reads obtained from standard ATAC-seq on FFPE samples (27-
103 42%) was much higher than that obtained through standard ATAC-seq on frozen samples
104 (2-6%) (**Fig. 1F, 1G**). Since all of the ATAC-seq libraries were prepared from purified nuclei,
105 the sequencing libraries should contain very limited amounts of mitochondrial DNA. The high
106 proportion of mitochondrial reads obtained through standard ATAC-seq on FFPE samples
107 may be due to the fact that library complexity from genomic DNA in FFPE samples is low,
108 and PCR amplification enriches a high percentage of mitochondria. **iv)** High TSS
109 enrichment scores (score number: 27-30) and high number fraction of reads in peaks (FRiP)
110 (over 40%) were obtained from standard ATAC-seq on FFPE samples (**Fig. 1F, 1G**,
111 **Supplemental Fig. S2E, S2F**). Standard ATAC-seq on FFPE samples also showed good
112 genome-wide correlation with the results of standard ATAC-seq on frozen tissue (mouse
113 liver: $R = 0.87$, mouse kidney: $R = 0.85$, **Fig. 1H, 1I**), and the distribution of sequencing
114 reads in the genome from standard ATAC-seq on FFPE samples was similar to the
115 distribution obtained by standard ATAC-seq on frozen tissue (**Supplemental Fig. S2G, S2H**).

116 In addition, a large proportion of the peaks obtained by standard ATAC-seq by standard
117 ATAC-seq on FFPE samples and standard ATAC-seq on frozen tissue overlapped
118 (**Supplemental Fig. S3A, S3B**). Exclusive peaks from standard ATAC-seq on FFPE
119 samples and standard ATAC-seq on frozen tissue are distributed randomly in the genome
120 and display similar enrichments of transcription factors (**Supplemental Fig. S3A, S3B**). **v)**
121 However, we noticed that a proportion of the accessible regions are much more open in

122 frozen samples than in FFPE samples (**Fig. 1H, 1I**). On differential peak analysis (Log_2 (fold
123 change) > 3 , $p < 0.01$) (Supplemental Code)(Love et al. 2014), many more accessible
124 chromatin regions were identified in the frozen samples ($n = 1598$ in mouse liver and $n =$
125 495 in mouse kidney), but almost no more accessible chromatin regions were identified in
126 the FFPE samples ($n = 0$ in mouse liver and $n = 3$ in mouse kidney) (**Fig. 1H, 1I**),
127 suggesting that standard ATAC-seq on FFPE samples failed to detect a proportion of the
128 accessible chromatin sites. To further investigate whether the more accessible regions in
129 standard ATAC-seq on frozen samples represent sites at which DNA breaks occurred in the
130 FFPE samples, we calculated the number of sequencing reads obtained for those regions in
131 standard ATAC-seq on FFPE samples and found that for 66.33% (1060/1598) of those
132 regions in FFPE mouse liver and 55.77% (256/459) of those regions in FFPE mouse kidney,
133 no sequencing reads were detected (**Supplemental Table S1, Supplemental Table S2**).
134 This strongly suggests that DNA breaks potentially occur at those sites in FFPE samples
135 and further hamper PCR amplifications of them. We also noticed that the more accessible
136 regions in standard ATAC-seq on frozen samples were mainly located at regions distal (>10
137 kb) to the TSS (**Fig. 1H, 1I**).

138 Taken together, our results show that the transposase-mediated technology, ATAC-seq, can
139 be applied to FFPE samples consisting of nuclei isolated through an optimized procedure.
140 However, we learned that DNA breaks at accessible chromatin sites in FFPE samples
141 potentially hamper PCR amplification of these regions when standard ATAC-seq is used.
142 We concluded that standard ATAC-seq libraries on FFPE samples have lower library
143 complexity and a lower proportion of long DNA fragments, and lack a proportion of the
144 accessible chromatin sites compared with libraries prepared by standard ATAC-seq on
145 frozen samples.

146 **The design of FFPE-ATAC**

147 To increase the library complexity and rescue lost accessible regions in standard ATAC-seq
148 on FFPE samples, we developed FFPE-ATAC to decode chromatin accessibility in FFPE

149 tissues by combining Tn5-mediated transposition and T7 *in vitro* transcription (IVT) (**Fig. 1J**).

150 During Tn5 transposition in FFPE samples, Tn5 adaptors are inserted into the genome after

151 FFPE sample preparation; they are therefore unlikely to undergo the DNA breakage that

152 occurs during reverse cross-linking of FFPE samples and should therefore remain at the

153 ends of broken accessible chromatin sites after reverse cross-linking. We reasoned that by

154 adding a T7 promoter sequence to the Tn5 adaptor (**Fig. 1J, Supplemental Fig. S4A**) we

155 could use IVT to convert the two ends of the broken DNA fragments to RNA molecules

156 before preparing sequencing libraries from the IVT RNAs, and further decode the Tn5

157 adaptors' insertion sites in the genome (**Fig. 1J**). Through this strategy, we could decode the

158 flanking sequences of the accessible chromatin despite the fact that there were breaks

159 between adjacent pairs of T7-T5 adaptor insertion sites. It was found that Tn5 activity is very

160 robust, given the different sequence modifications on the Tn5 adaptor(Chen et al. 2016; Sos

161 et al. 2016; Chen et al. 2017; Xie et al. 2020; Payne et al. 2021). Thus, we designed,

162 produced, and optimized a Tn5 adaptor with an added T7 promoter sequence, termed T7-

163 Tn5 (**Supplemental Fig. S4A, see Methods**). T7-Tn5 retains the activity of the standard

164 Tn5 (**Supplemental Fig. S4B, see Methods**). To test our hypothesis that the T7-Tn5

165 adaptors remain at the ends of the accessible chromatin DNA fragments despite the DNA

166 breaks that result from reverse cross-linking, we performed IVT on single nuclei obtained

167 from FFPE samples of mouse liver and kidney after T7-Tn5 transposition. We found that

168 RNA fractions from these two FFPE tissues contained both short and long RNA

169 (**Supplemental Fig. S4C**). This result suggests that the T7 promoter is still present at the

170 ends of the broken accessible chromatin sites in the long-term fixed FFPE samples after

171 reverse cross-linking and that the insertion sites of T7-Tn5 adaptors in the genome could be

172 decoded in RNA molecules from IVT even when only one T7-Tn5 adaptor was present at the

173 end of the broken DNA molecules. Our results indicate that use of a combination of Tn5

174 transposition and T7 IVT could be of value for performing FFPE-ATAC and that it potentially

175 rescues broken DNA fragments in FFPE samples at accessible chromatin regions.

176 **Proof of concept of FFPE-ATAC with mouse FFPE liver and kidney samples**

177 Next, we proved the principle of FFPE-ATAC using sets of 500-50 000 nuclei purified from
178 individual FFPE tissue sections of mouse liver or mouse kidney sectioned at various
179 thicknesses (**Fig. 2A-M, Supplemental Fig.S5-11**).

180 First, we cut a mouse liver into two parts, one part was frozen, and the other was prepared
181 as an FFPE block (**Supplementary Fig. S5A, see Methods**). We performed FFPE-ATAC on
182 nuclei purified from frozen mouse liver and FFPE mouse liver (**Supplemental Fig. S5A, see**
183 **Methods**). Sequencing libraries obtained from frozen mouse liver by FFPE-ATAC had good
184 genome-wide reproducibility (**Supplemental Fig. S5B**). The sequencing reads of the
185 libraries were enriched at TSS (**Supplemental Fig. S5C**), but the TSS enrichment score was
186 1.5-2.5-fold lower than those of libraries obtained by standard ATAC-seq on frozen samples
187 (**Fig. 2B**). However, the sequencing library complexity obtained from FFPE-ATAC on frozen
188 mouse liver is much higher than that obtained from standard ATAC-seq on frozen mouse liver
189 (**Fig. 2C**). The reason for the lower complexity of standard ATAC-seq libraries compared
190 with FFPE-ATAC libraries is that standard ATAC-seq is a PCR-based method, and it
191 requires two correct pairs of Tn5 adaptor insertions (Buenrostro et al. 2013). One insertion
192 event or unpaired Tn5 adaptor insertions from Tn5 tagmentation could not be amplified
193 through PCR in standard ATAC-seq but could be captured with FFPE-ATAC. FFPE-ATAC
194 on frozen mouse liver and standard ATAC-seq on frozen mouse liver exhibited high similarity
195 at the level of chromatin accessibility at individual gene loci (**Fig. 2D**), and in the distribution
196 of sequence reads across the genome (**Supplemental Fig. S5D, S5E**). The two libraries
197 also showed good genome-wide correlation ($R = 0.72$, **Fig. 2E**) and displayed a large
198 number of overlapping ATAC peaks (53 043 overlapping peaks, **Supplemental Fig. S6A**).
199 Some differential peaks are detected between FFPE-ATAC on frozen mouse liver and
200 standard ATAC-seq on frozen mouse liver ($\text{Log}_2(\text{fold change}) > 3$, $p < 0.01$) (**Supplemental**
201 **Code**) ($n = 262$ in FFPE-ATAC, and $n = 1789$ in standard ATAC-seq, **Fig. 2E**), which
202 indicates that there are potentially different technical biases between FFPE-ATAC and

203 standard ATAC-seq. Our results suggested that FFPE-ATAC could accurately profile
204 chromatin accessibility in frozen samples with higher library complexity than standard ATAC-
205 seq. Next, compared the sequencing libraries obtained using FFPE-ATAC with FFPE mouse
206 livers and frozen mouse livers. We found high similarity at the level of library complexity (**Fig.**
207 **2B**), TSS enrichment score (**Fig. 2C**), chromatin accessibility at individual gene loci (**Fig. 2D**),
208 and sequence read distribution across the genome (**Supplemental Fig. S5D, S5E**). There
209 was also a good genome-wide correlation ($R = 0.75$, **Fig. 2F**), and a large number of
210 overlapping ATAC peaks (49530 overlapping peaks, **Supplemental Fig. S6B**). At the same
211 time, we found that the TSS enrichment scores obtained by FFPE-ATAC on frozen mouse
212 liver and FFPE mouse liver were similar to each other but 1.5- to 2.5-fold lower than the
213 scores obtained by standard ATAC-seq on frozen mouse liver. This could be due to the
214 different designs of FFPE-ATAC and standard ATAC-seq. Differential peak analysis showed
215 that only 95 more accessible chromatin regions were captured from FFPE-ATAC on frozen
216 mouse liver, but 969 more accessible chromatin regions were detected from FFPE-ATAC on
217 FFPE mouse liver (**Fig. 2F, Supplemental Table S3**). The similar levels of library complexity
218 obtained through FFPE-ATAC on FFPE mouse liver and FFPE-ATAC on frozen mouse liver
219 and the very limited number ($n = 95$) of more accessible chromatin regions detected from
220 FFPE-ATAC on frozen mouse liver suggest that FFPE-ATAC can potentially decode all
221 accessible chromatin sites in the genome by rescuing broken DNA fragments in FFPE
222 mouse liver. However, the FRiP from FFPE-ATAC on FFPE mouse liver (approximately 13%)
223 was much lower than that from FFPE-ATAC on frozen mouse liver (approximately 29%) (**Fig.**
224 **2B**); this could be due to the harsh chemical treatments used during the preparation of FFPE
225 samples. Finally, we compared the sequencing libraries obtained by FFPE-ATAC on FFPE
226 mouse liver and by standard ATAC-seq on FFPE mouse liver, and found that the library
227 complexity obtained from FFPE-ATAC was much higher than that obtained from standard
228 ATAC-seq (**Fig. 2C**). Even though there is high similarity between FFPE-ATAC on FFPE
229 mouse liver and standard ATAC-seq on FFPE mouse liver based on multiple comparisons
230 (**Fig. 2C-D, 2G, Supplemental Fig. S6C**), we identified 15 062 more accessible chromatin

231 regions in FFPE-ATAC on FFPE mouse liver, and these were mainly distributed in regions
232 distal to TSS (**Fig. 2G**). However, only 18 more accessible chromatin regions were detected
233 in standard ATAC-seq on FFPE mouse liver (**Fig. 2G**). We reasoned that if those large
234 numbers of more accessible regions in FFPE-ATAC on FFPE mouse liver are located at
235 sites of DNA breakage in FFPE mouse liver, no sequencing reads from those regions would
236 be detected in libraries prepared from FFPE mouse liver by standard ATAC-seq. Indeed,
237 among the more accessible regions identified through FFPE-ATAC on FFPE mouse liver,
238 71.83% (10819/15062) of those regions had no PCR amplicons in the libraries obtained by
239 standard ATAC-seq on FFPE mouse liver (**Supplemental Fig. S6D, Supplemental Table**
240 **S4**); this strongly indicates that FFPE-ATAC can be used to rescue accessible regions at
241 DNA breakage sites in FFPE mouse liver samples. Taken together, our results demonstrate
242 that the accessible chromatin profiles obtained using FFPE-ATAC on FFPE mouse liver are
243 very similar to the accessible chromatin profiles in frozen mouse liver. The strategy used in
244 FFPE-ATAC can thus rescue accessible regions that are lost due to DNA breaks when
245 standard ATAC-seq of FFPE samples is used, resulting in greater library complexity and
246 higher coverage of accessible chromatin profiles.

247 Second, following the same strategy that was used with mouse liver, we performed FFPE-
248 ATAC on frozen mouse kidney and on FFPE mouse kidney, and conducted cross-
249 comparisons among libraries prepared by FFPE-ATAC on FFPE mouse kidney, FFPE-ATAC
250 on frozen mouse kidney, standard ATAC-seq on FFPE mouse kidney and standard ATAC-
251 seq on frozen mouse kidney (**Fig. 2H-M, Supplemental Fig. S7, Supplemental Fig. S8,**
252 **and Supplemental Table S5-8**). We also obtained high-quality FFPE-ATAC results from
253 mouse FFPE kidneys (**Supplemental Fig. S7B-D**). The FFPE-ATAC on FFPE mouse
254 kidney and that on frozen mouse kidney from the same mouse kidney also exhibited high
255 similarity in library complexity (**Fig. 2I**), chromatin openness at the level of individual gene
256 loci (**Fig. 2J**) and genomic features of ATAC peaks (**Supplemental Fig. S7D, S7E**). There
257 was a good genome-wide correlation ($R = 0.81$, **Fig. 2L**) and a large number of overlapping

258 ATAC peaks (63 259 overlapping peaks, **Supplemental Fig. S8C**) in the results obtained
259 from FFPE-ATAC on FFPE mouse kidney and FFPE-ATAC on frozen mouse kidney. A very
260 limited number of differential peaks ($n = 19$ in FFPE-ATAC on frozen mouse kidney, $n = 8$ in
261 FFPE-ATAC of FFPE mouse kidney, **Fig. 2 L**) between FFPE-ATAC on frozen mouse
262 kidney and that on FFPE mouse kidney were identified, indicating that the chromatin profiles
263 captured with FFPE-ATAC on FFPE mouse kidney are very similar to those captured with
264 FFPE-ATAC on frozen mouse kidney. Differential peak analysis of FFPE-ATAC on FFPE
265 mouse kidneys and standard ATAC-seq on FFPE mouse kidney showed that 3886 more
266 accessible chromatin regions were decoded in FFPE-ATAC on FFPE mouse kidney
267 (Supplemental Code), whereas only 541 more accessible chromatin regions were captured
268 in standard ATAC-seq on FFPE mouse kidney (**Fig. 2 M**). For 61.65% (2396/3886) of the
269 more accessible chromatin regions captured in FFPE-ATAC on FFPE mouse kidney, no
270 sequencing reads were detected in those regions from libraries obtained by standard ATAC-
271 seq on FFPE mouse kidney (**Supplemental Fig. S8D, Supplemental Table S8**). These
272 results further demonstrate that FFPE-ATAC can profile accessible chromatin with better
273 library complexity and rescue accessible regions at sites of DNA breakage in FFPE samples
274 compared with standard ATAC-seq on FFPE samples. Analysis of FFPE-ATAC libraries
275 generated from both FFPE mouse liver and FFPE mouse kidney revealed a large number of
276 peaks that overlap with the peaks listed in the Encyclopedia of DNA Elements (ENCODE)
277 mouse liver or kidney DNase-seq; there were 39 378 overlapping peaks for mouse liver
278 (**Supplemental Fig. S9A**) and 64 612 overlapping peaks for mouse kidney (**Supplemental**
279 **Fig. S9B**).

280 Third, we tested the sensitivity of FFPE-ATAC using various numbers of nuclei (ranging from
281 500 to 50 000) purified from FFPE mouse kidney tissue (**Supplemental Fig. S10A**, see
282 **Methods**) Based on a comprehensive comparison of chromatin accessibility obtained using
283 50 000 nuclei, including TSS enrichment scores (**Supplemental Fig. S10B**), FRiP values
284 (**Supplemental Fig. S10B**), library complexity (**Supplemental Fig. S10C**), genome-wide

285 correlation (**Supplemental Fig. S10D-F**), and sequencing read distribution across the
286 genome (**Supplemental Fig. S10G, S10H**), we concluded that FFPE-ATAC resulted in good
287 accessible chromatin profiles of FFPE samples when as few as 500 nuclei were used.

288 Fourth, we determined the minimum thickness of FFPE tissue sections needed for the
289 FFPE-ATAC by performing the FFPE-ATAC with 50 000 nuclei isolated from the 5-, 7-, and
290 10- μ m thick mouse FFPE kidney tissue sections (**Supplemental Fig. S11A, S11B**; see
291 **Methods**). The diameter of a mammalian cell nucleus is 6-10 μ m (Webster et al. 2009),
292 whereas the FFPE tissue sections used in routine clinical practice are 4-50 μ m thick. We
293 therefore investigated whether satisfactory FFPE-ATAC results could be obtained using
294 FFPE tissue sections of different thicknesses. We found that the TSS enrichment score,
295 library complexity and other parameters of the libraries obtained from mouse kidney FFPE-
296 ATAC remained adequate when 5 μ m thick tissue sections were used (**Supplemental Fig.**
297 **S11C-F**, see **Methods**). However, FRiP values of FFPE-ATAC libraries obtained from 5-, 7-,
298 and 10- μ m-thick mouse FFPE kidney tissue sections, ranging from 2.4% to 7.5%
299 (**Supplementary Fig. 11C**), were all lower than that of the FFPE-ATAC library obtained from
300 a 20- μ m-thick mouse FFPE kidney tissue section (~11%). In addition, the total number of
301 accessible peaks in the libraries prepared from these thin sections was much lower than the
302 number of accessible peaks in the libraries prepared from 20- μ m-thick mouse FFPE kidney
303 tissue sections (**Supplemental Fig. S11G**). Since the diameter of the mammalian nucleus is
304 6-10 μ m (Webster et al. 2009), we reasoned that nuclei isolated from 5- to 10- μ m-thick
305 FFPE tissue sections contain a large proportion of nonintact nuclei. We suspected that the
306 structure of the chromatin in nonintact nuclei could be affected during the isolation procedure,
307 resulting in low-quality accessible chromatin profiles. Thus, we concluded that FFPE tissue
308 sections with thickness greater than the diameter of nucleus should be used in FFPE-ATAC.

309 Taken together, accurate mapping of the accessible genome from mouse FFPE liver and
310 kidney tissue sections demonstrates that FFPE-ATAC can be used to identify the accessible
311 chromatin landscape using low number of cells obtained from FFPE tissue sections.

312 **Use of combination of FFPE-ATAC and H&E staining to decipher chromatin**
313 **accessibility in a region of interest in FFPE tissue sections**

314 Next, we deciphered chromatin accessibility in the mouse cerebellum by using hematoxylin
315 and eosin (H&E) staining to identify mouse cerebellum in FFPE tissue sections of the mouse
316 brain (**Fig. 3A-D**). H&E staining is a standard method that is used in clinical diagnostics to
317 facilitate the assessment of tumor morphology and composition (Martina et al. 2011). We
318 used H&E staining of a 5- μ m-thick FFPE mouse brain tissue section to find the location of
319 the cerebellum; we then isolated the cerebellar region from the immediately adjacent 20- μ m-
320 thick FFPE mouse brain tissue section, and purified the nuclei from the isolated cerebellum
321 for use in FFPE-ATAC (**Fig. 3A, Supplemental Fig. S12A, see Methods**). The resulting
322 FFPE-ATAC profiles of the mouse cerebellum had good TSS enrichment scores, FRiP
323 values (**Supplemental Fig. S12B**), library complexity (**Supplemental Fig. S12C**) and
324 genomic features (**Supplemental Fig. S12D, S12E**). The chromatin accessibility of
325 regulatory elements of cerebellum-specific genes such as *Gabrb2*, was high (**Fig. 3B**). The
326 technical replicates for FFPE-ATAC libraries from the cerebellum exhibited good
327 reproducibility of genome-wide correlation, showing with numerous overlapping peaks ($R =$
328 0.86, 58 277 overlapping ATAC-seq peaks, **Fig. 3C**). Gene Ontology term enrichment and
329 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the top 10 000
330 FFPE-ATAC peaks identified major terms and pathways that clearly represent relevant gene
331 pathways of the cerebellum (Sato et al. 2008) (**Fig. 3D, Supplemental Fig. S12F**).

332 **Application of FFPE-ATAC to clinically archived FFPE samples**

333 Finally, we applied the FFPE-ATAC method to colorectal cancer (CRC) FFPE tissue
334 sections from seven patients, including two cases of rectal cancer and five cases of colon
335 cancer (**Fig. 4A-G, Supplemental Fig. S13A, S13B, Supplemental Table S9**). These CRC
336 FFPE tissue blocks had been preserved for 6 to 10 years (**Supplemental Table S9**). The
337 FFPE-ATAC libraries obtained from these CRC samples had good reproducibility

338 **(Supplemental Fig. S13C, S13D, R ranged from 0.86 to 0.97)**, good TSS enrichment
339 scores **(Supplemental Fig. S14A)**, library complexity **(Supplemental Fig. S14B)**, and
340 similar distributions of genomic features **(Supplemental Fig. S14C, S14D)**. However,
341 diverse ranges of FRiP, ranging from 5.78% to 17.74%, was observed in the libraries from
342 these clinical samples **(Supplemental Fig. S14A)**; this could be due to variation in the
343 procedures used for FFPE sample preparation. When we derived nonnegative matrix
344 factorization (NMF) clusters using all the 7 CRC FFPE-ATAC peaks (Brunet et al. 2004), we
345 found that two clusters were the best to characterize the 7 cases of CRC **(Fig. 4B,**
346 **Supplemental Fig. S15A)**; samples from three of the colon cancer patients were in cluster 1,
347 while samples from the two rectal cancer patients and samples from the other two colon
348 cancer patients were in cluster 2. The promoter regions of the CRC-specific gene marker
349 *LRCH4* (Uhlen et al. 2015), were open in both cluster 1 and cluster 2 **(Fig. 4C)**. Comparing
350 the open chromatin sites within these two clusters, we identified 4 186 unique ATAC peaks
351 for cluster 1 and 4 392 unique ATAC peaks for cluster 2 (fold change > 2, false discovery
352 rate < 0.01, **Fig. 4D, 4E, Supplemental Fig. S15B, S15C, Supplemental Table S10,**
353 **Supplemental Table S11, Supplemental Code)**. We also found that the unique regulatory
354 elements in these two clusters had similar genomic features **(Supplemental Fig. S15D,**
355 **S15E)**, but the ranking of transcription factors (TFs) enriched in the cluster-specific peaks
356 are different between the two clusters **(Fig. 4F, 4G, Supplemental Table S12,**
357 **Supplemental Table S13)**; the top-ranking TFs for cluster 1 were ZIC1, TAL1, and NANOG,
358 while the top-ranking TFs for cluster 2 were FOSL2, FOSL1 and JUN. It has been reported
359 that AP-1 TFs play a dominant role in the progression of CRC (Ashida et al. 2005). We found
360 8 of the top 10 enriched TFs in cluster 2 are all from the AP-1 TF family (**Supplemental Fig.**
361 **S15F)**. A similarly high enrichment of AP-1 TF family members was not observed in cluster 1,
362 likely reflecting a role of AP-1 TFs in some cases of CRC but not in others.

363 In summary, FFPE-ATAC allows the profiling of chromatin accessibility in specific regions of
364 interest when combined with the use of H&E staining to identify the cell analyzed. This

365 approach serves to identify unique distal regulatory elements and TF enrichments using low
366 numbers of nuclei prepared from single clinically archived FFPE tissue sections preserved
367 for extended periods of time.

368 **DISCUSSION:**

369 FFPE tissue samples represent a large source of materials for epigenetic analysis in both
370 basic research and clinical translational studies(Gaffney et al. 2018), but such samples have
371 not been widely used in chromatin studies to date due to the lack of sufficiently sensitive
372 techniques. The broad application of ATAC-seq in biomedical research has offered us a
373 potential strategy for profiling chromatin accessibility in FFPE samples with high sensitivity
374 (Buenrostro et al. 2013; Buenrostro et al. 2015; Cusanovich et al. 2015; Chen et al. 2016;
375 Corces et al. 2017; Corces et al. 2018). However, the presence of DNA damage in FFPE
376 samples hampers the direct application of standard ATAC-seq to these samples(Chin et al.
377 2020). Using an optimized nuclei isolation protocol with FFPE tissue sections, we showed
378 that transposase-mediated technology, ATAC-seq, could be applied to FFPE samples.
379 However, standard ATAC-seq libraries of FFPE samples have lower library complexity and a
380 smaller proportion of long DNA fragments, and lack a proportion of accessible chromatin
381 sites compared with libraries obtained by standard ATAC-seq on frozen samples. To
382 increase library complexity and rescue accessible regions that are lost in standard ATAC-
383 seq on FFPE samples, we developed FFPE-ATAC, which used a combination of Tn5-
384 mediated transposition and T7 IVT to decode chromatin accessibility in FFPE tissues. We
385 demonstrated that the accessible chromatin profiles derived from FFPE samples by FFPE-
386 ATAC are very similar to the accessible chromatin profiles of frozen samples. We learned
387 that the TSS enrichment scores obtained after FFPE-ATAC of frozen samples and FFPE
388 samples are similar to each other but 1.5-2.5-fold lower than those obtained through
389 standard ATAC-seq of frozen samples; this could be due to the different designs of FFPE-
390 ATAC and standard ATAC-seq. At the same time, we observed that the proportion of
391 sequencing signals in peaks from FFPE-ATAC of FFPE samples was lower than the

392 proportion in peaks from FFPE-ATAC of frozen samples but fell in a range similar to that in
393 peaks from standard ATAC-seq of frozen samples; this could be due to the use of harsh
394 chemical treatments during preparation of the FFPE samples. FFPE-ATAC is more labor-
395 intensive than the more simply designed standard ATAC-seq method. However, through use
396 of the FFPE-ATAC strategy, it was possible to rescue many of the accessible regions that
397 are lost in standard ATAC-seq on FFPE samples, resulting in better library complexity. The
398 better library complexity and higher coverage of accessible chromatin profiles that can be
399 obtained using FFPE-ATAC on FFPE samples compared with standard ATAC-seq of FFPE
400 samples will be valuable for accessible chromatin profiling of clinically archived FFPE
401 materials.

402 We demonstrate here that FFPE-ATAC is a robust tool that can be used to decode
403 chromatin accessibility with high sensitivity using 500-50 000 nuclei prepared from single
404 FFPE tissue sections. The use of a combination of FFPE-ATAC and H&E staining to
405 decipher chromatin accessibility in a region of interest in FFPE tissue sections and
406 successful profiling of disease-associated chromatin regulation from the clinically archived
407 human colorectal cancer FFPE samples make FFPE-ATAC a powerful tool for use in
408 preclinical studies and precision medicine. In addition, FFPE-ATAC can potentially be used
409 to extend our current understanding of the cancer epigenome and in pathological diagnosis
410 through combination of other omics data obtained from the same FFPE materials with
411 clinicopathological records. FFPE-ATAC can find broad applications both in basic research
412 and in clinical settings. In the future, it will be of great interest to extend the resolution of
413 FFPE-ATAC to the single-cell level.

414

415

416 **METHODS**

417 **Nuclei isolation from FFPE tissue sections:** Mouse FFPE kidney, liver, and brain tissue
418 blocks were sectioned into 5-, 7-, 10-, and 20- μm thick section using a microtome. The
419 human CRC samples were cut into 10 μm thick section. One curved tissue section was
420 deparaffined with 1ml of xylene (HistoLab, 02070) 5mins, twice. Rehydration was done by
421 sequential ethanol washes, started with 100% ethanol 5 min twice, 95%, 70%, 50%, 30%
422 ethanol, 5 min each. After deparaffinization and rehydration, tissue was washed with 1 ml
423 water, then 1ml 0.5mM CaCl_2 (Alfa Aesar, J63122). Then the tissue was subjected to
424 microdissection under a stereo microscope first, then centrifuged at 3000g for 10 mins. After
425 centrifugation, supernatant was removed and 1ml enzymatic cocktail (3mg/ml of
426 Collagenase (Sigma-Aldrich, C9263) and 300 U/ml of hyaluronidase (Merk Millipore,
427 HX0154-1)) was added to the tissue pellet. Then the mixture was incubated at 37°C for 16
428 hours by adding 100 μg of Ampicillin (Serva, 69-52-3) and 50 μg of sodium azide (Merck
429 Millipore,26628-22-8). After the enzyme digestion, 400 μl NST buffer (146mM NaCl
430 (Invitrogen, 00648496), 10mM Tris pH 7.8 (Invitrogen, 15568-025), 1mM CaCl_2 (Alfa Aesar,
431 J63122), 21mM of MgCl_2 (Invitrogen, AM9530G), 0.05% BSA (Miltenyi Biotech MACS, 130-
432 091-376), 0.2% Igepal CA-360 (Sigma-Aldrich, 13021-50)) was added to the mixture, and
433 the tube was centrifuged at 2800g for 10 minutes. After the centrifugation, the supernatant
434 was aspirated and discarded, then the pellet was resuspended in 800 μl NST buffer
435 containing 0.1% DNase free RNase A (Thermo Scientific, EN0531), and 10% fetal bovine
436 serum (Life Technologies,10108-105). The mixture was passed through the 27G needle
437 syringe 30 times and filtered with a 30 μM filter (MiltenyiBiotechMacs,130-098-458). Then
438 the passed-through nuclear suspension was centrifugation at 2800 g for 10 minutes, and the
439 nuclei were resuspended in PBS, checked and counted.

440

441 **Nuclei isolation from the Mouse cerebellum FFPE tissue section:** The mouse
442 cerebellum area was identified with H&E staining from the adjacent tissue sections, and
443 labeled with marker pen under the stereo microscope. The tissues from the cerebellum area

444 were moved to the Eppendorf tube and the nuclei isolation from the selected area was
445 performed with the protocol stated as above.

446

447 **Human CRC sample collections and FFPE block preparation:** The regional ethical
448 research committee at the Uppsala University approved the study (Dnr 2015/419 and
449 2018/490). The FFPE tissue blocks of CRC were prepared at the Dept. of Clinical Pathology,
450 Uppsala University Hospital, Uppsala, Sweden, according to standard procedures. Briefly,
451 tissue from surgical specimens of colon and rectal samples were fixed in buffered formalin
452 for 24-72 hours. The pieces were then examined by a pathologist, excised and placed in
453 plastic cassettes. The fixed tissue was then dehydrated in an automated system (Tissue-
454 Tek® VIP®) where the tissue was immersed in ethanol of varying concentrations (70%, 95%,
455 99.5%) followed by xylene and finally paraffin (Histowax™, Histolab) over a period of
456 approximately 12 hours. Finally, the paraffin embedded tissue piece was oriented in a
457 cassette, liquid paraffin was poured over it and allowed to set, forming the FFPE block. The
458 FFPE block was then sectioned on a microtome at a thickness of 10- μ m.

459

460 **FFPE-ATAC on FFPE tissue and frozen tissue:** For FFPE-ATAC on FFPE tissue, 500-50
461 000 isolated FFPE nuclei were used in each FFPE-ATAC reaction, where nuclei were
462 isolated following nuclei isolation protocol stated in section of nuclei isolation from FFPE
463 tissue sections. For FFPE-ATAC on frozen tissue, 50 000 isolated nuclei were used in each
464 reaction, where nuclei were isolated following nuclei isolation protocol in section of standard
465 ATAC-seq on frozen tissue. In brief, nuclei were counted using the cell counter and pelleted
466 at 2800g for 10 minutes at room temperature. 50 μ l of lysis buffer (10mM Tris-HCl pH7.4
467 (Invitrogen, 15567-027), 10mM NaCl (Invitrogen, AM9759), 3mM MgCl₂ (Invitrogen,
468 AM9530G), 0.1% Igepal CA-360 (Sigma-Aldrich, 13021-50)) was added to the nuclei pellet
469 and the nuclei suspension was immediately centrifuged at 2800g for 10 minutes at room
470 temperature. After the supernatant was discarded, the nuclei pellet was resuspended in 50
471 μ l of transposase master mixture (25 μ l of 2X TD buffer (20mM Tris-HCl pH 7.6 (Invitrogen,

472 15568-025), 10mM MgCl₂ (Invitrogen, AM9530G), 20%Dimethyl Formamide), 22.5 μl
473 Nuclease free water (Invitrogen, AM9932) and 2.5μl of 2μM T7-Tn5), then incubated at 37°C
474 for 30 minutes. After the incubation, 50 μl of 2X reverse crosslinking solution (100mM Tris-Cl
475 PH 8.0 (Invitrogen, 15568-025), 2mM EDTA (Invitrogen, AM9290G), 2% SDS (Invitrogen,
476 15553-035), 0.4M NaCl (Invitrogen, AM9759)) and 10ng/ul Proteinase K (Thermo Scientific,
477 EO0491) was directly added into the tagmentation reaction mixture, then mixture was
478 incubated overnight at 65°C with 1200rpm shaking. Next day, the incubation mixture is
479 purified with MinElute PCR Purification kit (QIAGEN, 28004) and DNA is eluted in 20μl of
480 elution buffer, then 20 μl of 2X PCR master mix (New England Biolabs, New England
481 Biolabs, M0541S) was added to the samples. The mixture was incubated in a thermocycler
482 at 72°C for 5 minutes. The sample was first purified with MinElute PCR Purification kit
483 (QIAGEN, 28004), then re-purified with SPRI beads with 1:1 ratio (Beckman Coulter,
484 B23317), and eluted in 25 μl of water (Invitrogen, AM9932).

485

486 Next, *In Vitro* Transcription (IVT) was performed with T7 high yield RNA synthesis kit (New
487 England Biolabs, E2040S). The RNA from the IVT was purified using TRIzol first (Ambion,
488 15596026), then ZYMO RNA Clean & Concentration kit (Zymo, R1013). Next, 1 μl DNase I
489 (New England Biolabs, M0303L) was added into the RNA and the mixture was incubated for
490 15 minutes at 37°C. The RNA was purified with the ZYMO RNA Clean & Concentration kit
491 (Zymo, R1013) again and eluted in 15 μl of nuclease-free water. The IVT RNA is transferred
492 into cDNA with random primers with SMART MMLV kit by following the manufactory protocol
493 (TaKaRa, 639524). 100ng RNA was used for each library preparation. In brief, the mixture
494 was incubated at 42°C for 60 minutes and 70°C for 15 minutes, then 2 μl of RNase H buffer
495 and 0.2 μl RNase H enzyme (Thermo Scientific, EN0201) was added and incubated at 37°C
496 for 20 minutes. The cDNA was purified using RNA XP beads purification with 1:1.8x ratio of
497 sample to beads (Beckman Coulter, A63987) and eluted in 24.5 μl water. Next, the cDNA
498 was converted to double-stranded DNA with pre-PCR (98°C for 10 seconds, 63°C for 30

499 seconds, 72°C for 1 minute, 10°C hold in one cycle) by adding 25 μ l of 2X PCR master mix
500 (New England Biolabs, M0541S) and 0.8 μ l of Ad 2.X reverse primer. Then sample was
501 purified using MinElute PCR Purification kit (QIAGEN, 28004) and eluted in 20 μ l water.

502

503 The sequencing library was prepared with standard Tn5 tagmentation. In short, the double-
504 strand DNA samples were subjected to the tagmentation by adding 25 μ l of 2X TD-Buffer
505 (20mM Tris-HCl pH 7.6 (Invitrogen, 15567-027), 10mM MgCl₂ (Invitrogen, AM9530G), 20%
506 Dimethyl Formamide), 0.5 μ l 2uM standard Tn5, 4.5 μ l nuclease-free water (Invitrogen,
507 AM9932) and incubated at 55°C for seven minutes, then samples were purified using Qiagen
508 MinElute PCR Purification kit (QIAGEN, 28004) and eluted in 20 μ l elution buffer. The library
509 amplification PCR was performed by adding 25 μ l of 2X PCR master mix, 0.4 μ l of barcodes
510 forward primer i5 25 μ M, 0.4 μ l of barcodes reverse Primer i7 25 μ M, 4.2 μ l of nuclease-free
511 water to the sample, with the following PCR protocol (72 °C 5mins first, 20 cycles of 98°C for
512 10 seconds, 63°C for 30 seconds, 72°C for 1 minute), then sample was purified using
513 Qiagen MinElute PCR Purification kit (QIAGEN, 28004) and eluted in 20 μ l. At last, the DNA
514 library with the length of 220-1000bp was selected with PAGE gel purification for sequencing.
515 The FFPE-ATAC libraries were sequenced on Illumina NovaSeq 6000, and at least 40
516 million 150bp single end or paired end sequencing reads were generated for each library.

517

518 **Hyperactive Tn5 transposase production:** Hyperactive Tn5 was produced as previously
519 described(Picelli et al. 2014). In brief, pTXB1-Tn5 plasmid (Addgene, 60240) was
520 introduced into T7 Express LysY/Iq *E. coli* strain (NEB, C3013). 10 ml of overnight cultured *E.*
521 *coli* was inoculated to 500 ml LB medium. After incubation for 1.5 hrs at 37°C, bacteria was
522 incubated about 2.5 hrs at room temperature. When the OD600 = 0.9, Tn5 protein was
523 induced by adding 0.25 mM IPTG for 4 hrs. *E. coli* pellet was resuspended in lysis buffer (20
524 mM HEPES-KOH pH 7.2, 0.8 M NaCl, 1mM EDTA, 10% glycerol, 0.2% Triton X-100,
525 complete proteinase inhibitor (Roche,11697498001)) and lysed by sonication. 10% PEI was

526 added to supernatant of lysate to remove bacterial genomic DNA. 10 ml chitin resin (NEB,
527 S6651L) was added to the supernatant and incubated with rotating for 1 hr at 4°C. The resin
528 washed by lysis buffer extensively. In order to cleave Tn5 protein from intein, lysis buffer
529 containing 100 mM DTT was added to the resin and stored in 4°C. After 48 hrs, protein was
530 eluted by gravity flow and collected in 1ml fractions. 1 μ l of each fraction was added to
531 detergent compatible Bradford assay (Thermo Fisher Scientific, 23246) and peaked fractions
532 were pooled and dialyzed against 2X dialysis buffer (100 mM HEPE-KOH pH7.2, 0.2 M NaCl,
533 0.2 mM EDTA, 2 mM DTT, 0.2% Triton X-100, 20% glycerol). Dialyzed Tn5 protein was
534 concentrated by using ultracel 30-K column (Millipore, UFC903024) and the quantity of Tn5
535 was measured by Bradford assay and visualized on NuPAGE Novex 4-12% Bis-Tris gel
536 (Thermo Fisher Scientific, NP0321) followed by Coomassie blue staining.

537

538 **T7-Tn5 and Tn5 adaptor sequences:** The oligonucleotides for Tn5 and T7-Tn5
539 transposase adaptor were synthesized at INTERGATED DNA TECHNOLOGIES (IDT), and
540 the sequences of oligonucleotide are as follows:

541 Tn5MErev, 5'-[phos]CTGTCTCTTATACACATCT-3';

542 T7-Tn5ME: 5'/CAT GAG ATT AAT ACG ACT CAC TAT AGG GAG AAG ATG TGT ATA
543 AGA GAC AG-3';

544 Tn5ME-A: 5' - TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3';

545 Tn5ME-B: 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'.

546 **PCR primer sequences:** The PCR primers were synthesized at INTERGATED DNA
547 TECHNOLOGIES (IDT), and the sequences of primers were used by referring to the
548 previous report (Buenrostro et al. 2015).

549

550 **Tn5 and T7-Tn5 transposase assembly:** The assembly of Tn5 and T7-Tn5 transposase
551 were performed as described (Picelli et al. 2014). Briefly, oligonucleotides (T7-Tn5ME,
552 Tn5MErev, Tn5ME-A, Tn5ME-B) were resuspended in water to a final concentration of 100

553 μ M each. Equimolar amounts of Tn5MErev/Tn5ME-T7, Tn5MErev/Tn5ME-A and
554 Tn5MErev/Tn5ME-B were mixed in separate 200 μ l PCR tubes. These oligo mixtures were
555 denatured on a thermocycler for 5 min at 95°C and cooled down slowly on the thermocycler
556 by turning off the thermocycler. The T7-Tn5 transposase was assembled with the following
557 components: 0.25 vol Tn5MErev/Tn5ME-T7 (final concentration of each double-strand oligo
558 is now 50 μ M each), 0.4 vol glycerol (100% solution), 0.12 vol 2X dialysis buffer (100 mM
559 HEPES-KOH at pH 7.2, 0.2 M NaCl (Invitrogen, AM9759), 0.2 mM EDTA (Invitrogen,
560 AM9290G), 2 mM DTT, 0.2% Triton X-100 (Sigma-Aldrich, T8787), 20% glycerol (Sigma-
561 Aldrich, G9012-500)), 0.1 vol SL-Tn5 (50 μ M), 0.13 vol water. The reagents were mixed
562 thoroughly but gently, and the solution was left on the bench at room temperature for 1 h to
563 allow annealing of oligos to Tn5. The Tn5 transposase was assembled with same procedure
564 as T7-Tn5 transposase but with following oligos: 0.25 vol Tn5MErev/Tn5ME-A and 0.25 vol
565 Tn5MErev/Tn5ME-B.

566

567 **T7-Tn5 transposase activity assay:** The activity of the assembled T7-Tn5 and Tn5
568 transposase was checked as described below. The mixture of 10 μ l of 2X TD buffer (20mM
569 Tris-HCl pH 7.6 (Invitrogen, 15568-025), 10mM MgCl₂ (Invitrogen, AM9530G), 20%Dimethyl
570 Formamide), 50ng human genomic DNA (Promega, G304A), 2 μ M assembled T7-Tn5
571 transposase or Tn5 transposase was incubated at 55 °C for 7minutes. After incubation, the
572 mixture was purified by Qiagen MinElute PCR Purification kit (QIAGEN, 28004) and eluted in
573 10 μ l of elution buffer. Then eluted DNA was mixed with 2 μ l 6X loading dye and ran on a 1.2%
574 agarose gel to check the length distribution of the DNA.

575

576 **Standard ATAC-seq on FFPE samples:** 50, 000 isolated FFPE nuclei (mouse liver and
577 mouse kidney) were used in each reaction following standard ATAC-seq protocol as previous
578 reported (Buenrostro et al. 2013). The reverse-crosslinking was used after Tn5 tagmentation
579 following the protocol of ATAC-seq in fixed cells (Chen et al. 2016). Briefly, 50,000 cells

580 were centrifuged 500 g 5 min at room temperature. The cell pellet was resuspended in 50 μ l
581 lysis buffer (10 mM Tris-Cl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.01% Igepal CA-630) and
582 centrifuged immediately 500 g for 10 min at 4 °C. The cell pellet was resuspended in 50 μ l
583 transposase mixture (25 μ l 2X TD buffer, 22.5 μ l dH₂O and 2.5 μ l Tn5 transposase) and
584 incubated at 37 °C 30 min. After the transposase reaction, a reverse crosslink solution was
585 added (with final concentration of 50 mM Tris-Cl, 1mM EDTA, 1% SDS, 0.2M NaCl, 5 ng/ml
586 Proteinase K) up to 200 μ l. The mixture was incubated at 65 °C with 1000 rpm shaking in a
587 heat block overnight, then purified with Qiagen Mini-purification kit and eluted in 10 μ l
588 Quiagen EB elution buffer. Sequencing libraries were prepared following the original ATAC-
589 seq protocol (Buenrostro et al. 2013).

590

591 **Standard ATAC-seq on frozen samples:** Single nuclei were isolated from frozen tissue
592 with Dounce homogenization by following the nuclei isolation protocol in Omni-ATAC(Corces
593 et al. 2017). In brief, green bean size frozen tissue incubated in the ice-cold 800 μ l of 1X
594 homogenization unstable buffer(5 mM CaCl₂ (Alfa Aesar, J63122), 3 mM Mg(Ac)₂ (Sigma-
595 Aldrich, M5661), 10 mM Tris pH 7.8 (Invitrogen, 15568-025), 0.01667 mM PMSF (Sigma-
596 Aldrich, P7626), 0.1667 mM β -mercaptoethanol (Sigma-Aldrich, M-6250), 320 mM Sucrose
597 (Sigma-Aldrich, 84097-250), 0.1mM EDTA (Invitrogen, AM9290G), 0.1% Igepal CA-630
598 (Sigma-Aldrich, 13021-50)) for 5 minutes on ice. . Tissue was homogenized through 10
599 strokes with a loose pestle and 20 strokes with a tight pestle, then 400 μ l of the
600 homogenized sample was mixed with 400 μ l of 50% OptiPrep Density Gradient Medium
601 (Sigma-Aldrich, D1556-250), to make a final concentration of 25% of OptiPrep Density
602 Gradient Medium (Sigma-Aldrich, D1556-250) with homogenized tissue. After preparation of
603 tissue mixture, a fresh 2ml low binding vial was taken and layered 35% of OptiPrep Density
604 Gradient Medium (Sigma-Aldrich, D1556-250), 29% of OptiPrep Density Gradient Medium
605 (Sigma-Aldrich, D1556-250), and 25% of OptiPrep Density Gradient Medium (Sigma-Aldrich,
606 D1556-250) mixed with the sample, on the top of each other. The layered vial was
607 centrifuged at 3000g for 20 minutes at 4°C. After gradient centrifugation, the top 1300 μ l was

608 discarded and the 200 μ l of the nuclei region was carefully collected in a fresh vial. Then 800
609 μ l of ice cold PBS was added and centrifuged at 500g for 10 mins, followed by resuspended
610 in ice cold 500 μ l PBS. 50,000 nuclei were used for each reaction and prepared library by
611 using standard ATAC protocol as stated in the section of standard ATAC-seq on FFPE
612 samples (Buenrostro et al. 2013). The components for the solutions are as follows: 6X
613 Homogenization Buffer Stable Master Mix: 30mM CaCl₂ (Alfa Aesar, J63122), 18mM
614 Mg(Ac)₂, 60mM Tris-HCl pH 7.8 (Invitrogen, 15568-025); 6X Homogenization Buffer
615 Unstable Solution: 6X Homogenization Buffer Stable Master Mix, 0.1mM PMSF, 1mM β -
616 mercaptoethanol (Sigma-Aldrich, M-6250); 1X Homogenization Buffer Unstable Solution:
617 1X Homogenization Buffer Stable Master Mix, 320mM Sucrose (Sigma-Aldrich, 84097-250),
618 0.1mM EDTA (Invitrogen, AM9290G), 0.1% Igepal CA-360 (Sigma-Aldrich, 13021-50); 50%
619 OptiPrep Density Gradient Medium (Sigma-Aldrich, D1556-250) Solution:
620 1X Homogenization Buffer Stable Master Mix, 50% OptiPrep Density Gradient Medium
621 (Sigma-Aldrich, D1556-250) Solution; 29% OptiPrep Density Gradient Medium (Sigma-
622 Aldrich, D1556-250) Solution: 1X Homogenization Buffer Stable Master Mix, 160mM sucrose,
623 29% OptiPrep Density Gradient Medium (Sigma-Aldrich, D1556-250) Solution; 35%
624 OptiPrep Density Gradient Medium (Sigma-Aldrich, D1556-250) Solution:
625 1X Homogenization Buffer Stable Master Mix, 160mM sucrose (Sigma-Aldrich, 84097-250),
626 35% OptiPrep Density Gradient Medium (Sigma-Aldrich, D1556-250) Solution. The ATAC-
627 seq libraries were sequenced on Illumina NovaSeq 6000, and at least 20 million 150bp
628 paired-end sequencing reads were generated for each library.

629

630 **Genomic DNA purification from frozen and FFPE tissue nuclei:**

631 For FFPE-ATAC samples, single nuclei were isolated following nuclei isolation protocol
632 stated in section of nuclei isolation from FFPE tissue sections. For frozen samples, nuclei
633 were isolated following nuclei isolation protocol in section of standard ATAC-seq on frozen
634 tissue. For genomic DNA purification, 1 million isolated nuclei were spined down at 3000 g
635 for 10 mins, then resuspended with 100 μ l of lysis buffer (50 mM Tris-HCl pH=7.5 (Invitrogen,

636 15567027), 1 mM EDTA (Invitrogen, AM9260G), 1% SDS (Invitrogen, 1553-035), 200 mM
637 NaCl (Invitrogen, AM9759) and 200 μ g/mL Proteinase K (Thermo Scientific, EO0491). Nuclei
638 suspension was incubated at 65 °C with 1200 rpm shaking in a heat block overnight. On the
639 next day, the mixture was purified with Qiagen MiniElute Purification kit (QIAGEN, 28004)
640 and eluted in 20 μ l of elution buffer. Purified genomic DNA was measured, and run on a 1.5%
641 agarose gel (Lonza, 50004) to check size distribution.

642 **Animals:** The mouse brain, liver, and kidney tissues were from the 8-week-old Mice FVBN
643 mice, housed in individually ventilated cages (3-5 animals per cage) in accordance with
644 Uppsala University regulations on mice with appropriate organic bedding, paper house
645 enrichments, food and water *ad libitum* and 12/12-hour light/dark cycle. All experiments were
646 performed in accordance with national guidelines and regulations, and with the approval of
647 the animal care and use committees at Uppsala University.

648

649 **Mouse Tissue collection:** 8-week-old Mice were sacrificed via inhalation euthanasia, and
650 mouse organs (brains, livers and kidneys) were collected. For frozen sample, livers and
651 kidneys were snap-frozen on dry ice and stored at -80 °C. For FFPE sample, mouse brains,
652 livers and kidneys were fixed with formalin overnight, and then washed with phosphate-
653 buffered saline (PBS) and kept in 70% ethanol for paraffin embedding. Fixed mouse brains,
654 livers and kidneys were routinely processed, and paraffin embedded.

655

656 **Primary data processing for the FFPE-ATAC and standard ATAC-seq**

657 All scripts (available in Supplemental Code) are deposited in the following link:

658 <https://github.com/pengweixing/FFPE-ATAC>. For sequencing libraries of FFPE-ATAC, the T7

659 promoter sequences and Tn5 transposase sequences from the Illumina single end

660 sequencing reads were trimmed using cutadapt software with slightly modifications (Martin

661 2011) and in-house script, which was deposited in following link:

662 <https://github.com/pengweixing/FFPE-ATAC>. For sequencing libraries of standard ATAC-seq,

663 the Tn5 transposase sequences from the Illumina paired end sequencing reads were

664 trimmed with in-house script. After the adaptor trimming, the sequencing reads were mapped
665 to the reference genome (mm9 or hg19) with Bowtie 2 using parameters -very
666 sensitive(Langmead et al. 2009). The duplicate reads were removed with Picard v1.79
667 (<http://picard.sourceforge.net>). The mapping for FFPE-ATAC on FFPE samples, FFPE-
668 ATAC on frozen samples, standard ATAC-seq on FFPE samples and standard ATAC-seq
669 on frozen samples was all performed with same parameters, thus, using GRCh38 and
670 GRCm38 (mm10) as refence genome for mapping would not significantly affect the
671 conclusions. SAMtools v1.9 software was used to sort and filter BAM files(Li et al. 2009).
672 The bigWig file was generated from BAM file using deepTools v3.5 software with the option
673 “bamCoverage”(Ramirez et al. 2014). The TSS enrichment score was calculated using
674 deepTools with the option “computeMatrix”(Ramirez et al. 2014). The peak calling was
675 performed using MACS2 in the parameters of -q 0.01 -nomodel -shift 0 (Zhang et al. 2008).
676 The read counts within peaks for each sample were calculated using BEDTools v2.29.2 with
677 the option “multicov”(Quinlan and Hall 2010). Genomic annotation and distance of peaks
678 relative to TSS were calculated using CHIPseeker R package (Yu et al. 2015). Sequencing
679 library complexity was calculated using Preseq v3.1.2 (Daley and Smith 2014). Differential
680 peak analysis was performed with DESeq2 software (Love et al. 2014) and differential peaks
681 were filtered with Log_2 (fold change) >3 and false discovery rate <0.01. The insert size
682 distribution for nucleosome-free region and mononucleosome were calculated using
683 ATACseqQC package (Ou et al. 2018). The sequencing coverage was visualized in the
684 Integrative Genomics Viewer (IGV) (IGV) (Thorvaldsdottir et al. 2013). Transcriptional factors
685 enrichments were performed using HOMER v4.11 with “findMotifsGenome” tool (Heinz et al.
686 2010). The gene annotation was analyzed using CHIPseeker package(Yu et al. 2015). The
687 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis
688 were performed with DAVID(Huang et al. 2007).

689

690 **Differential peak analysis of CRC FFPE-ATAC**

691 The Nonnegative matrix factorization (NMF) method(Brunet et al. 2004) was used to cluster
692 the 7 cases of CRC FFPE-ATAC with default algorithm. The differentially FFPE-ATAC peaks
693 from two clusters of CRC were identified with DEseq2(Love et al. 2014), following the
694 parameter of fold-change > 2 and false discovery rate < 0.01. HOMER was used to calculate
695 the significant transcriptional factors enrichment from the differentially FFPE-ATAC
696 peaks(Heinz et al. 2010).

697

698 **ENCODE DNase-seq data:** 8-week-old mouse liver and kidney ENCODE DNase-seq data
699 were downloaded from NCBI GEO with accession numbers: GSM1014195 (liver) and
700 GSM1014193 (kidney).

701

702 **DATA ACCESS:** All raw and processed sequencing data generated in this study have been
703 submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>)
704 under accession number GSE163306.

705 **ACKNOWLEDGMENTS:** We thanks for the critical comments of our manuscript from Dr. Ulf
706 Landgren. This work is supported by grants to X.C. from the Swedish Research Council (VR-
707 2016-06794, VR-2017-02074.), Åke Wibergs stiftelse (M20-0007), Beijer Foundation,
708 Jeassons Foundation, Petrus och Augusta Hedlunds Stiftelse, Göran Gustafsson's prize for
709 younger researchers, Vleugel Foundation, and Uppsala University). Part of this work was
710 facilitated by the Protein Science Facility at Karolinska Institute, Stockholm.

711 **AUTHOR CONTRIBUTIONS:** F. J. W., T. S. and X. C. conceived and designed the study. H.
712 Z., V. K. P., M. Z., L. M., L. Z., G. R. performed experiments. P.X., and H. Z. performed all
713 the data mining in the study. X. C. wrote the manuscript with input from all authors. X. C.
714 supervised all aspects of this work.

715 **DISCLOSURE DECLARATION:** X. C., V. K. P., and L. Z. have filed patent applications
716 related to the work described here. The title of the patent application is "Method of preparing
717 DNA from formalin-fixed-paraffin-embedded (FFPE) tissue samples". The Swedish

718 Provisional Application was filed on June 28, 2021, Patent Application No. 2150823-9 in
719 Sweden. The authors declare no competing financial interests.

720 REFERENCES:

- 721 Ashida R, Tominaga K, Sasaki E, Watanabe T, Fujiwara Y, Oshitani N, Higuchi K, Mitsuyama S,
722 Iwao H, Arakawa T. 2005. AP-1 and colorectal cancer. *Inflammopharmacology* **13**:
723 113-125.
- 724 Brunet JP, Tamayo P, Golub TR, Mesirov JP. 2004. Metagenes and molecular pattern
725 discovery using matrix factorization. *P Natl Acad Sci USA* **101**: 4164-4169.
- 726 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native
727 chromatin for fast and sensitive ... *Nat Methods* **10**: 1213-1218.
- 728 Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY,
729 Greenleaf WJ. 2015. Single-cell chromatin accessibility reveals principles of
730 regulatory variation. *Nature* **523**: 486-490.
- 731 Cejas P, Li L, O'Neill NK, Duarte M, Rao P, Bowden M, Zhou CW, Mendiola M, Burgos E, Feliu
732 J et al. 2016. Chromatin immunoprecipitation from fixed clinical tissues reveals
733 tumor-specific enhancer profiles. *Nat Med* **22**: 685-691.
- 734 Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, Xie XS. 2017. Single-cell whole-genome
735 analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**: 189-
736 194.
- 737 Chen X, Litzenburger UM, Wei Y, Schep AN, LaGory EL, Choudhry H, Giaccia AJ, Greenleaf WJ,
738 Chang HY. 2018. Joint single-cell DNA accessibility and protein epitope profiling
739 reveals environmental regulation of epigenomic heterogeneity. *Nat Commun* **9**: 4590.
- 740 Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC,
741 Ghosh RP, East-Seletsky A et al. 2016. ATAC-seq reveals the accessible genome by
742 transposase-mediated imaging and sequencing. *Nat Methods* **13**: 1013-1020.
- 743 Chin HG, Sun Z, Vishnu US, Hao P, Cejas P, Spracklin G, Esteve PO, Xu SY, Long HW, Pradhan
744 S. 2020. Universal NicE-seq for high-resolution accessible chromatin profiling for
745 formaldehyde-fixed and FFPE tissues. *Clin Epigenetics* **12**: 143.
- 746 Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong
747 CK, Cho SW et al. 2018. The chromatin accessibility landscape of primary human
748 cancers. *Science* **362**.
- 749 Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S,
750 Satpathy AT, Rubin AJ, Montine KS, Wu B et al. 2017. An improved ATAC-seq
751 protocol reduces background and enables interrogation of frozen tissues. *Nat*
752 *Methods* **14**: 959-962.
- 753 Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ,
754 Trapnell C, Shendure J. 2015. Multiplex single cell profiling of chromatin accessibility
755 by combinatorial cellular indexing. *Science* **348**: 910-914.
- 756 Daley T, Smith AD. 2014. Modeling genome coverage in single-cell sequencing.
757 *Bioinformatics* **30**: 3159-3165.
- 758 Fanelli M, Amatori S, Barozzi I, Soncini M, Dal Zuffo R, Bucci G, Capra M, Quarto M, Dellino
759 GI, Mercurio C et al. 2010. Pathology tissue-chromatin immunoprecipitation, coupled
760 with high-throughput sequencing, allows the epigenetic profiling of patient samples.
761 *Proc Natl Acad Sci U S A* **107**: 21535-21540.

- 762 Fox CH JF, Whiting J, Roller PP. 1985. Formaldehyde fixation. *J Histochem Cytochem* **33**:
763 845-853.
- 764 Gaffney EF, Riegman PH, Grizzle WE, Watson PH. 2018. Factors that drive the increasing use
765 of FFPE tissue in basic and translational cancer research. *Biotech Histochem* **93**: 373-
766 386.
- 767 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK.
768 2010. Simple combinations of lineage-determining transcription factors prime cis-
769 regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-
770 589.
- 771 Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW,
772 Lane HC, Lempicki RA. 2007. The DAVID Gene Functional Classification Tool: a novel
773 biological module-centric algorithm to functionally analyze large gene lists. *Genome*
774 *Biol* **8**: R183.
- 775 Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, Ni B, Sklar J, Przytycka TM, Childs R et al. 2015.
776 Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue
777 samples. *Nature* **528**: 142-146.
- 778 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment
779 of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- 780 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
781 Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and
782 SAMtools. *Bioinformatics* **25**: 2078-2079.
- 783 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
784 RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- 785 Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L, Piscuoglio S,
786 Chadalavada K, Nanjangud G, Ng CK et al. 2017. Whole-genome single-cell copy
787 number profiling from formalin-fixed paraffin-embedded samples. *Nat Med* **23**: 376-
788 385.
- 789 Martin M. 2011. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
790 Reads. *EMBnet Journal*: 10-12.
- 791 Martina JD, Simmons C, Jukic DM. 2011. High-definition hematoxylin and eosin staining in a
792 transition to digital pathology. *J Pathol Inform* **2**: 45.
- 793 Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, Zhu LJ. 2018. ATACseqQC: a
794 Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC*
795 *Genomics* **19**: 169.
- 796 Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao CC, Markoulaki S, Earl
797 AS, Labade AS, Jaenisch R et al. 2021. In situ genome sequencing resolves DNA
798 sequence and structure in intact biological samples. *Science* **371**.
- 799 Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase
800 and tagmentation procedures for massively scaled sequencing projects. *Genome Res*
801 **24**: 2033-2040.
- 802 Qu K, Zaba LC, Satpathy AT, Giresi PG, Li R, Jin Y, Armstrong R, Jin C, Schmitt N, Rahbar Z et
803 al. 2017. Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and
804 Dynamic Response to HDAC Inhibitors. *Cancer Cell* **32**: 27-41 e24.
- 805 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
806 features. *Bioinformatics* **26**: 841-842.
- 807 Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for
808 exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-191.

- 809 Sato A, Sekine Y, Saruta C, Nishibe H, Morita N, Sato Y, Sadakata T, Shinoda Y, Kojima T,
810 Furuichi T. 2008. Cerebellar development transcriptome database (CDT-DB): profiling
811 of spatio-temporal gene expression during the postnatal development of mouse
812 cerebellum. *Neural Netw* **21**: 1056-1069.
- 813 Sos BC, Fung HL, Gao DR, Osothprarop TF, Kia A, He MM, Zhang K. 2016. Characterization of
814 chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq)
815 assay. *Genome Biol* **17**: 20.
- 816 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-
817 performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-
818 192.
- 819 Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A,
820 Kampf C, Sjostedt E, Asplund A et al. 2015. Proteomics. Tissue-based map of the
821 human proteome. *Science* **347**: 1260419.
- 822 Waldron L, Simpson P, Parmigiani G, Huttenhower C. 2012. Report on emerging
823 technologies for translational bioinformatics: a symposium on gene expression
824 profiling for archival tissues. *BMC Cancer* **12**: 124.
- 825 Webster M, Witkin KL, Cohen-Fix O. 2009. Sizing up the nucleus: nuclear shape, size and
826 nuclear-envelope assembly. *J Cell Sci* **122**: 1477-1486.
- 827 Xie L, Dong P, Chen X, Hsieh TS, Banala S, De Marzio M, English BP, Qi Y, Jung SK, Kieffer-
828 Kwon KR et al. 2020. 3D ATAC-PALM: super-resolution imaging of the accessible
829 genome. *Nat Methods* **17**: 430-436.
- 830 Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak
831 annotation, comparison and visualization. *Bioinformatics* **31**: 2382-2383.
- 832 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM,
833 Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**:
834 R137.
835
- 836

837 **Figure Legends**838 **Figure 1: Standard ATAC-seq on FFPE samples and design of FFPE-ATAC.**

839 (A), DNA damage on accessible chromatin sites in FFPE samples hampers PCR amplification in
840 standard ATAC-seq on FFPE samples.

841 (B-E), Comparison of DNA fragment size distribution (B, C) and library complexity (D, E) from
842 standard ATAC-seq on frozen mouse liver and kidney, and standard ATAC-seq on FFPE mouse liver
843 and kidney.

844 (F, G), Quality control metrics of standard ATAC-seq on frozen mouse liver (F) and kidney (G), and
845 standard ATAC-seq on FFPE mouse liver (F) and kidney (G). Lib size = total sequencing reads of
846 sequencing library (million); %Mito = percentage of mitochondria; TSS = enrichment score at
847 transcription start sites (TSS); FRiP = fraction of reads in peaks.

848 (H, I), Comparison of chromatin accessibility between standard ATAC-seq on frozen samples and
849 FFPE samples. Left: Genome-wide comparison of accessible chromatin regions. R = Pearson's
850 correlation. Middle: Differential peak analysis between standard ATAC-seq on frozen samples and
851 FFPE samples. FDR = false discovery rate. Right: Distribution of the more accessible regions from
852 frozen and FFPE mouse samples across transcription start sites (TSS).

853 (J), Design of FFPE-ATAC by combining T7-Tn5 transposase tagmentation and T7 *in vitro*
854 transcription.

855

856

857 **Figure 2: FFPE-ATAC decodes chromatin accessibility with low cell numbers obtained from**
858 **FFPE tissue sections.**

859 (A), Workflow of FFPE-ATAC.

860 (B), Quality control metrics of FFPE-ATAC on frozen mouse liver and FFPE mouse liver, and
861 standard ATAC-seq on frozen mouse liver and FFPE mouse liver. Lib size = total sequencing reads of
862 sequencing library (million); %Mito = percentage of mitochondria; TSS = enrichment score at
863 transcription start sites (TSS); FRiP = fraction of reads in peaks.

864 (C, D), Comparison of sequencing library complexity (C) and genome browser tracks (D) from FFPE-
865 ATAC on frozen mouse liver and FFPE mouse liver, and standard ATAC-seq on frozen mouse liver
866 and FFPE mouse liver. Chr. = Chromosome.

867 **(E-G)**, Comparison of chromatin accessibility from different conditions: standard ATAC-seq on frozen
868 mouse liver vs. FFPE-ATAC on frozen mouse liver **(E)**, FFPE-ATAC on frozen mouse liver vs. FFPE-
869 ATAC on FFPE mouse liver **(F)**, and FFPE-ATAC on FFPE mouse liver vs. standard ATAC-seq on
870 FFPE mouse liver **(G)**. Top: Genome-wide comparison of accessible chromatin regions. $R =$
871 Pearson's correlation. Middle: Differential peak analysis. FDR = false discovery rate. Bottom:
872 Distribution of the more accessible regions from each condition across transcription start sites (TSS).
873 **(H)**, Quality control metrics of FFPE-ATAC on frozen mouse kidney and FFPE mouse kidney, and
874 standard ATAC-seq on frozen mouse kidney and FFPE mouse kidney. Lib size = total sequencing
875 reads of sequencing library (million); %Mito = percentage of mitochondria; TSS = enrichment score at
876 transcription start sites (TSS); FRiP = fraction of reads in peaks.

877 **(I, J)**, Comparison of sequencing library complexity **(I)** and genome browser tracks **(J)** from FFPE-
878 ATAC on frozen mouse kidney and FFPE mouse kidney, and standard ATAC-seq on frozen mouse
879 kidney and FFPE mouse kidney. Chr. = Chromosome.

880 **(K-M)**, Comparison of chromatin accessibility from different conditions: standard ATAC-seq on frozen
881 mouse kidney vs. FFPE-ATAC on frozen mouse kidney **(K)**, FFPE-ATAC on frozen mouse kidney vs.
882 FFPE-ATAC on FFPE mouse kidney **(L)**, and FFPE-ATAC on FFPE mouse kidney vs. standard
883 ATAC-seq on FFPE mouse kidney **(M)**. Top: Genome-wide comparison of accessible chromatin
884 regions. $R =$ Pearson's correlation. Middle: Differential peak analysis. FDR = false discovery rate.
885 Bottom: Distribution of the more accessible regions from each condition across transcription start sites
886 (TSS).

887

888

889 **Figure 3: FFPE-ATAC decodes chromatin accessibility from the mouse cerebellum with the aid**
890 **of H&E staining.**

891 **(A)**, Hematoxylin and eosin staining (H&E) of a mouse FFPE brain tissue section, where the location
892 of the cerebellum is illustrated with a dotted line.

893 **(B)**, Genome browser tracks of results from FFPE-ATAC analyses of isolated mouse FFPE
894 cerebellum. Chr. = Chromosome.

895 (C), Reproducibility of FFPE-ATAC analyses of mouse FFPE cerebellum. Left, the genome-wide
896 correlation from the FFPE-ATAC reads. Right: the overlapping peaks from the FFPE-ATAC in the two
897 technical replicates. R = Pearson's correlation.

898 (D), Enrichment of Gene Ontology terms for the top 10 000 FFPE-ATAC peaks for the mouse FFPE
899 cerebellum.

900

901 **Figure 4: FFPE-ATAC decodes chromatin accessibility from clinical archived tumor samples.**

902 (A), Schematic image showing the location of human colorectal cancer (CRC) samples: colon and
903 rectum.

904 (B), Nonnegative matrix factorization (NMF) of chromatin accessibility with FFPE-ATAC from 2 cases
905 of rectal cancer and 5 cases of colon cancer, identifying two clusters.

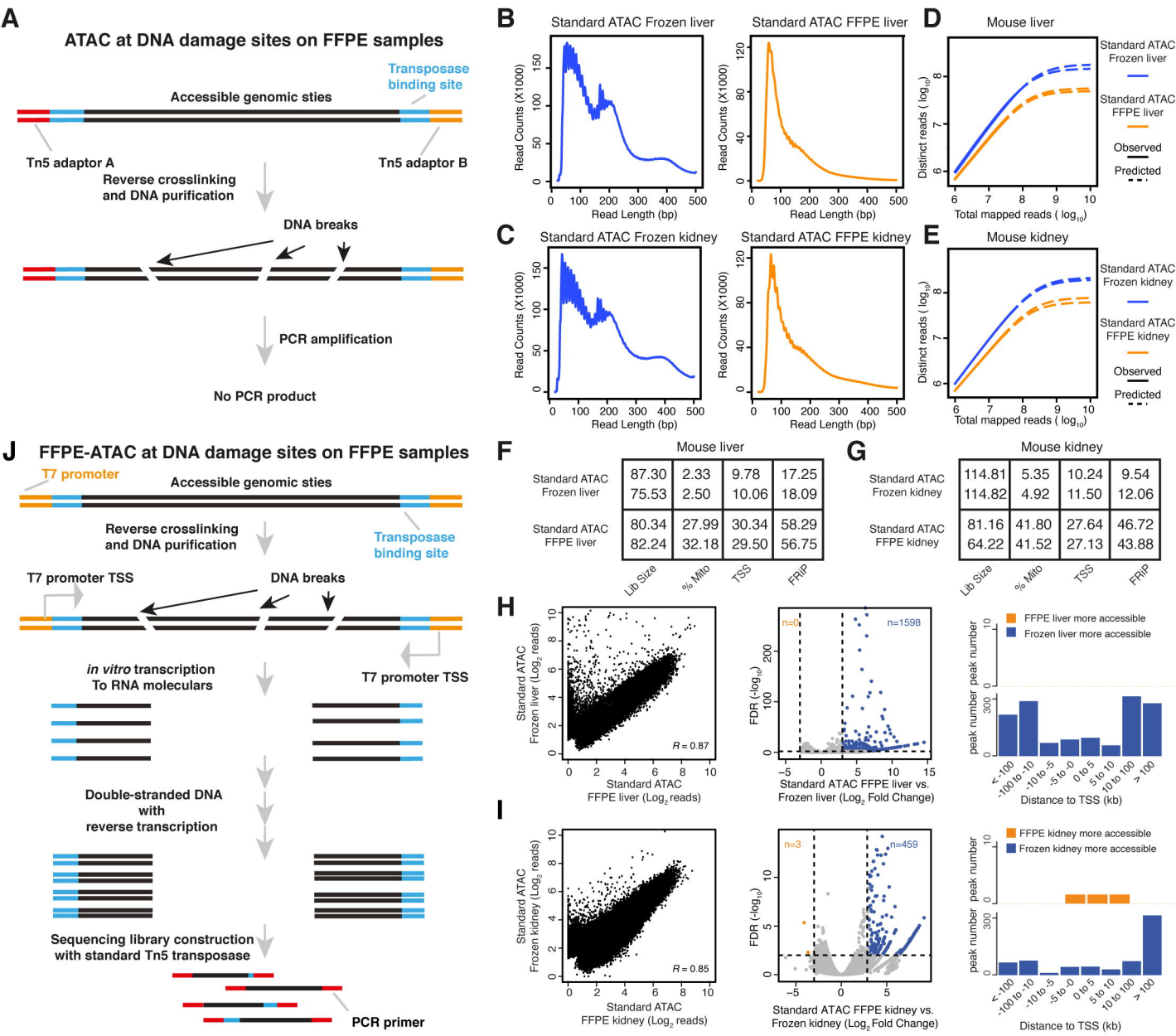
906 (C), Regulatory elements of the CRC marker gene *LRCH4* are accessible in both clusters.

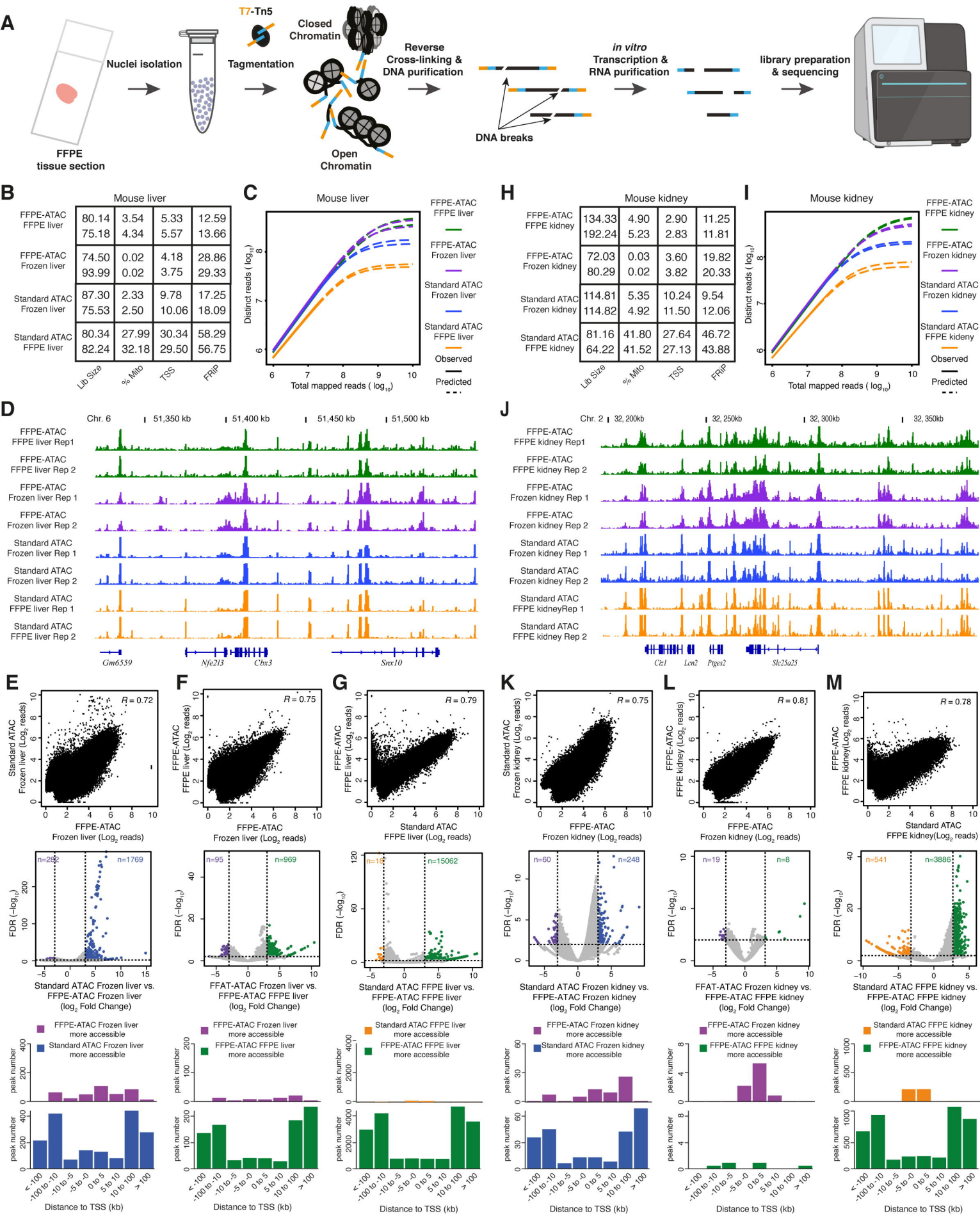
907 (D), Representative gene loci that are more accessible in cluster 1, as seen from the differential
908 FFPE-ATAC peaks.

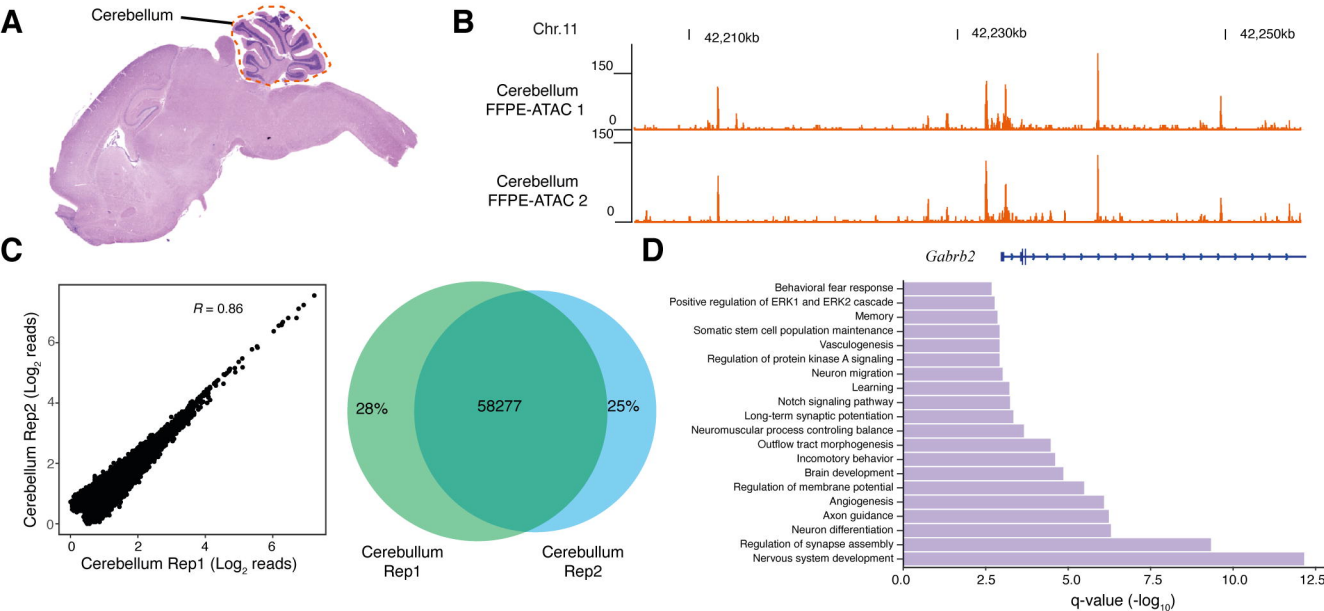
909 (E), Representative gene loci that are more accessible in cluster 2, as seen from the differential
910 FFPE-ATAC peaks.

911 (F), Ranked transcription factors significantly enriched in the specific regulatory elements from cluster
912 1.

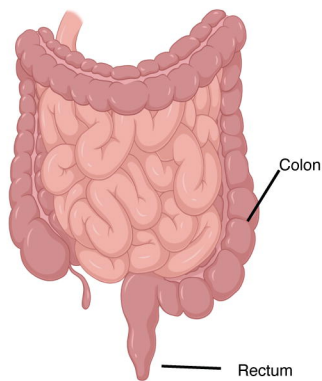
913 (G), Ranked transcription factors significantly enriched in the specific regulatory elements from cluster
914 2.



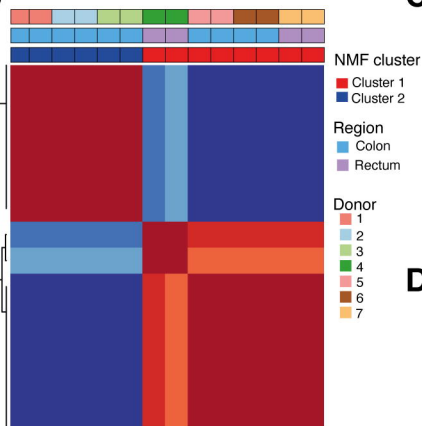




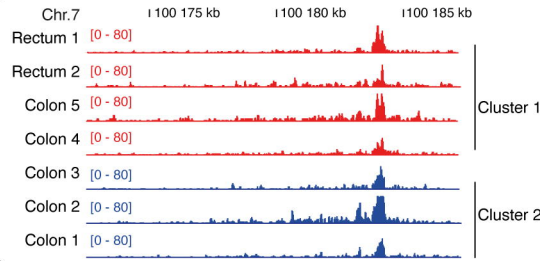
A



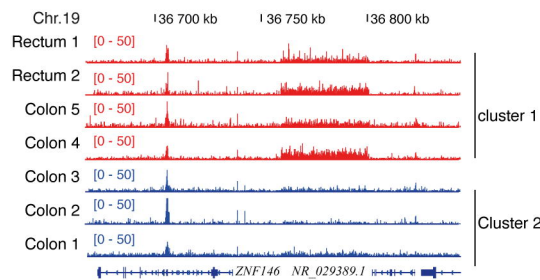
B



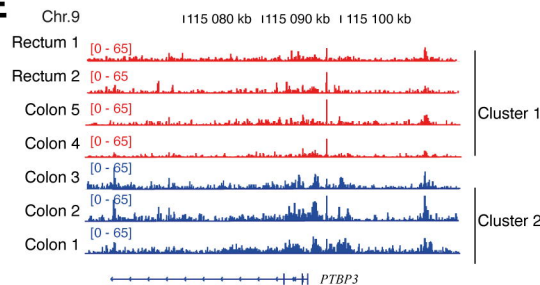
C



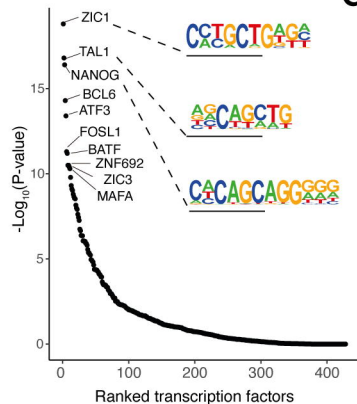
D



E



F



G

