A pVII positioning code determines the dynamics of Adenoviral nucleoprotein structure and primes it for early gene activation

Uwe Schwartz[a], Tetsuro Komatsu[b,c], Claudia Huber[a], Floriane Lagadec[c,d], Elisabeth Silberhorn[a], Edouard Bertrand[e], Harald Wodrich[c#] and Gernot Laengst[a#]

[a]Biochemie Zentrum Regensburg, University of Regensburg, Regensburg, Germany.
[b]Division of Transcriptomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan
[c]CNRS UMR 5234, Microbiologie Fondamentale et Pathogénicité, Université de Bordeaux, Bordeaux, France.
[d]Department of Molecular Biology, Faculty of Medicine, Göttingen Center of Biosciences (GZMB), Georg-August-University Göttingen, Humboldtallee 23, 37073 Göttingen, Germany.
[e]CNRS UMR 5355, Institut de Généthique Moléculaire de Montpellier, Montpellier, France

[#]Address correspondence: harald.wodrich@u-bordeaux.fr or gernot.laengst@vkl.uni-regensburg.de

Running title: Adenovirus nucleosome sequencing

SUMMARY

Inside the capsid adenovirus DNA is associated with the structural protein pVII. However, the viral DNA organisation, pVII positioning and the dynamic changes of viral packaging upon infection, to form a transcriptional active genome, are not known. We combined MNase-Seq and single genome imaging during early infection to provide the structure and time resolved dynamics of viral chromatin changes, correlated with gene transcription. pVII complexes form nucleosome-like arrays, being precisely positioned on DNA, creating a defined and unique adenoviral nucleoprotein-architecture. The structure renders the viral genome transcription competent with lower pVII densities at early gene loci, correlating with viral chromatin de-condensation upon infection. Nucleosomes specifically replace pVII at transcription start sites of early genes, preceding transcriptional activation. Our study suggests an underlying regulatory pVII nucleoprotein-architecture, required for the dynamic changes during early infection, including transcription related nucleosome assembly. We suggest that our study provides a basis for the development of recombinant adenoviral vectors exhibiting sustained expression in gene therapy.

INTRODUCTION

Most DNA viruses transport their genomes to the nucleus of uninfected target cells to initiate viral gene expression and genome replication. During cytoplasmic trafficking genomes are highly compacted inside the viral capsid to protect them from nucleic acid sensing molecules. Inside the nucleus genomes are decondensed and transcriptionally activated for viral gene expression and replication. Replicated genomes are then re-condensed for packaging into progeny. While this remarkable and reversible genome plasticity is highly efficient, little is known about the structural organization of the genome and their dynamic changes and regulatory cues of this process during virus entry.

Adenoviruses (Ad) are prototypic DNA viruses that contain a linear double stranded DNA genome of 30-40 kb, being incorporated into a ~ 90 nm non-enveloped icosahedral capsid shell. Ads enter cells by receptor-mediated endocytosis. Entry cues liberate the internal capsid protein VI and permit the virus to escape the endosomal compartment [1,2]. Cytoplasmic capsids are transported to the nuclear pore complex (NPC) where they dock, disassemble and release the viral genome for nuclear import [3,4]. Within hours after cell entry, nuclear genomes initiate gene expression from the immediate early promoters, as E1A. E1A encodes a transactivator driving early genes responsible for cell and immune modulation (E1B, E3, E4) and replication (E2). In

contrast activation of late genes (L1-L5) encoding most structural proteins, requires prior genome replication [5,6].

Packaged genomes are devoid of histones. Instead they associate with basic core proteins VII (pVII), V (pV) and polypeptide μ [7,8]. pVII is the most tightly bound and most abundant protein with over 500 copies per virion and packages the viral genome into a structure appearing as irregular "beads-on-a-string" by electron microscopy [9]. pVII, exhibits sequence similarities to protamines and may use its positive charges to neutralise and to compact the viral DNA [10–12]. No information on genomic binding sites of pVII, its structural organisation on DNA, or its dynamics during infection is available [13]. pVII remains associated with the imported genome while pV dissociates from the viral DNA [14]. The fate of the μ-peptide is not known.

Some studies suggest that pVII remains attached throughout the early phase of infection while others argue for a gradual and at least partial removal from the incoming genome [10,15–18]. pVII was shown to functionally interact with the viral transactivator E1A, pinpointing to a role in transcriptional activation [10]. It also was suggested that transcription would subsequently remove pVII from the genome [19]. In contrast chromatin immune precipitation (ChIP) analysis proposed that pVII remains associated with viral chromatin for several hours [10,15–17,20].

Functional chromatin extraction by using high and low concentrations of micrococcal nuclease (MNase) to digest cellular chromatin, combined with high-throughput sequencing, is a way to analyse protein binding to DNA and to reveal dynamic changes in chromatin structure [21,22]. In this study we combined time resolved MNase and RNA sequencing and single molecule imaging to correlate chromatin structure changes with transcription at very early time points of adenoviral infection. This approach revealed the specific nucleoprotein architecture of the virus, which dynamically changes during early infection. pVII positioning is driven by sequence and is specifically organized at early genes, correlating with the early de-condensation of these regions upon infection. Defined pVII nucleoprotein complexes are precisely exchanged by nucleosomes on early gene promoters, showing pVII remodeling and nucleosome assembly preceding transcriptional activation and not being its consequence. The time resolved single molecule imaging of transcribing genomes support our findings, identifying the morphological correlate at the single genome level at subcellular resolution. The defined positioning of pVII on the viral genome and its functional involvement in dynamic chromatin rearrangement during infection, suggest a key role in regulating viral life cycle. We propose that knowledge of the adenovirus nucleoprotein architecture could lead to the design of improved Ad vectors for gene-therapy or vaccination, with sustained expression.

RESULTS

Time resolved analysis of the early adenovirus transcriptome by RNA-Sequencing

To investigate the early spatio-temporal reorganization of viral chromatin and transcription activation we infected H1299 lung epithelial cells with a partially E3-deleted but replication competent HAd-C5. The inoculum was replaced with fresh medium after 30 min defining time point zero of the experimental setup (Fig. 1A). On time point 0, 0.5, 1, 2, and 4 hours post infection (hpi) infected cells were subjected to immune fluorescence analysis, RNA-sequencing (RNA-seq) and functional chromatin extraction combined with high-throughput sequencing of protected DNA fragments (MNase-seq) (Fig. 1A). To verify the progression of infection, capsid protein VI (pVI) exposure during endosomal escape was monitored by immunofluorescence [23]. The subsequent import of the nuclear genome, was monitored by immunofluorescence of the genome associated pVII (Fig. 1B) [18]. pVI and pVII quantification showed that endosomal escape occurred within 30 min of virus addition (Fig. 1B and 1C) followed by rapid nuclear delivery of viral genomes starting at 30 min post inoculum removal (0.5 hpi), with most genomes being imported 1 hpi (Fig. 1B and 1D). The data confirms a synchronized infection, covering pre- and post-nuclear genome delivery phases resembling previous dynamic entry analysis [24].

To investigate viral transcription, total RNA was isolated in two independent experiments. A polyA library was prepared to enrich for mRNA transcripts and sequenced on Illumina HiSeq platform yielding on average 26 Mio reads per sample (Table S1). The reads were mapped to an extended version of the human genome amended by the annotation of the HAd-C5 genome. First Ad transcriptional activity could be detected 2 hpi, and strongly increased levels were observed 4 hpi (Fig. 1E and Table S1, whole genome profile shown in Fig. S1). Considering that the first Ad genomes in the nucleus were detected at 0.5 hpi (Fig. 1B and 1C), Ad transcription initiation and mRNA synthesis exhibited a significant lag-time after nuclear entry, suggesting the existence of a remodelling process being required for the assembly of the transcription machinery. Consistent with the spatial organisation of the Ad transcription program [25,26], transcription is initiated at both ends of the Ad genome covering the early genes (Fig. S1). We detected E1A as the first viral transcript [27], along with significant levels of E3 at 2 hpi (Fig. 1E). At 4 hpi transcription is markedly enhanced with high levels of transcripts originating from all early genes and significant signals arising from the E2A region. Taken together we observed an efficient utilization of the host cell transcription machinery after nuclear entry, revealing a transcriptional competent viral template. Transcription initiates with

123 a delay of 1 hour after nuclear entry opening the question of which structural changes may be
124 required for assembly of the transcriptional machinery.
125
126 **Chromatin architecture of the invading Ad genome**
127 To address the Ad pVII nucleoprotein organisation, we partially hydrolysed viral and cellular
128 genomic DNA with the endo-nuclease MNase (Fig. 2A). This enzyme does preferentially
129 hydrolyse DNA in the linker region between nucleosomes and is restricted by stable protein-
130 DNA complexes [28]. Incubation of cellular chromatin with increasing concentration of MNase
131 resulted in the appearance of a typical MNase ladder, reflected by the monomer and multimers
132 of the nucleosomal DNA fragments (Fig. 2A). MNase concentrations were titrated such that
133 one sample exhibited less than 20% mono-nucleosomal DNA (low digestion) and the other
134 sample was fully hydrolysed to the mono-nucleosomal DNA (high digestion; Fig. S2A). The
135 latter is the common condition used for MNase-seq experiments to date [29–31]. High and low
136 digestion conditions were used for each time point to isolate the mono-nucleosomal DNA from
137 two replicates. The nucleosomal DNA of 17 individual experiments was used for library
138 preparation and paired-end high-throughput sequencing, yielding in total 3.8 billion paired-end
139 reads (Table S2).
140 DNA fragments were aligned to the human and the Ad genome, first analysing the read length
141 distribution on both genomes at time point 0 hpi (Fig. 2B). When comparing the virus and host
142 genome digested at the high MNase concentration, we observed a clear difference in the read-
143 length distribution on the two genomes (Fig. 2B). Whereas the majority of DNA fragments
144 exhibited the typical nucleosomal DNA size on the human genome (147 bp), the protected DNA
145 fragments on the Ad genome were significantly shorter with a mean size of 60 to 70 bp. The
146 MNase-seq was capable to reveal a stable adenoviral packaging unit that most probably
147 represents pVII bound to DNA. Previous studies showed that nuclease digestion of Ad
148 nucleoprotein resulted in discrete populations of protected DNA fragments, suggesting a
149 chromatin-like configuration with sizes varying between 180 bp and 200 bp, between 30 and
150 90 bp, or between 30 and 60 bp depending on the conditions used [8,10,32]. Assuming that larger
151 DNA fragments represent partial nuclease hydrolysis products, the 60 to 70 bp long Ad-DNA
152 fragment mirrors the previously described pVII-DNA protein complexes. In contrast it is not
153 known if these pVII-DNA complexes are precisely positioned or if they are placed randomly.
154 Detection of the pVII bound DNA allows us for the first time to analyse the occupancy and
155 positioning of Ad pVII. We used the DANPOS2 package [33] to further analyse the viral
156 nucleoprotein structure. We identified defined peaks over the Ad genome, showing non-

157 random, specific pVII-DNA positions on the viral DNA (Fig. 2D). The MNase profiles and
158 peak calling was highly reproducible between biological replicates and at later time points,
159 suggesting that pVII occupies specific Ad genome positions like the nucleosome on the host
160 genome (Fig. S2C-E). The adenovirus exhibits a defined nucleoprotein architecture that differs
161 along its functional regions. Interestingly, a central region of the viral genome (Late3) exhibits
162 almost no peaks, revealing domains of structural heterogeneity that could reflect functional
163 domains of the virus genome e.g. when packaged inside the virion, or being required for gene
164 activation (Fig. 2D).
165
166 **pVII exhibits DNA sequence dependent positioning patterns**
167 The observation of a defined nucleoprotein architecture for the adenoviral genome was
168 surprising because it can be *a priori* replaced by unrelated vector DNA of similar length [34]. We
169 next asked whether intrinsic DNA sequences patterns may establish pVII peaks. Such a
170 sequence dependent positioning code was shown for the nucleosome [35]. Sequence specific
171 bending of the DNA reduces the energy requirements for wrapping the DNA around the histone
172 octamer surface, revealed by characteristic di-nucleotide repeats in the nucleosomal DNA
173 sequence at 10 bp intervals [36]. We observed this described di-nucleotide pattern when analysing
174 our nucleosomal DNA fragments annotated to the human genome (Fig. 2C – upper panel).
175 Interestingly, we also observed a specific di-nucleotide pattern, when computing the di-
176 nucleotide frequency of the mid-point centered pVII fragments, suggesting that a specific DNA
177 code underlies pVII positioning on the Adenovirus genome (Fig. 2C – lower panel). However,
178 to our surprise, a periodic pattern of 5.4 bp was obtained for A/T- and G/C-dinucleotides, much
179 shorter than the nucleosomal DNA repeat pattern, implying a distinct packaging mode. Such a
180 repeat pattern indicates a spring-like winding of the DNA around pVII, similar to protamines
181 (see discussion for details). Even without knowledge of the exact DNA-pVII topology, the
182 sequence pattern and the specific binding sites of pVII on DNA clearly suggest the organisation
183 of the Ad-genome into a specific nucleoprotein-structure that could be essential for its function.
184 On average we detected 238 pVII peaks covering the Ad genome, which was consistent
185 throughout all timepoints (Fig. 2E). Next, we quantified the mean pVII-DNA particle repeat
186 length (pVII-RL) by measuring the center-center distance of adjacent pVII peaks in each sample
187 (Fig. 2F and S2B). The median pVII-RL measures 97.3 bp suggesting that we isolated defined
188 pVII-DNA particles spaced by accessible DNA linkers of 20 to 40 bp in length.
189 To address the presence of functionally distinct regions, we systematically subdivided the Ad
190 genome into 10 domains of similar size, named after the including gene and classified by the

191 timing of expression (color code: early to late = light to dark blue) (Fig. 2D). Plotting pVII-RL
192 showed a clear correlation of the repeat length with early expressing genes (Fig. 2G). Increased
193 repeat lengths in the regions of the early genes can serve to increase genome accessibility and
194 to enable the binding of regulatory factors. Strengthening this finding, the pVII occupancy and
195 stability, reflected by the peak height, did anti-correlate with early timing in gene expression
196 (Fig. 2H). In general, less pVII peaks (Fig. 2I) with larger repeat lengths and increased
197 instability were found at early expressed domains indicating an open chromatin conformation
198 (Fig S2D-E).
199 Our high-resolution analysis of Ad genome packaging provides for the first-time information
200 on pVII-DNA particle positioning and occupancy. The analysis revealed a defined viral
201 nucleoprotein architecture, with an open configuration at early expressed genes potentially
202 enabling immediate access of the host cell transcription machinery.
203
204 **Dynamic changes of Ad genome packaging**
205 Previous studies showed that histones are deposited on the Ad genome [37]. However, where and
206 whether functional nucleosomes were assembled on the viral DNA is not known. Analysing the
207 MNase-seq pattern of the time points 1, 2 and 4 hpi revealed the appearance of a 147 bp long,
208 protected DNA fragment at 2 and 4 hpi (Fig. 3A and biological replicate in S3A). The
209 nucleosome sized peaks increased in size, within the background of the pVII DNA fragments,
210 revealing a co-occupation of the viral genome by pVII and nucleosomes. To quantify the
211 fraction of nucleosomes on the viral genome, we simulated the peak sizes by computationally
212 adding nucleosomal reads to the pVII fragment data of time point 1 hpi (Fig. S3B). The analysis
213 suggested that at 2 hpi 8% and at 4 hpi 14% of the viral genomes are occupied with
214 nucleosomes. First nucleosomal occupancy is detected at 2 hpi, coinciding with the initiation
215 of viral transcription and indicating the requirement of nucleosome assembly for proper
216 transcription initiation.
217 In order to reveal the sites of nucleosome assembly, DNA fragments with a size between 137-
218 157 bp were extracted from time points 1, 2 and 4 hpi (Fig. 3A; indicated by dashed lines). The
219 time point 1 hpi was defined as background signal and we calculated the fold-change of
220 nucleosomal DNA arising at 2 hpi and 4 hpi. A cut-off of 5-fold enrichment was used to define
221 the genomic sites of preferential nucleosome assembly (Fig. 3B, raw signal is shown in Fig.
222 S3C). Strikingly, nucleosomal gain on the Ad-genome was not random. At 2 hpi we detected
223 12 hotspots of nucleosome assembly and additional 17 sites at 4 hpi. All sites showed a gain in
224 enrichment over time, suggesting site specific nucleosome assembly after nuclear entry. The

225 sites of nucleosome assembly are concentrated within the three regions of the early genes and
226 within the peculiar Late3 region, suggesting a role of nucleosome assembly in transcriptional
227 activation.
228
229 **Early genes gain a postioned +1 nucleosome**
230 Nucleosome assembly selectively occurs in the genomic regions associated with the early
231 genes, exhibiting a lower pVII protein occupancy and increased pVII repeat-length. An analysis
232 of the incoming nucleosome positions with respect to pVII occupancy, revealed that
233 nucleosomes directly formed over the central positions of the pVII-nucleoprotein particles (Fig.
234 3C) suggesting a specific mechanism to replace pVII.
235 A detailed analysis of the nucleosome assembly sites with respect to the early genes revealed
236 that nucleosomes occupy discrete positions. Remarkably nucleosome positions coincide with
237 the 5'-end of the synthesized E1A and E4-orf RNA, resembling the highly positioned +1
238 nucleosome found in active human genes (Fig. 3D and 3E) [29]. Additional nucleosomes are
239 detected in the coding region of the early genes. The +1 nucleosome marks the transcription
240 start sites of genes and flanks the regulatory sequences of the promoter that reside in the
241 upstream linker region of this nucleosome. The Ad-DNA upstream of the +1 nucleosome
242 exhibits either no pVII occupancy or long pVII repeat lengths, allowing binding of regulatory
243 factors for gene activation. The preferential assembly of nucleosomes on the early genes,
244 specifically located at the +1 position, prior to transcription initiation implies the requirement
245 of nucleosome assembly for transcription initiation.
246
247 **Dynamic changes of viral genome accessibility over time**
248 Assessing the nucleoprotein structure with high MNase concentrations is providing information
249 about stable protein particles occupying the DNA. In contrast, limited MNase concentrations,
250 in analogy to the analysis of regulatory sites with DNase I [38], would reveal genomic regions
251 that de-condense chromatin and identifies regulatory elements [21].
252 As shown below, only a fraction of about 20% of the viruses in the cell are transcriptionally
253 active, explaining why we observed only little variation between samples and the different
254 timepoints treated with high MNase concentrations (Fig. 4A and S4A). In contrast, at low
255 MNase concentration (Fig. 4B) clear changes over time are detected, signifying gross structural
256 changes of the virus during early infection times. The active viruses undergo de-compaction
257 that can be detected by limited MNase hydrolysis. When monitoring the pVII positioning at 0
258 hpi, the low MNase signal was well correlated with the high MNase signal (R = 0.91),

259 essentially showing the same pVII positions in both conditions (Fig. S4B and S4C). However,
260 when inspecting the viral maps at later time points, we detected specific structural changes,
261 with changes in peak intensities and the appearance of novel peaks in the low MNase signal
262 (Fig. 4C). Structural changes over time, were quantified by a linear regression fit. The average
263 signal intensity obtained in a 300 bp sliding window was quantified for each timepoint. The
264 slope of the regression was used as a score for DNA accessibility changes (accessibility score;
265 Fig. 4C and S5A). Genomic regions with a positive accessibility score are opening up, being
266 more MNase accessible, whereas negative values correlate with DNA remaining condensed. A
267 similar strategy was used in previous studies to measure local chromatin accessibility [22,39,40].
268 Correlating with the activity of the virus, we detected Ad-chromatin de-condensation at the
269 early gene regions reflecting gene activation (Fig. 4C and 4D). Interestingly, the score at E1A
270 is not as pronounced as at other early genes (Fig. 4D). As this is the first gene to be transcribed,
271 we suggest that this region already pre-exists in an open chromatin configuration, correlating
272 with the low pVII peak density and therefore not changing over time.
273 The Late3 domain also exhibited a positive accessibility score, suggesting Ad chromatin de-
274 condensation at this site (Fig. 4D). Additionally, in this domain we observe a discrete peak
275 appearing at low MNase concentrations at 4 hpi (Fig. S5B). This peak is located in a rather
276 MNase insensitive region and was only detected in the low MNase datasets. The occurance of
277 this local nuclease hypersensitive site suggests the binding of regulatory factors [39,41,42].
278 Transcription factor motif analysis revealed potential binding sites for EN1, CREM, C/EBP-
279 alpha and C/EBP-beta in the centre of the peak (Fig. S5B). Interestingly, the transcripts of EN1,
280 CREM and C/EBP-beta in the host cell were upregulated up to twofold upon viral infection
281 (Fig. S5C).
282 As shown above, we identified the early gene regions and Late3 regions as the sites of
283 preferential nucleosome assembly and of viral chromatin de-condesation (Fig. 4C). To quantify
284 the probability of this co-incidence, we randomly sampled the same number of nucleosome
285 peaks for 10.000 times and calculated the cumulative accessibility score for each run (Fig. 4E).
286 The cumulative scores are normally distributed around zero. The score of the detected
287 nucleosome peaks (red dashed line in Fig. 4E) was very rarely reached by the randomly sampled
288 peaks (p-value: 0.0002), indicating that targeted deposition mechanisms drive nucleosome
289 assembly at the accessible regions of the Ad genome.
290
291 ***In situ* detection of E1A mRNA identifies transcribing Ad genomes.**

292 Time resolved analysis using MNase-seq and RNA-seq showed strong correlation between
293 localized viral chromatin remodelling and subsequent gene activation. We next used RNAscope
294 technology for the direct detection of individual viral transcripts at the single cell level [43]. A
295 specific set of probes for the E1A mRNA was designed (ACDBio) and validated on cells
296 infected with either HAd-C5 virus or an E1-deleted vector (Ad5-GFP). At 4 hpi specific E1A
297 transcripts could be readily detected in virus- but not in vector-infected cells (Fig. 5A), nor in
298 non-infected control cells (not shown). We next repeated the time course of infection (Fig. 1A)
299 and detected at each time point viral capsids (not shown) and viral genomes using specific
300 antibodies as well as E1A transcripts using RNAscope (Fig. 5B). Most genomes were imported
301 until 1 hpi with no significant additional import after this time point. Normalization with cell
302 associated capsids revealed that only ~40% of genomes were imported into the nucleus (Fig.
303 5C). While the nuclear import of genomes was finished 1 hpi, E1A transcripts only appeared at
304 2 hpi and further increased at 4 hpi (Fig. 5D). E1A transcript accumulation could be related to
305 the continued expression from a stable pool of transcriptionally active genomes or by an
306 increasing number of genomes initiating transcription over time. We quantified the
307 colocalization of nuclear genomes with nuclear E1A transcripts to discriminate between both
308 possibilities. Colocalizations resulting from association by probability would augment over
309 time when active genome numbers remain constant and E1A transcript numbers increase over
310 time. However, our results clearly showed that E1A positive nuclear genomes rapidly plateau
311 at ~20% despite strong accumulation of nuclear E1A transcripts (Fig. 5E). It is shown that only
312 a relatively small subpopulation of imported genomes contributes to early viral gene expression
313 until 4 hpi.
314
315 ***In vivo* detection of E1A mRNA identifies a subpopulation of transcribing Ad genomes.**
316 The RNAscope analysis confirmed the RNA-seq data and shows that the fraction of active viral
317 genomes first reconstitute nucleosomes and only then activate early gene expression. To
318 confirm this observation and further characterize transcribing and non-transcribing genomes
319 we developed an imaging system for living cells. To mark E1A transcripts in living cells we
320 inserted 24x MS2 binding sites into the 3'UTR of the E1A gene of a replication competent
321 virus. MS2 binding sites are short stem-loop structures derived from the MS2 bacteriophage
322 and can be detected by binding of multiple copies of fluorophore tagged MS2 binding protein
323 (i.e. MS2BP-NLS-GFP) forming a fluorescent spot *in situ* [44]. Production and amplification of
324 the virus resulted in normal yields with unaltered infectivity or protein composition compared
325 to the parental virus (data not shown). To detect Ad genomes in living cells we used our

326 previously established life cell imaging system based on fluorescent TAF-Iβ, which associates
327 with incoming genomes also generating a fluorescent spot *in situ* [18]. A stable mCherry-TAF-Iβ
328 expressing cell line was transduced with an expression vector for MS2BP-NLS-GFP to generate
329 double fluorescent stable cell lines. Cells were left either uninfected, or infected with untagged
330 control virus, or infected with E1A-MS2 tagged replicative virus and at ~ 4 hpi analysed by life
331 cell imaging at high temporal resolution (1 frame/sec) (Movies S1-S4). To display the results
332 in a single image we superimposed individual movie frames (60 sec t-projections, Fig. 6A). In
333 non-infected control cells both, mCherry-TAF-Iβ and MS2BP-NLS-GFP localized to the
334 nucleoplasm, with some unspecific enrichment of the MS2 signal in nucleoli (Fig. 6A top row,
335 taken from Movie S1). When cells were infected with non-MS2 containing control virus,
336 genomes formed mCherry-TAF-Iβ fluorescent spots in the nucleoplasm with limited confined
337 motion as previously described [18] while the MS2 signal remained unchanged compared to non-
338 infected cells (Fig. 6A 2nd row taken from Movie S2). In contrast cells infected with virus
339 expressing MS2-repeat-tagged E1A mRNA formed mCherry-TAF-Iβ spots indicating viral
340 genomes but also rapidly moving GFP spots visible inside the nucleoplasm and in the cytoplasm
341 (Movie S3/S4). GFP spots accumulated in most cells around 3-5 hpi in agreement with the
342 RNAscope experiment. Upon merging the mCherry-TAF-Iβ channel (viral genomes) with the
343 MS2BP-NLS-GFP channel (viral E1AmRNA) using t-projections, some genomes were clearly
344 double positive suggesting transcribing genomes (Fig. 6A 3rd and 4th row, taken from Movie
345 S3 and S4). Similar observations were made when we used a different MS2BP construct
346 (MS2BP-NES-GFP) predominantly localizing to the cytosol or when we lowered the input MOI
347 resulting in fewer (<10) genome copies per imaging field (Fig. S6). To quantify the number of
348 E1AmRNA positive genomes we performed kymograph analysis using dual-channel t-
349 projections of individual cells. This approach was possible because the genome position was
350 stable over time allowing the identification of MS2 positive (transcribing) genomes (Fig. 6B).
351 Kymograph analysis of several cells showed that ~20% of detected genomes inside a given
352 nucleus are positive for MS2BP-NLS-GFP (Fig. 6C), which was very similar to the RNAscope
353 approach (Fig. 6D). Taken together the analysis of transcriptionally active genomes in living
354 cells supports our findings that early in infection transcription is restricted to a subpopulation
355 of imported genomes that actively remodel the viral chromatin for transcription onset.
356

357 **DISCUSSION**

358 **The adenovirus exhibits a defined nucleoprotein architecture**

359 Early studies suggested that pVII binds adenoviral DNA at a 1:1 ratio like histones and protects
360 DNA from endonuclease cleavage [45,46]. However, no clearly defined pVII-DNA particle has
361 been identified, as the protected DNA had a variable size range of 30 to 250 bp depending on
362 experimental conditions and the degree of endonuclease digestion [8,13,32]. In addition, it was not
363 known whether pVII binds to specific sites on the viral genome, creating a defined viral
364 nucleoprotein architecture, like the cellular nucleosomes. Here, our MNase-seq approach
365 revealed the 147 bp long DNA fragments of the host nucleosomes and clearly identified a
366 shorter, protected DNA species with a mean size of about 60 to 70 bp originating from the viral
367 sequences. With pVII being the major DNA binding protein of the adenovirus, we suggest that
368 we observe DNA fragments protected by pVII. The pVII digestion pattern is not as clearly
369 defined as the nucleosomal DNA fragments, suggesting a heterogeneity in pVII-DNA
370 complexes. The rather broad peak of protected DNA fragments (50 to 150 bp) is mirroring the
371 variable results of the early studies, suggesting a rather heterogenous association of pVII with
372 DNA. Alternatively, such a digestion pattern could also be the result of discrete pVII core
373 complex (60 bp) that tightly associates with neighbouring pVII complexes, partially protecting
374 the interconnecting linker DNA from endonuclease cleavage. Structures that are visible in early
375 electron microscopic studies [13].
376
377 **pVII organizes the viral DNA into a specific nucleoprotein structure**
378 Annotation of the pVII protected DNA fragments to the viral genome revealed for the first time
379 a specific positioning of pVII on the genome. We show that the Ad genome is covered by about
380 238 stable pVII particles located at defined positions that are spaced by a DNA linker of about
381 20 to 40 bp. Precise pVII positioning suggest a highly organized and defined nucleoprotein
382 architecture shared by all virus particles. DNA structure determines histone binding
383 preferences, as shown by the DNA sequence dependent nucleosome positioning code [35]. DNA
384 bending allows wrapping of the DNA around a protein core, requiring site specific A/T and
385 G/C di-nucleotide enrichments. A/T and G/C repeats differ in the wideness of the DNA
386 major/minor groove thereby bending the DNA [35]. If the sequence dependent DNA shape
387 follows the protein surface path it increases binding affinity and thereby determines binding
388 site preferences [36]. Whereas the nucleosomal di-nucleotide repeats are spaced by 10.3 bp, pVII
389 exhibits a much shorter repeat length of 5.4 bp. Nucleosomes and pVII-DNA complexes are
390 organized differently and we suggest that the DNA associates in form of a spiral around the
391 pVII core (Fig. 2C). This toroidal structure was proposed for protamines, to which pVII exhibits
392 sequence similarities [46].

393 Our results show that pVII folds Ad genomes into an organized higher order structure, driven
394 by its specific binding sites encoded by the Ad-DNA sequence. Taken together with the active
395 remodelling of individual pVII molecules at early genes, this findings have to be taken into
396 account for the optimal (sequence-)design of therapeutic DNAs packed into adenoviral vectors
397 to achieve high and sustained transgene expression. In analogy to eukaryotic chromatin
398 organisation, specific assembly factors and ATP dependent remodelling enzymes may be
399 required to adjust the precise positioning of pVII on DNA, in addition to the DNA assembly
400 code. Like nucleosomes, pVII does not bind as isolated protein to DNA but is forming
401 homotypic complexes. The number of pVII proteins per Ad genome was estimated to be about
402 500 copies [9] and early biochemical studies suggested the occurrence of pVII dimers [47,48]. Taking
403 into account the median number of 238 pVII positions we identified on the Ad genome (plus
404 19 on average for the 1871 bp deleted E3 region) our data would be compatible with two pVII
405 molecules forming dimers at 238 discrete sites. This number would also come close to the ~200
406 "beads-on-string" estimated for purified viral cores [49].
407
408 **Viral chromatin dynamics drives early gene expression**
409 Here we show that nuclear delivery of viral genomes takes about 30 min but requires an
410 additional hour to initiate early gene expression. In agreement with earlier observations full
411 activation of early gene expression is reached after 4 h (reviewed in [50]). Correlating gene
412 expression with our differential MNase digestion conditions, we show that early genes exhibit
413 reduced pVII density and increased pVII repeat lengths. Direct detection of genomes and of the
414 E1A mRNA showed that genomes imported to the nucleus require about two hours to activate
415 gene expression (i.e. E1A transcription). In this lag phase pVII is selectively replaced by
416 nucleosomes at early genes, preceding gene activation. We identified that 80% of the imported
417 Ad genomes are transcriptionally inactive, but the remaining 20% do actively rearrange the
418 chromatin structure to allow transcription factor and RNA Polymerase binding. Once entering
419 the nucleus, we showed that the active viral genomes do selectively de-condense the two early
420 gene regions and curiously a specific domain in the Late3 gene. The overlap of open viral
421 chromatin with early transcription, suggests an organisation that allows binding of regulatory
422 factors for gene activation. For the Late3 region we observe increased DNA accessibility
423 overlapping with binding motifs for transcriptional regulators that may serve as enhancing
424 factors.
425
426 **pVII is exchanged for nucleosomes at early genes**

427 Genome decompaction correlates with low pVII occupancy and coincided with a site-specific
428 exchange of pVII protein for histone octamers. pVII molecules were exchanged for positioned
429 nucleosomes at the +1 site of actively transcribed genes (i.e. the E1A and E4 gene, Fig. 4D and
430 4E). The calculated number of nucleosomes and their positions on DNA, as determined by the
431 DNA fragment length of 147 bp, corresponds to the number of ~20% transcribing genomes,
432 exhibiting single positioned nucleosomes at the +1 position of early genes. As nucleosome
433 exchange is almost complete at the time point 2 hpi when transcription activation starts, we
434 suggest that nucleosome positioning is a prerequisite for gene activation. At this point we can
435 only speculate why only a subpopulation of genomes is transcriptionally active. Reasons could
436 be that the few genomes undergoing nucleosome assembly and active transcription produce the
437 replication enzymes whereas the bulk of genomes enters replication without activation as an
438 elegant way to avoid repeated chromatinization. Alternatively, cellular defence mechanisms
439 could supress transcription.
440 Interestingly, pVII removal to place nucleosomes at early genes preceded transcription,
441 suggesting that pVII removal is driving transcription and not vice versa as previously reported
442 [19]. Future work has to reveal which machinery drives the pVII nucleosome exchange and
443 determines specificity. Generally, active promoters are free of nucleosomes (nucleosome
444 depleted region, NDR) and the transcription start site is typically located upstream of a well-
445 positioned '+1' nucleosome influencing recruitment of the transcriptional machinery and
446 initiation of transcription [51,52]. In case of the Ad genome our data show that the E1A promoter
447 region is already pVII free and exists in an accessible nucleoprotein conformation when
448 entering the cell, potentially allowing fast assembly of the +1 nucleosome and enabling
449 immediate gene activation. This organization might thus be defined in the producer cell during
450 genome packaging, which occurs in a sequential packaging sequence driven manner [53,54]. In
451 contrast, other early regions were actively de-compacted with nuclear entry of the virus,
452 correlating with slightly delayed transcriptional onset. Whether this requires E1A expression is
453 not known.
454 Taken together our analysis provides unprecedented insight into the adenoviral nucleoprotein
455 structure and its dynamic reorganization during the onset of viral gene expression. We reveal
456 genome sequence, pVII-DNA occupancy and nucleosome positioning as driving forces for
457 transcriptional activation. Reproducing such organization in adenoviral vectors could result in
458 efficient and sustained transgene expression.

459 **EXPERIMENTAL PROCEDURES**

460 **RNA library preparation and sequencing**

461 Total cellular RNA was isolated from infected cells at 0, 0.5, 1, 2, 4 hpi using TRI Reagent
462 (Molecular Research Center, Inc.). Two time series sets of total RNA were prepared from two
463 independent infections. Library preparation, comprising rRNA depletion, polyA enrichment,
464 fragmentation to ~270 nucleotide length, reverse transcription to cDNA using random hexamer
465 primers and adapter ligation, was accomplished using TruSeq RNA Sample Preparation Kit v2
466 (Illumina) according to the manufacturers protocol by EMBL GeneCore facility in Heidelberg.
467 Libraries were sequenced on Illumina HiSeq2000 platform resulting in 18-75 Mio 50 bp single-
468 end reads per sample.

469 **Isolation of mono-nucleosomes, library preparation and sequencing**

470 H1299 cells infected with Ad were permeabilized with the detergent IGEPAL-CA630 and
471 digested at 0, 0.5, 1, 2 and 4 hpi with 100U/600U of MNase for 4/5 min at 37°, denoted as
472 'low'/'high' MNase. Mnase treatment was performed as described in [21]. After fragmented DNA
473 has been separated on 1.3% agarose gels, mono-nucleosomal fractions were isolated. Paired-
474 end sequencing on Illumina HiSeq2000 platform (EMBL GeneCore facility, Heidelberg),
475 yielding about 220 Mio read pairs per sample, was accomplished following library preparation
476 using NEBNext DNA library prep Master Mix Kit (New England Bioloabs)

477 **Viruses and virus production**

478 Bacterial artificial chromosome (BAC) mutagenesis was used to generate recombinant
479 adenoviruses following the protocol described in [55] and [23]. Briefly, bp 1-3513 of the HAdV-C5
480 genome including the left inverted terminal repeat (ITR) and the whole E1 region was replaced
481 by an FRT (Flippase Recognition Target) site and maintained as BAC (termed B12) to perform
482 lambda red recombination essentially as described. Plasmid pO6-Ad5-E1-FRT (encoding the
483 left ITR and the complete E1 region followed by an FRT recombination site) was used to clone
484 24 copies of the MS2 binding site repeats into the 3'UTR of the E1A gene. Subsequently flp-
485 recombination was used [55] to recombine plasmids pO6-Ad5-E1-FRT-MS2 and BAC B12 to
486 reconstitute BxAd5-E1-MS2 a BAC encoding the whole AdV genome including the E1A gene
487 tagged with 24 copies of the MS2 binding site. To generate the virus the BxAd5-E1-MS2 BAC
488 was digested by *PacI* restriction enzyme to release the viral genome and transfected into
489 HEK293 cells to reconstitute the recombinant viruses. Two rounds of plaque purification were
490 performed followed by large scale production in HEK293 cells and purification through double
491 $CsCl_2$ gradients banding. Replicative E3-deleted HAd-C5 virus or E1/E3-deleted HAd-C5 GFP

492 expressing control virus were essentially produced in the same way. Purified viruses were
493 extensively dialyzed against PBS/10% glycerol and stored in aliquots at -80°C and quantified
494 as physical particles per µL following the method described in [56].

495 **Cell culture and Infection**

496 U2OS (ATCC #HTB-96) and H1299 (ATCC #CRL-5803) and HEK293 (kindly provided by
497 G. Nemerow, TSRI, La Jolla USA) cells were maintained at 37°C and 5% CO_2 in Dulbecco's
498 modified Eagle's medium (DMEM)-Glutamax supplemented with 10% fetal calf serum (FCS)
499 and 1% of Penicillin/Streptavidin. U2OS cells stably expressing mCherry-TAF-Iβ [57] were
500 transduced with lentiviruses expressing MS2BP-NLS-GFP and double positive cells were
501 sorted by fluorescence activated cell sorting (FACS). The resulting cell population was
502 maintained without further selection. For control experiments using cytosolic MS2BP (Fig. S6)
503 stable mCherry-TAF-Iβ cells were transfected with an expression vector for MS2BP-NES-
504 GFP. For RNAscope and immunofluorescene (IF) analysis H1299 cells were grown on 15mm
505 glass coverslips. Subsequently cells were infected with 3000 physical particles/cell of
506 replicative E3-deleted HAd-C5 virus or E1/E3-deleted HAd-C5 GFP expressing control virus,
507 washed once with PBS and fixed at different time points with paraformaldehyde 4%/PBS for
508 10min at room temperature. Fixed cells were washed 3 times with PBS and slowly dehydrated
509 in 50%/70%/100% ethanol for 5 min each and stored at -20°C in 100% ethanol. For the
510 RNAscope protocol cells were rehydrated for 2 min in 70%/50% ethanol and equilibrated for
511 10 min in 1x PBS.

512 **RNAscope**

513 RNAscope experiment were performed using reagents and protocols from the
514 RNAscope® Multiplex Fluorescent Assay essentially as indicated by the supplier (ACD Bio).
515 Briefly, cells were incubated with Protease III (freshly diluted 1:30 with PBS) for 15 min at
516 room temperature (RT) and washed 3 times with PBS. Hybridisation with the E1A RNAscope
517 probe was performed for 2h at 40° in a humidified chamber. Hybridized cells were washed
518 twice for 2 min and hybridized with signal amplifier probes (final amplifier Amp 4 AltB-FL)
519 following the manufacturer's instructions. Hybridized cells were washed twice in PBS followed
520 by incubation in IF buffer (PBS/10% FCS/0.1% Saponin) for 15 min at room temperature to
521 saturate unspecific binding sites followed by IF analysis.

522 **Immunofluorescence**

523 Cells (infected or controls) grown on cover slips were fixed in 4% EM-grade PFA at indicated
524 time points and stored in cold PBS until use. For IF, cover slips were initially incubated in IF

525 buffer (PBS/10% FCS/0.1% Saponin) for 15min and then replaced by antibody solution.
526 Primary antibodies (mouse anti-pVI, mouse anti-pVII and rabbit anti Ad5- capsids) were
527 diluted in IF buffer and incubated with cells for 1h at RT. Antibody was removed by washing
528 in PBS followed by incubation with secondary antibodies (mouse Alexa 647 and rabbit Alexa
529 594) diluted in IF buffer for 1h at RT. Cells were washed with PBS, ddH_2O and absolute
530 ethanol, air-dried and mounted in DAKO mixed with DAPI (1:1000).

531 **Live-cell imaging and image analysis**

532 For live-cell imaging, U2OS cells stably expressing mCherry-TAF-Iβ and MS2BP-NLS-GFP
533 were seeded in ibidi μ-slide $VI^{0.4}$ (Ibidi). The next day, the medium in the ibidi μ-slide was
534 replaced by imaging medium (FluoBrite DMEM supplemented with Prolong antifade reagent,
535 both Life technologies) containing 3000 physical particles/cell of replicative BxAd5-E1-MS2
536 for 30min followed by washing with imaging medium to remove the inoculum. Infection and
537 subsequent cultivation were performed at 37˚C. Cells were imaged using a Leica spinning-disk
538 microscopy system (x100 objective) equipped with an environmental chamber at a frame rate
539 of 1 frame per second using MetaMorph software. To generate kymographs, movies were
540 processed using the ImageJ KymographBuilder plugin
541 (https://imagej.net/KymographBuilder_:_Yet_Another_Kymograph_Fiji_plugin) on confocal
542 sections. Line width was adapted to cover all viral genomes throughout the length of the movie
543 and double positive genomes were scored as transcribing genomes. Mounted cells for IF and
544 RNAscope analysis were imaged using a SP8 confocal microscope (Leica) at the Bordeaux
545 Imaging Center (BIC) platform. Stacks (7x) were taken every 0,3µm using a pinhole of 1 at a
546 pixel size of 85 nm. Quantifications were done using an adapted semi-automated macro in
547 ImageJ (provided upon request). Briefly, channels were split and stacks were Z-projected and
548 the cell periphery was outlined manually. A threshold was applied to every channel and single
549 objects or colocalized objects exceeding a defined area of pixels were counted. Statistical
550 analysis was done using one-way ANOVA multicomparison test.
551

552 # QUANTIFICATION AND STATISTICAL ANALYSIS

553 **Virus annotation**

554 The reference genome sequence and gene annotation of Human Adenovirus C serotype 5
555 (accession: AY339865.1) was downloaded from NCBI. The annotation was adapted to the used
556 Adenovirus strain and nucleotides 28,593-30,464 were excised from the full version (available
557 at github project page).

558 **RNA-seq analysis**

559 STAR [58] was used to align the RNA-seq reads simultaneously to the human and the adenoviral
560 genome. Therefore, a STAR index was generated based on the hg19 reference, the modified
561 adenoviral genome (see section *Virus annotation* for details) and the GENCODE human
562 transcript annotation version 25. Read mapping was conducted using following parameters:

563 --outSAMmultNmax 1
564 --outMultimapperOrder Random
565 --outFilterType BySJout
566 --outFilterIntronMotifs RemoveNoncanonical
567 --outSAMstrandField intronMotif
568 --outFilterMultimapNmax 20
569 --alignSJoverhangMin 8
570 --alignSJDBoverhangMin 1
571 --alignIntronMin 20
572 --alignIntronMax 1000000

573

574 To quantify the RNA abundance for each gene the number of uniquely mapped reads per gene
575 was counted using the featureCounts function of the subread package [59]. The read counts were
576 normalized to transcripts per million (TPMs).

577 **MNase-seq analysis**

578 A bowtie2 index comprised of the human reference genome hg19 and the sequence of the
579 used Adenovirus strain (see section *Virus annotation* for details) was built (code available at
580 [60]) and paired-end reads were mapped to the index using bowtie2-aligner [61] with following
581 parameters:

582 --very-sensitive-local
583 --no-discordant

584

585 Reads were filtered to mapping quality ≥ 30 and only reads mapped in proper pair were kept
586 using samtools [62]. Reads aligned to the Adenovirus genome were extracted for further
587 analysis. The fragment size distribution was plotted using the CollectInsertSizeMetrics
588 function from the Picard toolkit (GitHub Repository. http://broadinstitute.github.io/picard/;
589 Broad Institute).

590 **pVII positioning analysis**

591 Fragments mapped to the Adenovirus reference with a maximum insert size of 139 bp were
592 selected and used for pVII positioning analysis. pVII positions were called using the
593 DANPOS2 toolkit [33]. First, a pVII occupancy profile for each sample was generated and
594 normalized to 20,000 reads using the dpos function with following parameters: -u 1e-5; -a 1; -
595 z 5; -p 0.001, --extend 35 and --mifrsz 0. Next, the profiles were normalized by a genome
596 wide quantile normalization to the first replicate at 0 hpi of each MNase condition (high/low)
597 using the wiq script. Finally, pVII positions were obtained from the normalized pVII
598 occupancy profiles using the dpos function with following parameters: -jd 50; -q 10; -z 5; -a
599 1.

600 **Rotational positioning analysis**

601 49 to 51 bp sized fragments mapped to the Adenovirus reference of all high MNase samples
602 were selected and pooled together. Frequencies of A/T or G/C di-nucleotides around the
603 fragment center were calculated using the annotatePeaks.pl script from the HOMER suit [63].
604 Noise was filtered using the Fast Fourier Transform build in the filterFFT function of the
605 nucleR Bioconductor package [64]. The cutoff was set with the option pcKeepComp=0.2.
606 Finally, to reflect the symmetry of the pVII binding the reverse di-nucleotide frequency was
607 added, and the total frequency was divided by 2. The same procedure was conducted to derive
608 the di-nucleotide periodicity of nucleosomal fragments, except that 147 bp fragments of the
609 first replica at 0 hpi were selected.

610 **Sites of nucleosome assembly**

611 The R Bioconductor package nucleR [64] was used to call sites of nucleosome assembly onto
612 the Ad genome. First, fragments specified to the Ad genome were loaded into the R
613 environment. Fragments between 137 to 157 bp from time points 1, 2 and 4 hpi of the high
614 MNase condition were extracted for further analysis. The coverage was normalized to reads
615 per million Ad mapped reads and noise was removed using the built-in Fast Fourier
616 Transform (filterFFT with option pcKeepComp=0.007). Potential nucleosome positions were
617 called based on the normalized profile using the peakDetection function with the option
618 width=147. Peaks showing at least a 5-fold enrichment compared to the background signal at
619 1 hpi were considered as nucleosome positions.

620 **Ad genome accessibility**

621 DNA fragments released under low MNase conditions were used to identify structural
622 changes of Ad chromatin over time. Therefore, a chromatin accessibility score was calculated

623 along the Ad genome using a 300 bp bin sliding along the Ad genome by 20 bp. The
624 accessibility score was obtained for each bin in the genome by fitting a linear regression on
625 the normalized read frequencies computed for each time point (average of both replicates).
626 The slope of the linear fit was used as a measure of dynamic changes of viral genome
627 accessibility.

628 **Comparison of nucleosome assembly sites and Ad genome accessibility**

629 The cumulative accessibility score at the sites of nucleosome assembly was calculated. To
630 compare how likely this score is obtained by chance, the nucleosome peaks were randomly
631 shuffled along the Ad genome and the cumulative accessibility score was computed. This
632 process was iterated 10.000 times.

633 **Transcription factor motif identification**

634 The Ad DNA sequence from position 19,890 to 19,940 was extracted and uploaded to the
635 PROMO web server [65]. Motif search was performed against known human and mouse motifs
636 stored in TRANSFAC data base using the SearchSites tool. The maximum dissimilarity rate
637 was set to 2 %. The obtained motifs were further filtered to have at least a recognition
638 sequence of 5 nucleotides and that the corresponding transcription is expressed in the infected
639 H1299 cells.
640
641

642 **ACCESSION NUMBERS**

643 The Gene Expression Omnibus accession number for the MNase-seq, and RNA-seq data
644 reported in this paper is GSE136550. The code used to analyze the sequencing data is
645 available at GitHub: https://github.com/uschwartz/AdVir

646

647 **SUPPLEMENTAL INFORMATION**

648 Supplemental Information includes 6 figures, 4 movies, 2 tables, and can be found with this
649 article online at http://xxxxx

650

651 **AUTHOR CONTRIBUTIONS**

652 Conceptualization H.W. and G.L.; Methodology F.L., T.K., E.S., C.H., E.B., H.W. and G.L.;
653 Formal Analysis F.L., U.S. and H.W.; Investigation T.K., C.H., F.L., E.S., H.W. and G.L.,
654 Data Curation U.S; Visualization F.L., U.S. and H.W.; Writing – Original Draft U.S. G.L. and
655 H.W.; Writing – Review & Editing E.B., T.K., U.S., G.L. and H.W., Resources E.B., H.W
656 and G.L; Supervision H.W and G.L; Funding Acquisition G.L. and H.W.

657

658 **ACKNOWLEDGEMENTS**

659 The study was financed by the BMBF 01DN17003 (to G.L.), the DFG SFB960 (to G.L.), the
660 Fondation pour la Recherche Médicale (Equipes FRM 2018 DEQ20180339229, to H.W.) and
661 by Grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports,
662 Science and Technology of Japan (19H04838, to T.K.). Some of the microscopy was done in
663 the Bordeaux Imaging Center, a service unit of the CNRS-INSERM, and Bordeaux
664 University, a member of the national infrastructure France BioImaging. HW is an INSERM
665 fellow.

666

667

668 **REFERENCES**

669 1. Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. Integrins alpha v beta
670 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell*
671 **73**, 309–19 (1993).
672 2. Wiethoff, C. M., Wodrich, H., Gerace, L. & Nemerow, G. R. Adenovirus protein VI
673 mediates membrane disruption following capsid disassembly. *J. Virol.* **79**, 1992–2000
674 (2005).
675 3. Strunze, S. *et al.* Kinesin-1-Mediated Capsid Disassembly and Disruption of the

676 Nuclear Pore Complex Promote Virus Infection. *Cell Host Microbe* **10**, 210–223
677 (2011).
678 4. Cassany, A. *et al.* Nuclear Import of Adenovirus DNA Involves Direct Interaction of
679 Hexon with an N-Terminal Domain of the Nucleoporin Nup214. *J. Virol.* **89**, 1719–
680 1730 (2015).
681 5. Flint, S. J. Expression of adenoviral genetic information in productively infected cells.
682 *Biochim. Biophys. Acta* **651**, 175–208 (1982).
683 6. Akusjarvi, G. Temporal regulation of adenovirus major late alternative RNA splicing.
684 *Frontiers in Bioscience* **13**, 5006–5015 (2008).
685 7. Brown, D. T., Westphal, M., Burlingham, B. T., Winterhoff, U. & Doerfler, W.
686 Structure and composition of the adenovirus type 2 core. *J. Virol.* **16**, 366–87 (1975).
687 8. Corden, J., Engelking, H. M. & Pearson, G. D. Chromatin-like organization of the
688 adenovirus chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 401–4 (1976).
689 9. Benevento, M. *et al.* Adenovirus composition, proteolysis, and disassembly studied by
690 in-depth qualitative and quantitative proteomics. *J. Biol. Chem.* **289**, 11421–30 (2014).
691 10. Johnson, J. S. *et al.* Adenovirus Protein VII Condenses DNA, Represses Transcription,
692 and Associates with Transcriptional Activator E1A. *J. Virol.* **78**, 6459–6468 (2004).
693 11. Keller, M., Tagawa, T., Preuss, M. & Miller, A. D. Biophysical characterization of the
694 DNA binding and condensing properties of adenoviral core peptide mu. *Biochemistry*
695 **41**, 652–9 (2002).
696 12. Anderson, C. W., Young, M. E. & Flint, S. J. Characterization of the adenovirus 2
697 virion protein, mu. *Virology* **172**, 506–12 (1989).
698 13. Mirza, M. A. & Weber, J. Structure of adenovirus chromatin. *Biochim. Biophys. Acta*
699 **696**, 76–86 (1982).
700 14. Puntener, D. *et al.* Stepwise Loss of Fluorescent Core Protein V from Human
701 Adenovirus during Entry into Cells. *J. Virol.* **85**, 481–496 (2011).
702 15. Haruki, H., Gyurcsik, B., Okuwaki, M. & Nagata, K. Ternary complex formation
703 between DNA-adenovirus core protein VII and TAF-Ibeta/SET, an acidic molecular
704 chaperone. *FEBS Lett.* **555**, 521–7 (2003).
705 16. Komatsu, T., Haruki, H. & Nagata, K. Cellular and viral chromatin proteins are
706 positive factors in the regulation of adenovirus gene expression. *Nucleic Acids Res.* **39**,
707 889–901 (2011).
708 17. Xue, Y., Johnson, J. S., Ornelles, D. A., Lieberman, J. & Engel, D. A. Adenovirus
709 Protein VII Functions throughout Early Phase and Interacts with Cellular Proteins SET

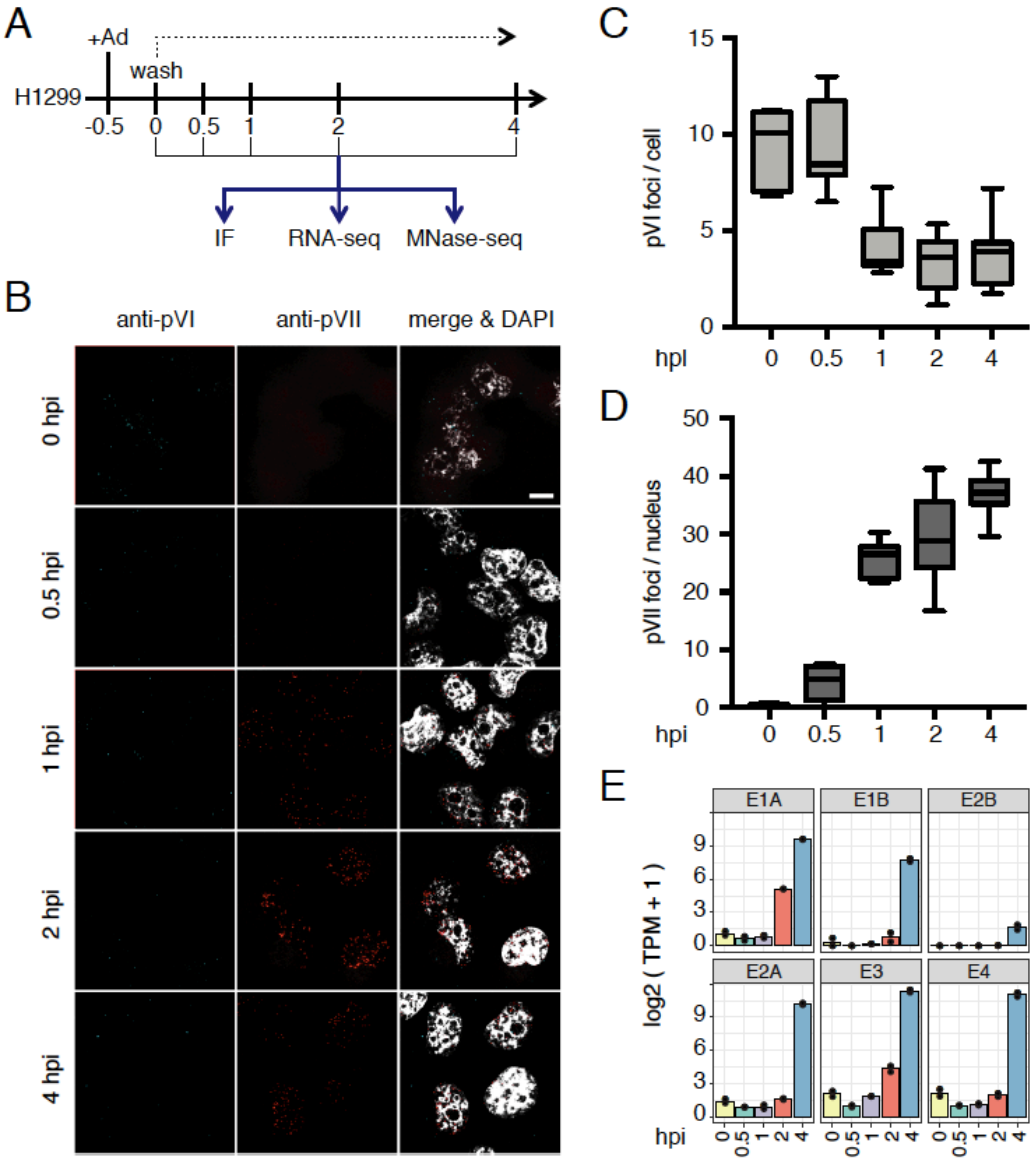
710 and pp32. *J. Virol.* **79**, 2474–2483 (2005).
711 18. Komatsu, T., Dacheux, D., Kreppel, F., Nagata, K. & Wodrich, H. A method for
712 visualization of incoming adenovirus chromatin complexes in fixed and living cells.
713 *PLoS One* **10**, e0137102 (2015).
714 19. Chen, J., Morral, N. & Engel, D. A. Transcription releases protein VII from adenovirus
715 chromatin. *Virology* **369**, 411–22 (2007).
716 20. Chatterjee, P. K., Vayda, M. E. & Flint, S. J. Adenoviral protein VII packages
717 intracellular viral DNA throughout the early phase of infection. *EMBO J.* **5**, 1633–1644
718 (1986).
719 21. Schwartz, U. *et al.* Characterizing the nuclease accessibility of DNA in human cells to
720 map higher order structures of chromatin. *Nucleic Acids Res.* **47**, 1239–1254 (2018).
721 22. Mueller, B. *et al.* Widespread changes in nucleosome accessibility without changes in
722 nucleosome occupancy during a rapid transcriptional induction. *Genes Dev.* **31**, 451–
723 462 (2017).
724 23. Martinez, R. *et al.* The Amphipathic Helix of Adenovirus Capsid Protein VI
725 Contributes to Penton Release and Postentry Sorting. *J. Virol.* **89**, 2121–2135 (2015).
726 24. Carlon-Andres, I. *et al.* Nup358 and transportin 1 cooperate in adenoviral genome
727 import. *J. Virol.* (2020). doi:10.1128/JVI.00164-20
728 25. Glenn, G. M. & Ricciardi, R. P. Detailed kinetics of adenovirus type-5 steady-state
729 transcripts during early infection. *Virus Res.* **9**, 73–91 (1988).
730 26. Crisostomo, L., Soriano, A. M., Mendez, M., Graves, D. & Pelka, P. Temporal
731 dynamics of adenovirus 5 gene expression in normal human cells. *PLoS One* **14**,
732 e0211192 (2019).
733 27. Nevins, J. R., Ginsberg, H. S., Blanchard, J. M., Wilson, M. C. & Darnell, J. E.
734 Regulation of the primary expression of the early adenovirus transcription units. *J.*
735 *Virol.* **32**, 727–33 (1979).
736 28. Noll, M., Thomas, J. O. & Kornberg, R. D. Preparation of Native Chromatin and
737 Damage Caused by Shearing. *Science (80-.).* **187**, 1203–1206 (1975).
738 29. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human
739 Genome. *Cell* **132**, 887–898 (2008).
740 30. Diermeier, S. *et al.* TNFα signalling primes chromatin for NF-κB binding and induces
741 rapid and widespread nucleosome repositioning. *Genome Biol.* **15**, 536 (2014).
742 31. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells.
743 *Nature* **474**, 516–520 (2011).

744 32. Daniell, E., Groff, D. E. & Fedor, M. J. Adenovirus chromatin structure at different
745 stages of infection. *Mol. Cell. Biol.* **1**, 1094–1105 (1981).
746 33. Chen, K. *et al.* DANPOS: Dynamic analysis of nucleosome position and occupancy by
747 sequencing. *Genome Res.* **23**, 341–351 (2013).
748 34. Gao, J., Mese, K., Bunz, O. & Ehrhardt, A. State-of-the-art human adenovirus
749 vectorology for therapeutic approaches. *FEBS Letters* **593**, 3609–3622 (2019).
750 35. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome.
751 *Nature* **458**, 362–366 (2009).
752 36. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev.*
753 *Biophys.* **34**, 269–324 (2001).
754 37. Komatsu, T. & Nagata, K. Replication-Uncoupled Histone Deposition during
755 Adenovirus DNA Replication. *J. Virol.* **86**, 6701–6711 (2012).
756 38. Längst, G., Schätz, T., Langowski, J. & Grummt, I. Structural analysis of mouse
757 rDNA: Coincidence between nuclease hypersensitive sites, DNA curvature and
758 regulatory elements in the intergenic spacer. *Nucleic Acids Res.* **25**, 511–517 (1997).
759 39. Mieczkowski, J. *et al.* MNase titration reveals differences between nucleosome
760 occupancy and chromatin accessibility. *Nat. Commun.* **7**, 11485 (2016).
761 40. Chereji, R. V., Bryson, T. D. & Henikoff, S. Quantitative MNase-seq accurately maps
762 nucleosome occupancy levels. *Genome Biol.* **20**, 198 (2019).
763 41. Kent, N. A., Adams, S., Moorhouse, A. & Paszkiewicz, K. Chromatin particle
764 spectrum analysis : a method for comparative chromatin structure analysis using
765 paired-end mode next-generation DNA sequencing. *Nucleic Acids Res.* **39**, (2011).
766 42. Chereji, R. V., Ocampo, J. & Clark, D. J. MNase-Sensitive Complexes in Yeast:
767 Nucleosomes and Non-histone Barriers. *Mol. Cell* **65**, 565-577.e3 (2017).
768 43. Pheasant, K. *et al.* Nuclear-cytoplasmic compartmentalization of the herpes simplex
769 virus 1 infected cell transcriptome is co-ordinated by the viral endoribonuclease vhs
770 and cofactors to facilitate the translation of late proteins. *PLoS Pathog.* **14**, (2018).
771 44. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**,
772 437–445 (1998).
773 45. Lischwe, M. A. & Sung, M. T. A histone-like protein from adenovirus chromatin [37].
774 *Nature* **267**, 552–554 (1977).
775 46. Sato, K. & Hosokawa, K. Analysis of the interaction between DNA and major core
776 protein in adenovirus chromatin by circular dichroism and ultraviolet light induced
777 cross-linking. *J. Biochem.* **95**, 1031–9 (1984).

778 47. Everitt, E., Lutter, L. & Philipson, L. Structural proteins of adenoviruses. XII. Location
779 and neighbor relationship among proteins of adenovirion type 2 as revealed by
780 enzymatic lodination, immunoprecipitation and chemical cross-linking. *Virology* **67**,
781 197–208 (1975).
782 48. Chatterjee, P. K., Vayda, M. E. & Flint, S. J. Interactions among the three adenovirus
783 core proteins. *J. Virol.* **55**, 379–86 (1985).
784 49. Vayda, M. E., Rogers, A. E. & Flint, S. J. The structure of nucleoprotein cores released
785 from adenovirions. *Nucleic Acids Res.* **11**, 441–60 (1983).
786 50. Pied, N. & Wodrich, H. Imaging the adenovirus infection cycle. *FEBS Lett.* **593**, 3419–
787 3448 (2019).
788 51. Kubik, S. *et al.* Nucleosome Stability Distinguishes Two Different Promoter Types at
789 All Protein-Coding Genes in Yeast. *Mol. Cell* **60**, 422–434 (2015).
790 52. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through
791 genomics. *Nat. Rev. Genet.* **10**, 161–72 (2009).
792 53. Hearing, P., Samulski, R. J., Wishart, W. L. & Shenk, T. Identification of a repeated
793 sequence element required for efficient encapsidation of the adenovirus type 5
794 chromosome. *J. Virol.* **61**, 2555–8 (1987).
795 54. Hammarskjöld, M. L. & Winberg, G. Encapsidation of adenovirus 16 DNA is directed
796 by a small DNA sequence at the left end of the genome. *Cell* **20**, 787–795 (1980).
797 55. Ruzsics, Z., Lemnitzer, F. & Thirion, C. Engineering adenovirus genome by bacterial
798 artificial chromosome (BAC) technology. *Methods Mol. Biol.* **1089**, 143–58 (2014).
799 56. Mittereder, N., March, K. L. & Trapnell, B. C. Evaluation of the concentration and
800 bioactivity of adenovirus vectors for gene therapy. *J. Virol.* **70**, 7498–509 (1996).
801 57. Komatsu, T., Nagata, K. & Wodrich, H. An Adenovirus DNA Replication Factor, but
802 Not Incoming Genome Complexes, Targets PML Nuclear Bodies. *J. Virol.* **90**, 1657–
803 67 (2016).
804 58. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
805 (2013).
806 59. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program
807 for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
808 60. Schwartz, U. & Längst, G. Bioinformatic Analysis of ChIP-seq Data on the Repetitive
809 Ribosomal RNA Gene. *Encycl. Cell Biol.* **3**, 254–269 (2015).
810 61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.
811 Methods* **9**, 357–9 (2012).

812 62. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
813 2078–9 (2009).
814 63. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
815 cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**,
816 576–89 (2010).
817 64. Flores, O. & Orozco, M. nucleR: a package for non-parametric nucleosome
818 positioning. *Bioinformatics* **27**, 2149–2150 (2011).
819 65. Messeguer, X. *et al.* PROMO: detection of known transcription regulatory elements
820 using species-tailored searches. *Bioinformatics* **18**, 333–334 (2002).
821

822 **FIGURES AND LEGENDS**



823

824 **Figure 1**. Spatio-temporal analysis of adenovirus infection. **A)** Experimental design. Virus was
825 added (+Ad) and removed (Wash) after 30 min and cell were processed as indicated (hpi =
826 hours post infection) using immunofluorescence (IF), RNA sequencing (RNA-seq) and
827 nucleosome sequencing (MNase-seq). **B)** Images show representative confocal images at
828 indicated time points stained with anti-pVI (left column, cyan signal) and anti-pVII (middle
829 column, red signal) antibodies and merged with DAPI detection of the nucleus (right column,
830 grey signal). Scale bar is 10µm. **C)** Quantification of the pVI signal over time as indicated on
831 the x-axis. The distribution of pVI foci per cell are shown as Box and Whiskers plot (n > 44
832 cells per time point). **D)** As in (C) showing the quantification of pVII foci per nucleus from the
833 same experiment (n > 44 nuclei per time point). **E)** Time resolved analysis of Ad transcription
834 using RNA-seq. TPMs (Transcripts Per Kilobase Million) for early expressed genes were
835 calculated at each timepoint. The barplots show the average of two replicates and each replicate
836 is indicated by a black dot.

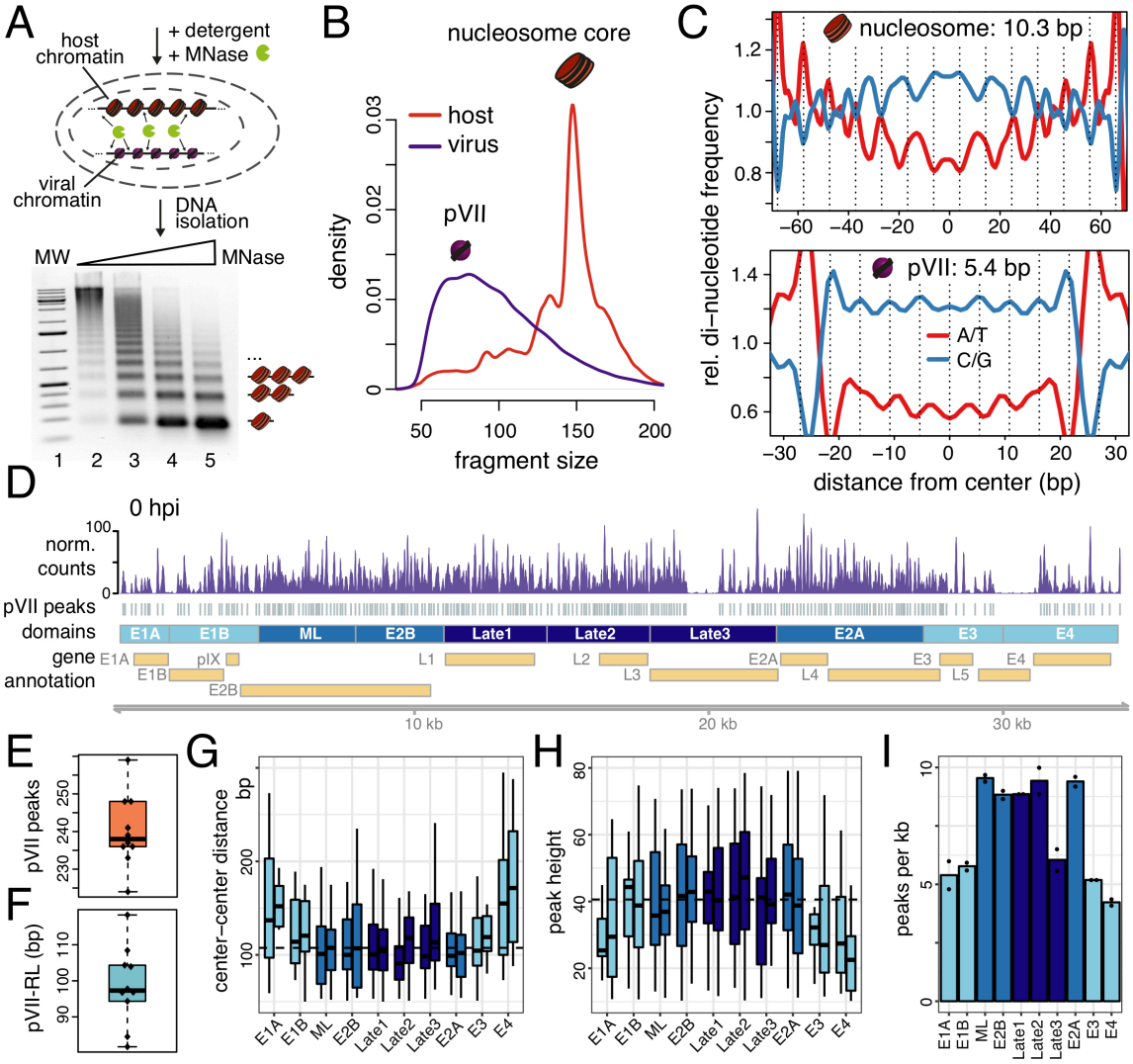


Figure 2. Characterization of Ad DNA packaging by pVII. **A)** Schematic view of MNase digestion of infected cells. Agarose gel shows nucleosome ladders of a MNase titration. Mono-, di- and tri-nucleosomes are indicated at the right. The mono-nucleosomal DNA fraction was isolated after MNase digestions and subjected to high-throughput sequencing. **B)** Fragment size distribution of paired-end reads either mapped to the human genome (red) or the Ad genome (purple). **C)** Average A/T- and G/C-di-nucleotide frequency of nucleosomal (upper panel) and pVII fragments (lower panel). Size-selected fragments were centred at the midpoint and the di-nucleotide frequency was calculated. Every 10.3 (nucleosome) / 5.4 (pVII) bp a dashed line highlights the periodicity of the signal. **D)** Profile illustrates Ad genome coverage by MNase-seq reads. pVII positions, which have been called with the DANPOS2 package [33], are represented by grey boxes. Partitioning of the genome into functional domains and the timing of their transcription is indicated by the color code (lightblue = early, darkblue=late). Genomic location and underlying gene annotation of HAd-C5 (NCBI accession: AY339865.1) are indicated at the bottom. **E)** Boxplot showing the number of detected pVII peaks for each sample. **F)** Boxplot showing the genome-wide pVII repeat length (pVII-RL) for each sample. **G)** and **H)** Boxplots showing the distribution of pVII center-center distances / peak heights in each domain. Individual replicates are shown side by side. **I)** Barplot showing the number of pVII peaks per kb detected in the domains. Individual replicates are shown as black dots.

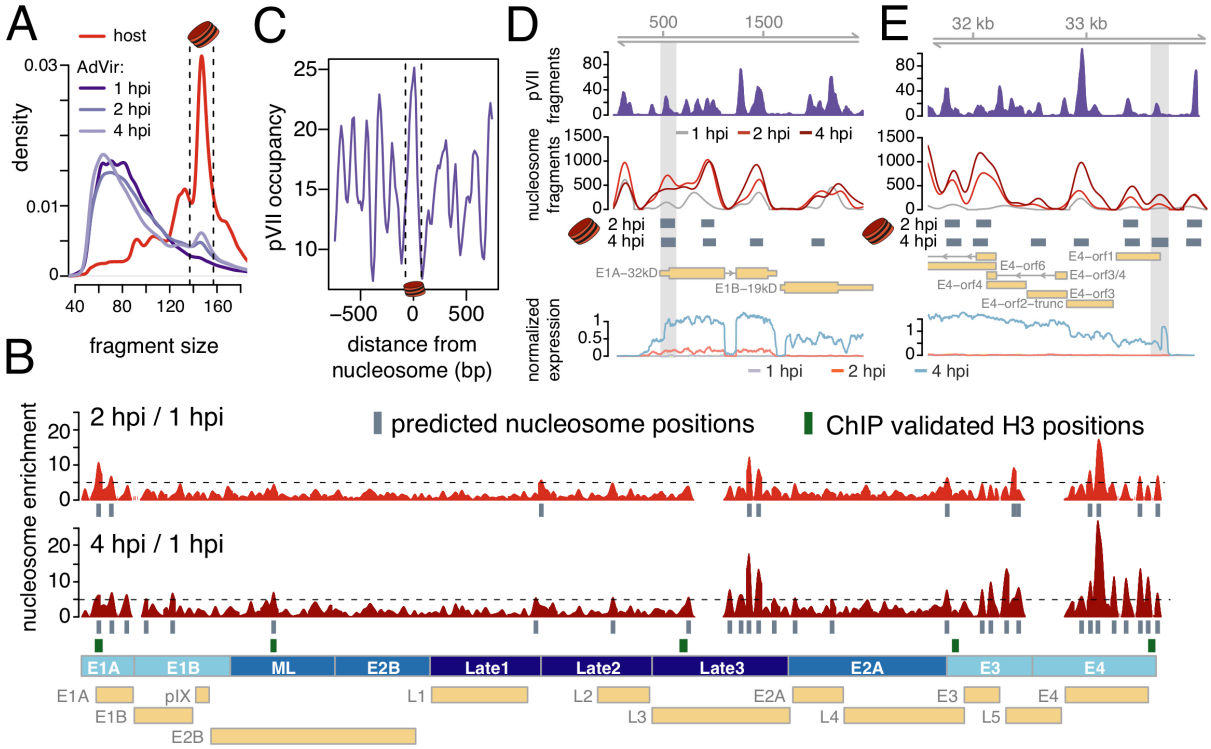


Figure 3. Nucleosome assembly onto the Ad genome. **A)** Fragment size distribution of paired-end reads either mapping to the human genome (red) or the Ad genome (shades of purple). Nucleosome sized fragments are highlighted between two dashed lines at 137 bp and 157 bp. **B)** Enrichment of nucleosome sized fragments across the Ad genome. Enrichment was calculated as fold-change of 137-157 bp sized fragments at 2 hpi / 4 hpi to a background level at 1 hpi. Dashed lines indicate the threshold used for histone peak calling. Nucleosome positions are represented by grey boxes. H3 ChIP-qPCR verified regions [37] are shown as green boxes. **C)** Average pVII occupancy at 0 hpi around predicted nucleosome positions. Dashed lines indicate nucleosome boundaries. **D)** and **E)** Genome browser tracks showing pVII occupancy at 0 hpi (upper panel), the abundance of nucleosome sized fragments (137-157 bp, middle track) from 1 to 4 hpi, predicted sites of nucleosome assembly (grey boxes) and expression levels of early genes (bottom panel).

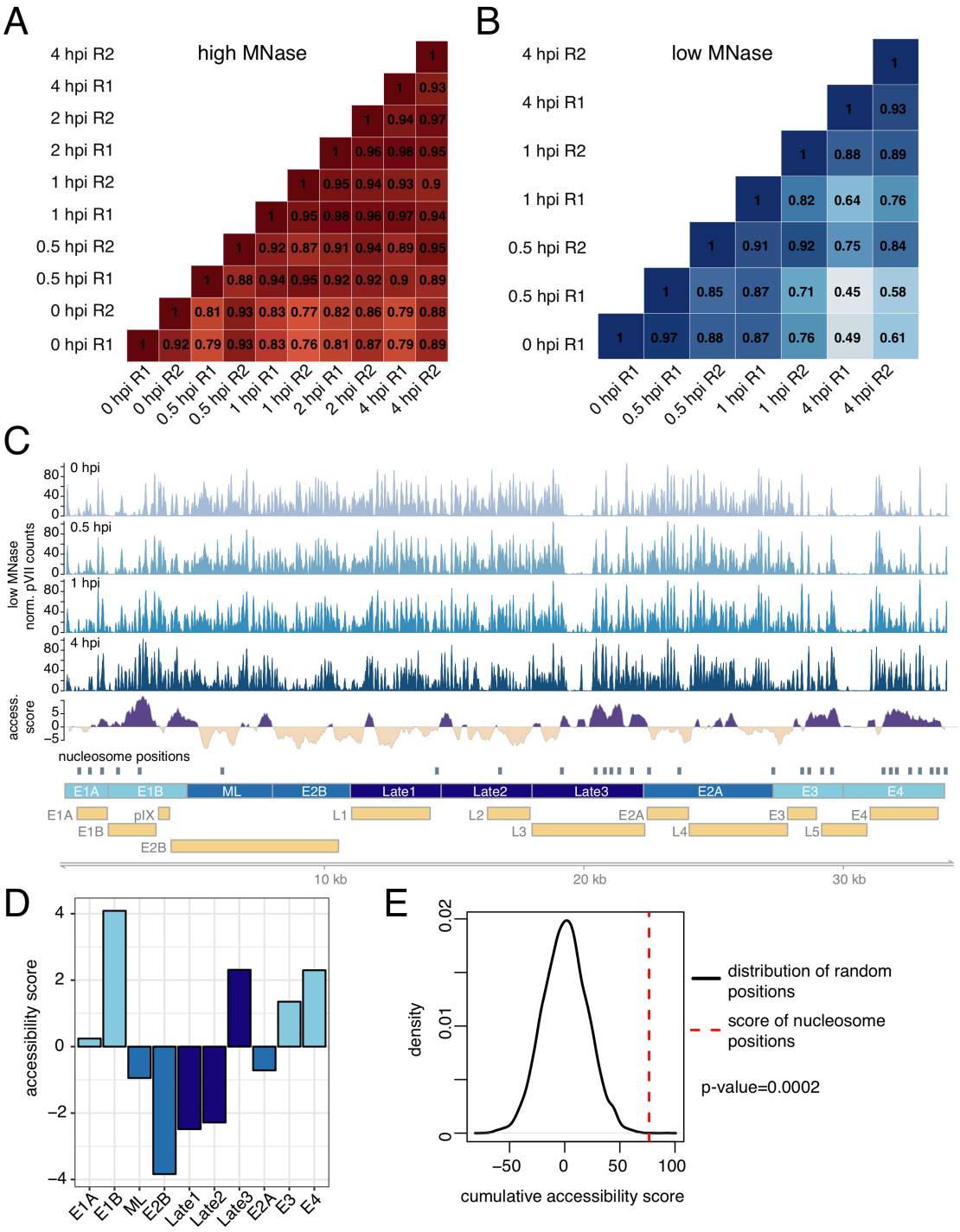


Figure 4. Early decondensation of Ad chromatin. **A)** and **B)** Sample-wise Peason correlations of **A)** high MNase or **B)** low MNase profiles. **C)** Profile illustrates Ad genome coverage by low MNase-seq fragments. The average of two replicates is shown, except at timepoint 0 hpi where only one replicate was available. The accessibility score was calculated as the slope of a linear regression fit of the average signal intensity across each timepoint. The score was assessed in 300-bp sliding windows. Nucleosome positions at 4 hpi derived from Fig. 4B are shown. **D)** Barplot showing the average accessibility score in each domain. **E)** Density distribution of the cumulative accessibility score of sampled peaks. The cumulative accessibility score of the estimated nucleosome positions at the Ad genome is highlighted by a dashed red line.

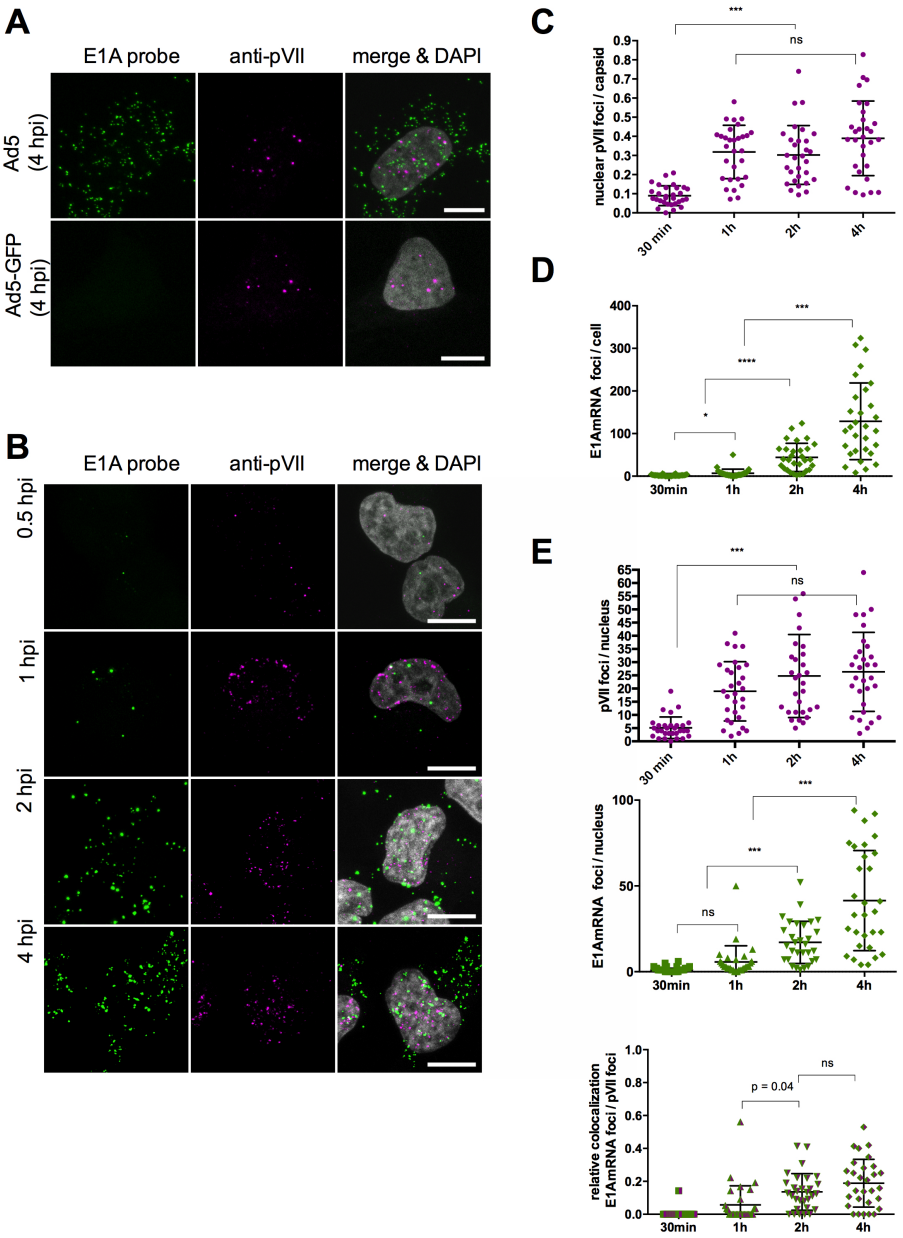


Figure 5. E1A transcription at the single cell level. **A)** Representative confocal images of cells infected with HAd-C5 (Ad5, top row) or an E1-deleted HAd-C5 vector (Ad5-GFP, bottom row). Detection of E1A transcript with specific RNAscope probes (left column, green signal), genomes using anti-pVII antibodies (middle column, magenta signal) and merged with DAPI detecting the nucleus (right column, grey signal). Scale bar is 10µm. **B)** Representative confocal images as in (A) of cells infected with HAd-C5 at different time points as indicated. Scale bar is 10µm. **C)** Quantification of the ratio of nuclear pVII foci normalized by cell associated capsid signals over time as indicated on the x-axis. The distribution of normalized pVII foci at each time point is shown as scatter plot (Mean +/- SD, n = 30 cells per time point). **D)** Graph as in (C) showing the distribution of E1A transcript foci per cell as scatter plot (Mean +/- SD, n = 30 cells per time point). **E)** Graph as in (C) showing the distribution of pVII foci per nuclei (top graph) and E1A transcript foci per nuclei (middle graph) as scatter plot (Mean +/- SD, n = 30 nuclei per time point). The bottom graph was generated by normalizing the number of E1A positive genomes by the total number of genomes per nucleus shown as scatter plot (Mean +/- SD, n = 30 nuclei per time point). Statistical analysis for C-E via one-way ANOVA multi-comparison test.

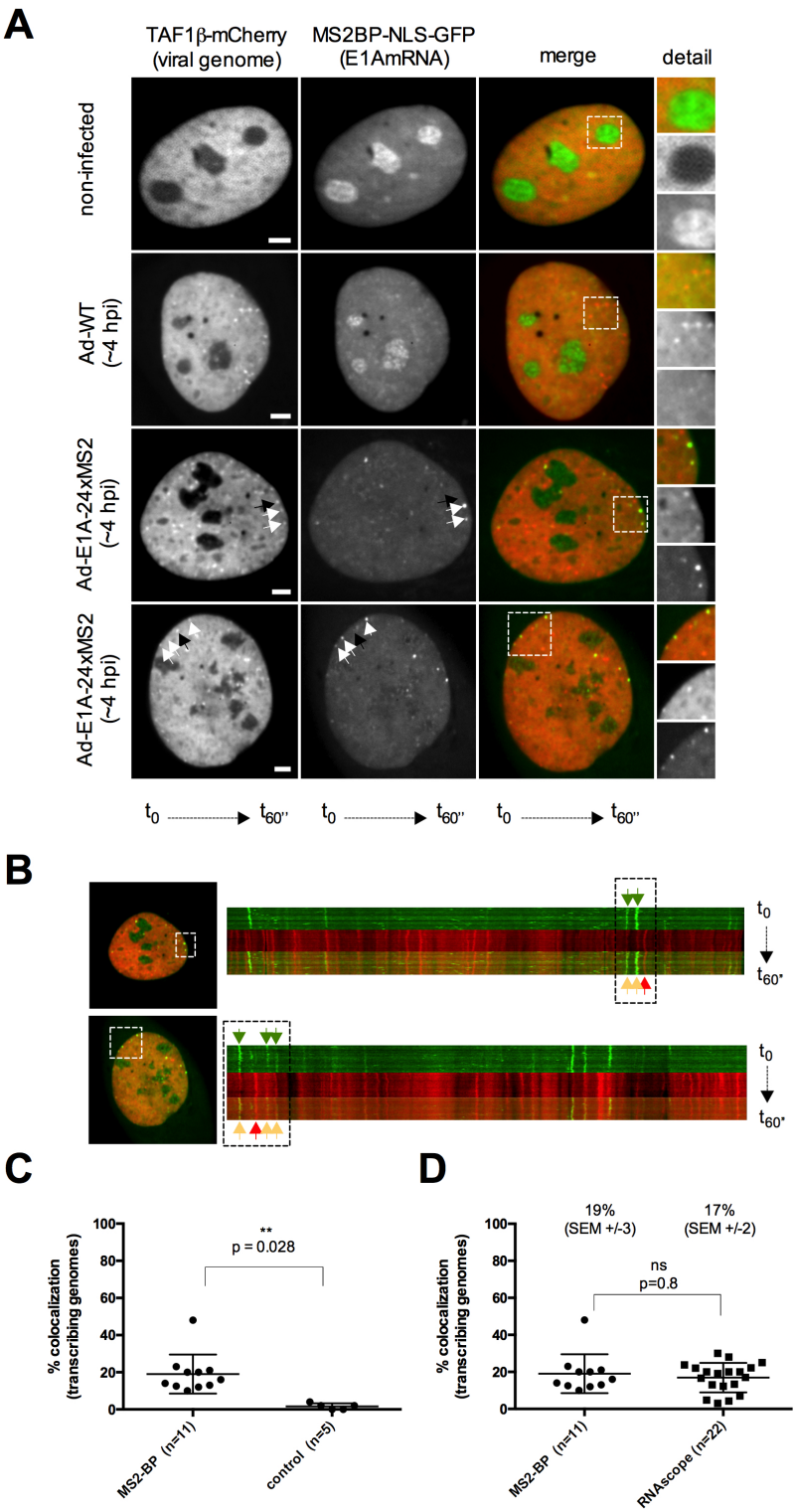


Figure 6. E1A transcription in living cells. **A)** Representative images showing t-projections of 60 frames (frame rate 1F/sec) imaged at ~4 hpi. Cells express mCherry-TAF-Iβ showing viral genomes (first column, red signal in merge) and MS2BP-NLS-GFP showing MS2-repeat tagged E1AmRNAs (second column, green signal in merge) and an overlay (third column, merge) of both channels. The first row shows non-infected control cells, the second row cells infected with untagged control virus and the third and fourth row examples of cells infected with virus expressing E1AmRNA tagged with MS2-repeats. Black arrows point at non-

transcribing and white arrows at transcribing genomes. Details are shown as magnifications of the boxed area. Scale bar is 10µm. **B)** Kymograph analysis of the bottom two cells shown in A and as described in the material and methods section. Cell overview to the left, kymograph separated into the individual channels and a superimposition of both channels to the right. Genome position over time appears as red vertical line, E1AmRNA signal as green vertical line. The boxed area in the overview corresponds to the boxed area in the kymograph. Green arrows point at the E1AmRNA signal, orange arrows at transcribing (double positive) and red arrow at non-transcribing (single positive) genomes. **C)** Quantitative kymograph analysis. Scatter plot representation of the percentage of MS2-BP positive genomes per cell (n = 11 cells) vs. control cells infected with viruses expressing untagged E1AmRNA (n = 5 cells) determined by kymograph analysis. **D)** Scatter plot representation comparing the percentage of E1AmRNA positive genomes per cell determined with the MS2-BP method (left, n = 11 cells) vs. the RNAscope method (right, n = 22 cells). Unpaired t-test was used for statistical analysis in C and D.