

1 LATENCY, EXPRESSION AND SPLICING DURING INFECTION WITH HIV
2 Scott Sherrill-Mix
3 A DISSERTATION
4 in
5 Genomics and Computational Biology
6 Presented to the Faculties of the University of Pennsylvania
7 in
8 Partial Fulfillment of the Requirements for the
9 Degree of Doctor of Philosophy
10 2015

11 Supervisor of Dissertation:

12 Frederic D. Bushman, Ph.D., Professor of Microbiology

13 Graduate Group Chairperson:

14 Li-San Wang, Ph.D., Associate Professor of Pathology and Laboratory Medicine

15 Dissertation Committee:

16 Nancy Zhang, Ph.D. Associate Professor of Statistics

17 Yoseph Barash, Ph.D., Assistant Professor of Genetics

18 Kristen Lynch, Ph.D., Professor of Biochemistry and Biophysics

19 Michael Malim, Ph.D., Professor of Infectious Diseases, King's College London

20 LATENCY, EXPRESSION AND SPLICING DURING INFECTION WITH HIV

21 © COPYRIGHT

22 2015

23 Scott A. Sherrill-Mix

24 This work is licensed under the

25 Creative Commons Attribution

26 NonCommercial-ShareAlike 3.0

27 License

28 To view a copy of this license, visit

29 <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Dedicated to William Maurer, Gayle Maurer & Michele Sherrill-Mix

ACKNOWLEDGEMENTS

32 I would like to thank

33 Rick Bushman

34 Christian Hoffmann wet lab

35 collaborators

36 Bushman lab Rithun Mukherjee, Karen Ocwieja, Nirav Malani, Troy Brady, Young Hwang,
37 Brendan Kelly, Kyle Bittinger, Rebecca Custers-Allen, Serena Dollive, Frances Male, Jacque
38 Young, Rohini Sinha, Sam Minot, Aubrey Bailey, Christopher Nobles, Stephanie Grunberg,
39 Vesa Turkki, Anatoly Dryga, Eric Sherman, Greg Peterfreund, Yinghua Wu, Alice Laughlin,
40 Sesh Sundararaman, Alexandra Bryson, Christel Chehoud, Erik Clarke, Arwa Abbas

41 My committee—Nancy Zhang, Yoseph Barash, Kristen Lynch and Michael Malim—have
42 provided guidance and encouragement. Many faculty of GCB mentoring and teaching.
43 Hannah Chervitz, Tiffany Barlow, Mali Skotheim, Caitlin Greig and Laurie Zimmerman
44 for managing everything and helping manage the layers of bureaucracy. Funding from the
45 HIV Immune Networks Team (HINT) consortium P01 AI090935 and NRSA computational
46 genomics training grant T32 HG000046.

47 Ram Myers and Mike James for great previous mentoring.

48 Xiaofen and Otto

49 ...

ABSTRACT

51 LATENCY, EXPRESSION AND SPLICING DURING INFECTION WITH HIV

52 Scott Sherrill-Mix

53 Frederic D. Bushman, Ph.D.

54 Over 35 million people are living with human immunodeficiency virus (HIV-1). The
55 mechanisms causing integrated provirus to become latent, the diversity of spliced viral
56 transcripts and the cellular response to infection are not fully characterized and hinder the
57 eradication of HIV-1. We applied high-throughput sequencing to investigate the effects of
58 host chromatin on proviral latency and variation of expression and splicing in both the host
59 and virus during infection.

60 To evaluate the link between host chromatin and proviral latency, we compared genomic and
61 epigenetic features to HIV-1 integration site data for latent and active provirus from five cell
62 culture models. Latency was associated with chromosomal position within individual models.
63 However, no shared mechanisms of latency were observed between cell culture models. These
64 differences suggest that cell culture models may not completely reflect latency in patients.

65 We carried out two studies to explore mRNA populations during HIV infection. Single-
66 molecule amplification and sequencing revealed that the clinical isolate HIV_{89.6} produces at
67 least 109 different spliced mRNAs. Viral message populations differed between cell types,
68 between human donors and longitudinally during infection. We then sequenced mRNA
69 from control and HIV_{89.6}-infected primary human T cells. Over 17 percent of cellular genes
70 showed altered activity associated with infection. These gene expression patterns differed
71 from HIV infection in cell lines but paralleled infections in primary cells. Infection with
72 HIV_{89.6} increased intron retention in cellular genes and abundance of RNA from human
73 endogenous retroviruses. We also quantified the frequency and location of chimeric HIV-host
74 RNAs. These two studies together provided a detailed accounting of both HIV_{89.6} and host

75 expression and alternative splicing.

76 A more cost-effective method of detecting viral load would aid patients with poor access to
77 healthcare. We developed improved methods for assaying HIV-1 RNA using loop-mediated
78 isothermal amplification based on primers targeting regions of the HIV-1 genome conserved
79 across subtypes. Combined with lab-on-a-chip technology, these techniques allow quantitative
80 measurements of viral load in a point-of-care device targeted to resource-limited settings.

81 This work disclosed novel HIV-host interactions and developed techniques and knowledge
82 that will aid in the study and management of HIV-1 infection.

TABLE OF CONTENTS

84	ABSTRACT	v
85	TABLE OF CONTENTS.....	vii
86	LIST OF TABLES	ix
87	LIST OF ILLUSTRATIONS	x
88	CHAPTER 1 : Introduction	1
89	1.1 The HIV epidemic	1
90	1.2 The HIV virus	3
91	1.3 Integration and latency	8
92	1.4 HIV splicing.....	8
93	1.5 Host cell interactions	12
94	1.6 HIV detection.....	12
95	1.7 Contribution summaries	14
96	CHAPTER 2 : HIV latency and integration site placement in five cell-based models	15
97	2.1 Abstract	15
98	2.2 Background	16
99	2.3 Methods	18
100	2.4 Results.....	22
101	2.5 Conclusions	33
102	2.6 Availability of supporting data	35
103	2.7 Author's contributions	35
104	2.8 Acknowledgements	36
105	CHAPTER 3 : Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing	37
107	3.1 Abstract	37
108	3.2 Introduction.....	38
109	3.3 Materials and methods.....	40
110	3.4 Results.....	47
111	3.5 Discussion	53
112	3.6 Acknowledgements	57
113	CHAPTER 4 : Gene activity in primary T cells infected with HIV _{89.6} : intron retention and induction of distinctive genomic repeats	58
114	4.1 Abstract	58
116	4.2 Background	59
117	4.3 Methods	60
118	4.4 Results.....	64

119	4.5 Discussion	82
120	4.6 Conclusions	86
121	4.7 Availability of supporting data	87
122	4.8 Acknowledgements	87
123	CHAPTER 5 : A reverse transcription loop-mediated isothermal amplification assay	
124	optimized to detect multiple HIV subtypes	88
125	5.1 Abstract	88
126	5.2 Introduction.....	89
127	5.3 Methods	90
128	5.4 Results.....	92
129	5.5 Testing different primer designs	96
130	5.6 Discussion	100
131	5.7 Acknowledgments.....	102
132	CHAPTER 6 : Conclusions and future directions.....	103
133	6.1 Latency and integration location	103
134	6.2 HIV-1 alternative splicing.....	104
135	6.3 Host expression during HIV infection	107
136	6.4 LAMP PCR and lab-on-a-chip	108
137	APPENDICES	112
138	A.1 Generalized linear models of changes in use of mutually exclusive HIV-1 splice acceptors	112
139	A.2 Reproducible report of HIV integration sites and latency analysis	118
140	BIBLIOGRAPHY	150

LIST OF TABLES

143	TABLE 2.1 : Integrations from <i>in vitro</i> models of latency	21
144	TABLE 2.2 : Genomic data available for comaprison to integration sites	23
145	TABLE 4.1 : Samples and RNA-Seq sequencing coverage	65
146	TABLE 4.2 : Data used for meta-analysis of expression changes in HIV	66

LIST OF ILLUSTRATIONS

148	FIGURE 1.1 : The HIV replication cycle.....	4
149	FIGURE 1.2 : The HIV-1 genome	4
150	FIGURE 2.1 : Correlations of genomic features and latency	24
151	FIGURE 2.2 : Lasso regressions predicting latency	25
152	FIGURE 2.3 : Cellular expression and latency	27
153	FIGURE 2.4 : Strand orientation and latency	28
154	FIGURE 2.5 : Genes and latency	29
155	FIGURE 2.6 : Alphoid repeats and latency	30
156	FIGURE 2.7 : Acetylation and latency	32
157	FIGURE 2.8 : Shared expression status between near neighbors.....	32
158	FIGURE 3.1 : Mapping the splice donors and acceptors of HIV _{89.6}	39
159	FIGURE 3.2 : Spliced transcripts produced from HIV _{89.6}	46
160	FIGURE 3.3 : Novel transcripts utilizing acceptor A8c	51
161	FIGURE 3.4 : Temporal, cell type and donor variability in accumulation of HIV-1 messages.....	54
163	FIGURE 4.1 : Comparisons among studies quantifying cellular gene expression after HIV infection	67
164	FIGURE 4.2 : Comparisons of the effect of HIV infection on gene expression to studies comparing subsets of immune cells	69
165	FIGURE 4.3 : Changes in the abundance of intronic regions with HIV infection ...	71
166	FIGURE 4.4 : Repeat categories enriched upon infection with HIV	73
167	FIGURE 4.5 : Characteristics of LTR12C sequences associated with induction upon infection with HIV _{89.6}	74
168	FIGURE 4.6 : Estimating relative abundance of HIV _{89.6} message size classes using RNA-Seq data	76
169	FIGURE 4.7 : Transcription and splicing of the HIV _{89.6} RNA	78
170	FIGURE 4.8 : Chimeric RNA sequences containing both human and HIV sequences	82
175	FIGURE 5.1 : Amplification results for all RT-LAMP primer sets tested	94
176	FIGURE 5.2 : Subtype-agnostic RT-LAMP primers design	95
177	FIGURE 5.3 : Performance of the AceIN-26 primer set with different starting RNA concentrations	98
178	FIGURE 5.4 : Replicate tests of the ACeIN-26 primer set over six HIV subtypes..	99
179	FIGURE 6.1 : Ebola RT-LAMP primers design	110

CHAPTER 1: Introduction

182 1.1 The HIV epidemic

183 In 1981, physicians began to notice a mysterious increase, often clustered in men who
184 had sex with men or intravenous drug users, in the occurrences of Kaposi's sarcoma and
185 pneumocystis pneumonia^{1–6}.

186 Kaposi's sarcoma was, until 1981, a rare cancer in the US found largely in elderly men with
187 Jewish or Mediterranean ancestry⁷. Kaposi's sarcoma had also been seen in immunocom-
188 promised individuals^{8–10} and there were suggestions that it was a virus-associated cancer¹¹
189 although the causative human herpesvirus 8 would not be discovered for another decade^{12,13}.

190 Pneumocystis pneumonia was known to be caused by infection of the alveoli with the yeast-
191 like fungus *Pneumocystis jirovecii*^{14,15}. Pneumocystis pneumonia was almost exclusively
192 seen only in patients with suppressed immune systems or immune disorders and rarely, if
193 ever, in immunocompetent individuals¹⁵.

194 The mechanism for this spike of opportunistic infections was clarified when researchers found
195 severe T cells depletion and decreases in cellular immunity in these patients^{4–6,16,17}. This
196 disease was eventually labeled acquired immunodeficiency syndrome (AIDS). However, the
197 underlying cause remained unclear.

198 Potential transmissions by transfusion^{18–20}, injection drug use^{4,17,21}, maternal transmission²²
199 and both homosexual^{16,23} and heterosexual^{17,24} contact pointed towards an infectious agent.
200 In 1983, a virus later named human immunodeficiency virus type 1 (HIV-1) was isolated
201 from patient samples^{25–28} and soon detected in most immunodeficient patients^{28–31}.

202 Reports of AIDS and associated opportunistic infections in sub-Saharan Africa soon revealed
203 widespread endemic infection^{32–35} and a great diversity of viruses^{36–41}. Retrospective studies
204 suggested that the virus had been present, at least sporadically, in Europe and the USA
205 for decades^{42,43} and circulating for even longer in Africa^{33,44–48}. Archived patient samples

206 containing HIV-1 genome fragments from as early as 1959 were found in what is now
207 Kinshasa, in the Democratic Republic of Congo⁴⁶. These samples showed extensive genome
208 diversification already present in the 1960s, suggesting that HIV-1 had been circulating in
209 humans for some time^{47,48}. Phylogenetic analyses adding in contemporary HIV-1 type M
210 sequences estimated a most recent common ancestor in the early 1900s^{48–53}.

211 A virus similar to HIV causing AIDS in monkeys was soon discovered in macaques^{54?} and
212 many other primates⁵⁵. HIV-1 appeared most similar to virus found in chimpanzees^{54,56}
213 and surveys of wild chimpanzees revealed a closely related simian immunodeficiency virus
214 infecting chimpanzees in southeast Cameroon^{57–59}.

215 The ancestor of HIV-1 was likely transmitted from a chimpanzee to a human, perhaps during
216 harvesting of chimpanzees for food^{60–65}, in the forests of southeastern Cameroon then virus
217 was transported down the Sangha River⁶⁶ to the city of Kinshasha, where the virus began its
218 global spread^{38,48,53,67}. A combination of social upheaval, increased mobility, urbanization
219 and mass vaccination campaigns with unsterilized needles appear to have provided fuel for
220 the growing epidemic^{53,68–70}. A virus appears to have been carried from Africa to Haiti in
221 the 1960s, perhaps by workers returning home from an exchange program^{35,67}, and into the
222 US in the 1970s⁷¹ before being detected in the US in 1981. In the past 34 years, HIV-1 has
223 spread to over 78 million people and caused over 35 million deaths⁷².

224 In the early days of the epidemic, there were no tests to detect the virus, and no treatments.
225 The presence of the virus was often revealed by the onset of AIDS. Opportunistic infections⁷³
226 and death usually followed soon after. The median survival time after diagnosis with AIDS
227 was about 1 year^{74,75}.

228 Isolation of the virus allowed the detection of the virus through assays of antibody response.
229 Testing revealed that, from infection, patients had a median survival time of around a
230 decade^{76–79}.

231 In 1987, the successful trial of the reverse transcriptase inhibitor azidothymidine provided

232 the first hope for treatment^{80–82} but it soon became apparent that the fast mutation rate
233 of HIV^{83–89} and strong selection by drug therapy could quickly create drug-resistant forms
234 of virus in patients receiving single drug therapy^{90–99}. Median survival time from AIDS
235 diagnosis rose to only about 2 years with therapy^{75,81,100,101}.

236 Additional antiretrovirals, again targeting reverse transcriptase, were developed¹⁰². Sequential
237 or alternating administration of different antiretroviral drugs did not greatly improve
238 prognosis^{103–107}. Simultaneous treatment with two reverse transcriptase inhibitor offered
239 modest benefits but viral escape was still common^{108–112}.

240 Development of drugs targeting other stages of the HIV replication cycle allowed synergistic
241 combinations of antiretroviral drugs^{113–118}. The difficulty for HIV of evolving multiple
242 drug resistant mutations^{119,120} meant that therapy using simultaneous combinations of
243 drugs finally began to offer patients more hope of long term survival^{121–125}. With early
244 triple therapy, median survival times rose to 20 years^{78,126}, rising to $\frac{2}{3}$ of uninfected life ex-
245 pectancy^{127,128} and finally approaching control populations with uninterrupted antiretroviral
246 treatment^{129–131}.

247 HIV rebounds quickly after cessation of therapy¹³² Groups and subtypes?

248 early establishment of latent infection¹³³ within 3 days¹³⁴ SAHA induces latent¹³⁵ new
249 drug for latency disulfram¹³⁶ modest induction of latent provirus¹³² characterizing latent
250 reservoir rare cells present in resting actice and macro¹³⁷

251 1.2 The HIV virus

252 HIV is an enveloped single strand positive-sense retrovirus, an RNA virus which uses reverse
253 transcription to create a DNA intermediate in host cells^{138,139}. Viral encoded integrase
254 protein discovered¹⁴⁰ and mapped to 3' end of pol^{141–143} through mutation and loss of
255 function. Mutations at ends also results in defective viruses¹⁴⁴.

256 The HIV genome encodes genes for at least two polyproteins and seven proteins:

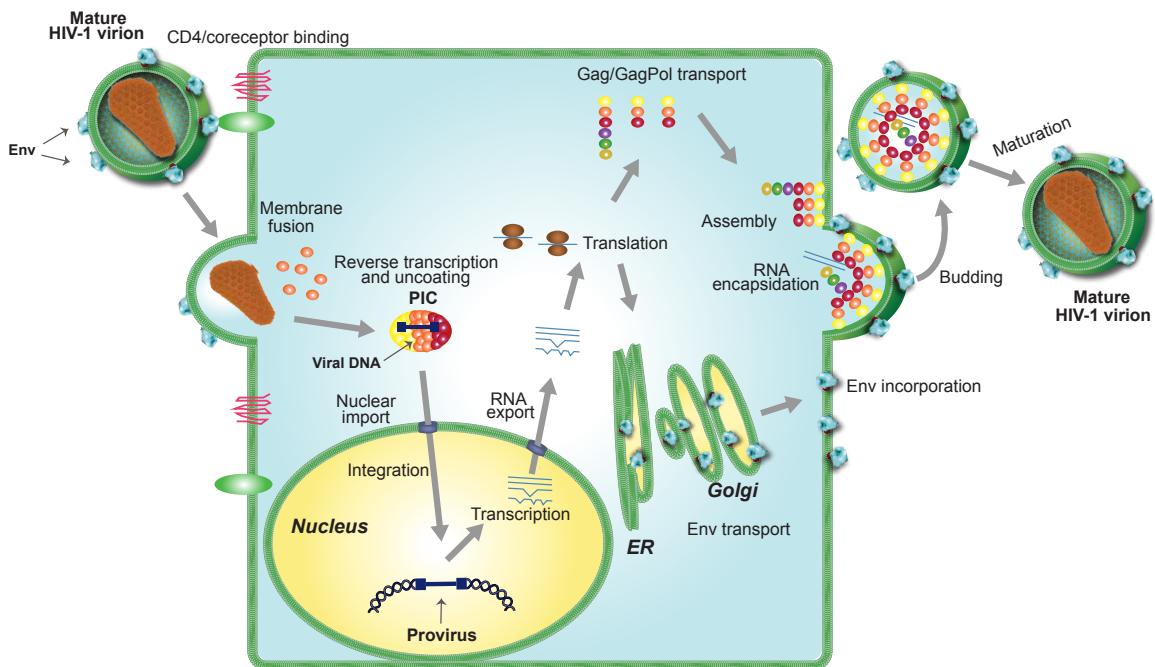


Figure 1.1: The HIV replication cycle

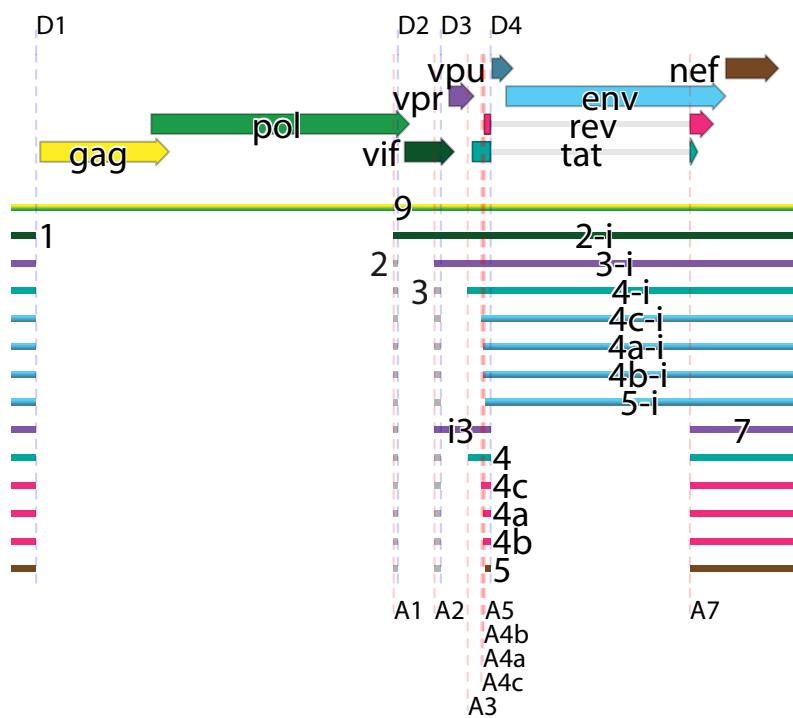


Figure 1.2: The HIV-1 genome. Arrows indicate open reading frames. Dashed lines show major splice acceptors (red) and donors (blue). Major spliceforms are shown as thin rectangles and colored according to their corresponding open reading frame.

257

258 **gag** Gag (group specific antigens) is a myristoylated membrane protein which is [[something]]
259 on the virion surface and cleaved by viral protease after virion budding to produce
260 matrix, capsid, nucleocapsid and p6 protein along with two small spacer peptides SP1
261 and SP2.

262 **MA** p17 MA (matrix) aids in transport of genome to the nucleus¹⁴⁵.

263 **CA** p24 CA (capsid) proteins assemble into to form a protective shell around RNA
264 genome of the virus. The viral capsid is composed of around 1500 copies of
265 CA arranged into hexameric rings interspersed with with 12 pentameric rings to
266 form fullerene cone^{146–149}. CA binds cellular CPSF6¹⁵⁰, cyclophilin A^{151,152} the
267 cyclophilin domain of RanBP2¹⁵³, perhaps to gain access to the nucleus^{153,154}
268 and to avoid premature uncoating and exposure of the viral genome to innate
269 immunity¹⁵⁵.

270 **NC** p7 NC (nucleocapsid)

271 **p6** p6 (protein 6 kda) is a small protein which appears to primarily recruit cellular
272 proteins to allow virion budding from the cell membrane^{156–158} and aid in the
273 packaging of Vpr in to particles¹⁵⁹.

274 **pol** Pol (polymerase) is cleaved by viral protease to produce reverse transcriptase, RNaseH,
275 integrase and HIV protease.

276 **RT** p51 RT (reverse transcriptase)

277 **RNase H** p15 RNase H

278 **IN** p31 IN (integrase) is a dimeric enzyme integrates the retroviral DNA into host
279 chromatin^{141,160–163}. Integrase removes two nucleotides from from 3' ends of the
280 viral DNA and inserts the pair of viral ends into host DNA¹⁶⁴.

281 **PR** PR (protease) is a dimeric aspartyl protease¹⁶⁵ that cleaves viral polyproteins
282 Gag and Pol^{166,167}.

283 **env** gp160 Env is a trimeric transmembrane protein that mediates entry through fusion
284 of viral and cellular membrane by binding its receptor CD4⁺^{168–172} and coreceptors
285 CXCR4¹⁷³ CCR3 and CCR5 coreceptor^{174,175}. gp160 is cleaved into two subunits gp41
286 and gp120¹⁷⁶ by cellular furin protease¹⁷⁷. The envelope protein is highly glycosylated
287 to form a ‘glycan shield’ against adaptive immune response¹⁷⁸. There are about 14
288 Env proteins per virion¹⁷⁹. Env is highly variable within and between patients^{180,181}
289 due to positive selection from host immune recognition^{182–184}.

290 **tat** Tat (transactivator) protein is a transactivator of expression from the HIV-1 long
291 terminal repeat^{185–187}. Virus does not replicate efficiently without this transactivation¹⁸⁸. and appears to regulate cellular expression such as downregulation of major
292 histocompatibility complex type I expression¹⁸⁹. Tat may suppress miRNA silencing
293 pathway^{190–192} but controversial¹⁹³.

295 **rev** Rev (regulator of expression of virion proteins) trans-activator protein shuttles between
296 the nucleus and cytoplasm¹⁹⁴ and causes the export of partially spliced and unspliced
297 viral transcripts^{195–199} from the nucleus through recognition of a structured RNA rev
298 response element^{200,201}.

299 **nef** Nef (negative factor) is a myristoylated membrane associated protein²⁰² that is involved
300 in multiple functions. Nef causes endocytosis of the HIV receptors CD4^{203–207} and
301 CCR5²⁰⁸ and viral surveillance major histocompatibility complex molecules^{209–212}.
302 Nef also induces T cell activation through interactions with signaling kinases and the
303 T cell receptor^{213–217}. In contrast, Nef in most other primate lentiviruses inhibits
304 activation and inflammation²¹⁸ perhaps indicating that the gain of *vpu* in HIV-1
305 and its simian relatives allowed the loss of the immune inhibition traits of *nef* and
306 contributes to the increased pathogenicity in these viruses^{219,220}.

307 ***vpr*** Vpr (viral protein R) is a 15 kDa protein^{221,222} with diverse functions that is incorporated
308 into virions^{223,224}. Vpr arrests the cell in the G2 phase of the cell cycle^{225–229}
309 and aids in transport of the viral genome to the nucleus¹⁴⁵. Vpr also appears to
310 trans-activate viral expression^{230,231} and induce apoptosis^{232,233} but these may be
311 linked to conditions caused by cell cycle arrest.

312 ***vif*** Vif (virion infectivity factor) counteracts the cellular restriction factor APOBEC3G²³⁴ by
313 excluding APOBEC3G from incorporation into the virion²³⁵ and causing APOBEC3G
314 to be ubiquitinated and degraded^{236–238}. APOBEC3G is otherwise packaged into
315 virions²³⁹ and deaminates the HIV genome during reverse transcription causing G-to-A
316 hypermutation^{239–242}.

317 ***vpu*** Vpu (viral protein U)^{243,244} is a small integral membrane protein which has two
318 known functions; degradation of CD4 and downregulation of BST-2 from the cell
319 membrane [[cites]]. Vpu causes cellular CD4 to be ubiquitinated and degraded^{245,246}
320 which prevents interactions between progeny virus Env and host cell CD4^{206,207,247,248}
321 and superinfection by other viruses²⁰⁴ while also releasing Env proteins from CD4
322 interactions in the endoplasmic reticulum^{249,250}. Vpu also counteracts the cellular
323 restriction factor BST-2, which would otherwise interfere with viral budding²⁵¹. Vpu
324 does not appear to be found in the virion²⁵².

325 In Chapter 3, we look for spliceforms potentially encoding more proteins and find two new
326 proteins.

327 HIV tropism infects macrophages²⁵³

328 HIV diversification within a single patient in env loop¹⁸⁰ positive selection^{182,184} positive
329 selection at same sites over time correlate with slower progression^{183,184} 1% distance increase
330 diversity in env per year (decreases later in infection)¹⁸¹

331 half life of 2 days and 10^9 CD4 T cells per day^{254–256} macrophages? half life 1–4 weeks^{256–258}

332 resting CD4 T cells^{137,259,260}
333 Retrovirus package two copies of RNA in each virion^{261–263}. Interstrand transfer during the
334 reverse transcription step allows recombination^{264–266} [[strand transfer rt refs 3–5 in²⁶⁴]]

335 1.3 Integration and latency

336 Integration into a host chromosome is an integral step in the retrovirus replication cycle.
337 cells can produce virus without cell death in vitro²⁶⁷ first latent HIV²⁶⁸ High throughput
338 integration site sequencing^{269,270}

339 1.4 HIV splicing

340 RNA splicing was first observed in adenovirus^{271,272}. Improved understanding of HIV and
341 other viruses offers medical benefits. Although HAART treatments have greatly improved
342 HIV prognosis, long-term survival of HIV patients remains reduced by at least a decade
343 compared to the general population²⁷³. In addition HAART does not provide a permanent
344 cure²⁷⁴ and the long-term costs of treatment runs into hundreds of thousands of dollars over
345 the lifetime of a patient^{275,276}. Induction or alteration of splicing has been suggested as a
346 potential treatment^{277,278} and differential splicing appears to be one factor limiting cross
347 species infection and the development of animal models of HIV²⁷⁹.

348 Rev balanced with weak 3' splice sites²⁸⁰.

349 Splicing changes due to splice factor changes in macrophages²⁸¹
350 Driven by a strong selective pressure for genome compactness^{282–284}, HIV and other
351 lentiviruses subvert host cell alternative splicing pathways to allow tight packing of their
352 genetic information. Through weak splice sites and overlapping reading frames (Figure 1.2),
353 the virus manages to produce precise quantities of at least nine proteins and polyproteins
354 from its single transcription start site and less than 10 kb genome²⁸⁵.

355 As such an integral part of the virus life cycle^{286,287}, alteration of splicing poses a tempting

356 therapeutic target. Inhibition of cellular splicing factors reduces viral reproduction in many
357 genome-wide siRNA screens^{288–290} and several members of the spliceosome interact with
358 viral proteins in affinity pulldowns²⁹¹. Open reading frames in uncharacterized transcripts
359 appear to produce epitopes useful for vaccine development²⁹². Potential treatments altering
360 viral splicing through small molecule inhibitors^{277,293} and gene therapy^{278,294} have restricted
361 viral replication *in vitro*. However without methods to quantify viral splicing or a thorough
362 quantification of splicing under varying conditions, the development of such treatments
363 remains limited.

364 Viral proteins also interact with components of the cellular splicing complex^{291,295,296}. These
365 interactions have been reported to change splicing in viral^{296–298} and cellular transcripts^{299,300}
366 and raise the possibility that the virus has evolved to alter host splicing. A genome-wide
367 study of changes in cellular splicing during HIV infection would greatly clarify this hypothesis
368 but no such study has been performed.

369 Alternative splicing, the differential inclusion of exons and removal of introns from primary
370 mRNA transcripts, allows rapid evolution of protein segments^{301–303} and drastic increases
371 in the number of proteins generated by a single DNA sequence³⁰⁴. Many viruses subvert
372 the splicing machinery of their eukaryotic hosts to modify their viral mRNA³⁰⁵.

373 In particular, it has previously been reported that HIV utilizes alternative splicing to
374 generate more than 40 mRNA transcripts encoding at least 9 proteins and polyproteins from
375 a genome smaller than 10kb³⁰⁶. A specific progression of viral transcripts appear in the
376 cytoplasm of the host cell as infection progresses allowing a shift from regulatory protein
377 production in early infection into virion production in late infection^{286,287,307}. Because
378 HIV has only a single transcription start site, these transcriptional changes are driven by
379 alternative splicing²⁸⁵.

380 Although it plays such an essential role for the virus, only a single detailed census of viral
381 splicing has been reported³⁰⁶. Due to limitations in technology, this study was limited

382 to only the most abundant transcripts in lab-adapted HIV strains in cell culture³⁰⁶. Yet
383 rare transcripts may play an important role in immune response²⁹² and encode unknown
384 proteins³⁰⁸; lab adapted HIV can differ markedly from viruses actually found in patients³⁰⁹;
385 cell cultures often do not reflect *in vivo* conditions³¹⁰; and splicing can vary between
386 humans^{311,312} and cell types^{313,314}. Without a fuller characterization of transcripts under
387 these relevant conditions, many aspects of viral splicing will remain poorly understood.

388 Alternative splicing may also play an unappreciated role in HIV-host interactions. Viral
389 proteins interact with the splicing complex^{291,295,296} and alter splicing of some cellular
390 transcripts^{299,300}. Yet, although infection has been shown to cause genome-wide changes in
391 the expression of cellular genes^{315–319}, no genome-wide study of cellular alternative splicing
392 during HIV infection has ever been reported. Such a genome-wide study of splicing changes
393 might reveal a distortion of diverse cellular splicing which is adaptively advantageous to the
394 virus.

395 Current sequencing advancements allow a much broader and deeper investigation of viral
396 splicing. Targeted amplification with RainDance droplet PCR offer the potential to reduce
397 size bias inherent in bulk PCR³²⁰. RNA-seq with Illumina sequencing allows extremely deep
398 sequencing of cellular and viral transcripts with billions of bases of short read sequence^{321,322}.
399 Single molecule sequencing with Pacific Biosciences provides reads approaching 20,000
400 bases^{323,324} that could characterize entire viral transcripts in one continuous read. By
401 combining these technologies, viral and cellular transcripts could be interrogated to an
402 unprecedented level.

403 A better understanding of viral splicing and viral effects on host splicing may bring therapeutic
404 benefits. siRNA inhibition of splicing factors reduces HIV replication in many genome-
405 wide screens^{288–290}. Alteration of viral splicing through small molecule inhibitor of SR
406 protein kinases²⁷⁷ and Splicing Factor 2²⁹³, shRNA against spliceosomal U7 snRNP²⁹⁴
407 and expression of modified spliceosomal U1 snRNP²⁷⁸ show treatment potential *in vitro*.
408 In addition, rare uncharacterized HIV transcripts and their encoded proteins appear to

409 produce potent immune response in HIV patients²⁹² thus offering potential targets for
410 vaccine development. Yet without methods to characterize viral RNA and measure the
411 effects of treatments on viral splicing, further development is inhibited.

412 Inclusion and exclusion of a particular stretch of RNA into an mRNA is determined by a bal-
413 ance of RNA secondary structure^{325–327}, chromatin structure³²⁸, nucleosome positioning³²⁹,
414 histone marks³³⁰, previous splicings³³¹, order of intron removal^{332,333} and enhancers³³⁴
415 and suppressors³³⁵ that bind specific motifs³³⁶. Together these factors create a precise
416 controllable splicing code^{314,337,338}.

417 In HIV, splicing occurs between at least four splice donors and eight splice acceptors²⁸⁵.
418 Two splice donors, D1 and D4, are relatively strong while the remaining donors and all
419 acceptors are fairly weak³³⁹. Several exonic splicing silencers^{340,341} and exon splicing
420 enhancers^{342,343} and a single intronic splicing silencer³⁴⁴ in the viral genome interact with
421 many human splicing factors, including hnRNPs A1^{341,344} H, F, 2H9, and A2³²⁶ and SR
422 proteins SRp40^{342,345}, SRp75³⁴⁵, ASF/SF2³⁴² and SC35³²⁶, to alter viral splicing^{285,346}.

423 Several viral proteins affect mRNA abundances. Rev causes export of unspliced viral mRNA
424 that would otherwise be trapped in the nucleus³⁴⁷ to be exported^{305,348} and may also interact
425 with splicing factors to alter viral splicing²⁹⁵. The HIV protein Tat is best known for its
426 transactivation of viral transcription^{185,349} and triggering apoptosis in uninfected cells^{350,351}
427 but Tat also appears to independently affect alternative splicing of viral transcripts^{296–298,352}.

428 Viral protein Vpr is known to cause cell cycle arrest²²⁸ and mediate nuclear import of the viral
429 preintegration complex³⁵³. Vpr also appears to alter alternative splicing of some cellular
430 transcripts^{299,300} and interact with the SMN complex²⁹¹, which assembles spliceosomal
431 snRNP³⁵⁴. Although all three of these proteins modify viral splicing, whether they also
432 cause widespread alterations in cellular splicing is unknown.

433 Despite the critical role alternative splicing plays in viral replication, no genome-wide studies
434 of lentiviral effects on cellular splicing or detailed censuses of viral alternative splicing in

435 biologically relevant settings have been published.

436 RNA-seq offers a much broader view of alternative splicing than previously possible^{355,356}
437 but Illumina sequencing has not yet been applied to the study of differential splicing in
438 host RNA of HIV-infected cells. There have been many studies of cellular expression using
439 microarrays^{315–318,352} and Sage^{357,358} but only a single study using Illumina RNA-seq and
440 alternative splicing changes were not reported³¹⁹. Thus a potentially significant aspect of
441 HIV-host interactions remains unknown.

442 The most extensive survey of HIV transcripts to date was published in 1993³⁰⁶. Technology
443 at the time necessitated the use of Northern blots and RNA protection assays³⁰⁶ which can
444 not distinguish multiple similarly sized transcripts or detect rare transcripts. This study
445 also focused on a single lab adapted HIV_{NL4-3} strain in HeLa cell culture.

446 Many previous studies of viral splicing have used lab-adapted strains of HIV which often differ
447 from patient isolates³⁰⁹ in cell cultures which often differ from primary cells³¹⁰. Selection
448 under cell culture conditions may quickly alter splicing patterns to down regulate proteins
449 unneeded *in vitro*. Characterization of alternative splicing in biologically relevant cell types
450 infected with clinical isolates of HIV are sorely needed.

451 1.5 Host cell interactions

452 1.6 HIV detection

453 Immunoassays provide cheap immediate testing of HIV infection in patients. These tests are
454 based on the enzyme-linked immunosorbent assay (ELISA), using an enzyme linked to an
455 antibody to produce a detectable signal in the presence of antigen^{359–361}.

456 The isolation of HIV^{25–28} allowed the production of large quantities of virions. These virions
457 were bound to a substrate, sera from patients added and any patients antibodies sensitive
458 to HIV allowed to bind. Any unbound antibodies were washed away. Then a peroxidase
459 enzyme-labeled antibody targeted to human antibody bound was added, allowed to bind

460 and the unbound antibodies again washed away. Any HIV-targeted patient antibodies
461 would bind the antigen and be bound in turn by the peroxidase-labeled antibody and the
462 peroxidase would then change the color of media^{30,31,362}. These tests had a large false
463 positive rate and the standard procedure was to perform multiple ELISA tests follow by a
464 Western blot test^{363,364} but false positives were still prevalent³⁶⁵. More conservative criteria
465 and cleaner lab procedures reduced false positives³⁶⁶. These assays have been developed to
466 fourth generation³⁶⁷ with more sensitive and specific detection of patient antibodies and
467 earlier detection using antibodies against the HIV capsid protein^{368,369}.

468 Slightly less specific but rapid immunoassays providing results in 30 minutes have been
469 developed to allow point-of-care testing with many fewer patients lost to follow up prior
470 to delivery of results^{370–372}. Rapid tests detecting HIV in oral fluids have been developed
471 and obviate the need for a blood draw^{373–375}. These rapid care tests allow self testing at
472 home^{376,377}.

473 Reverse transcription and PCR amplification offers another alternative^{378,379} but is not
474 currently cost effective for primary patient screening³⁸⁰.

475 Tests allowing point-of-care qualitative HIV detection are now widespread but point-of-care
476 assays for viral load in a patient exist. In addition, existing laboratory-based tests are rela-
477 tively expensive and require specialized equipment making access difficult in resource-limited
478 settings^{381,382}. Without viral load measures, CD4⁺ T cell counts or clinical presentation
479 are used to infer the emergence of drug. These criteria are not specific or sensitive enough
480 without viral load measures so many patients are unnecessarily switched to second line
481 therapy^{383,384} or switched too late leading to accumulations of drug resistant mutations³⁸⁵.
482 In Chapter 5, we design loop-mediated isothermal amplification methods that can be used
483 with microfluidics to create a point-of-care assay of infection and viral load in resource-limited
484 settings.

485 1.7 Contribution summaries

486 Much of this work was performed as part of a large collaboration. It would not tell a
487 complete story in isolation. Therefore, I have preserved the chapters in published form and
488 detailed my contribution to each project at the start of the chapter.

489

CHAPTER 2: HIV latency and integration site placement in five cell-based models

490

This chapter was originally published as:

S Sherrill-Mix, MK Lewinski, M Famiglietti, A Bosque,
N Malani, KE Ocwieja, CC Berry, D Looney, L Shan
et al. 2013. HIV latency and integration site place-
ment in five cell-based models. *Retrovirology*, 10:90. doi:
10.1186/1742-4690-10-90

491

I led the computational analysis, with assistance from CC Berry and N Malani. MK Lewinski, D Looney and J Guatelli analyzed integration sites using IonTorrent sequencing. M Famiglietti, A Bosque and V Planelles prepared DNA from latent and activated T cells using the Central Memory CD4 + model. L Shan, RF Siliciano, MJ Pace, LM Agosto, KE Ocwieja and U O'Doherty contributed data and suggestions. FD Bushman and I planned the overall study. I produced the figures. FD Bushman and I wrote the paper.

Additional files are available at [http://www.retrovirology.com/
content/10/1/90/additional](http://www.retrovirology.com/content/10/1/90/additional)

492

2.1 Abstract

493

Background: HIV infection can be treated effectively with antiretroviral agents, but the persistence of a latent reservoir of integrated proviruses prevents eradication of HIV from infected individuals. The chromosomal environment of integrated proviruses has been proposed to influence HIV latency, but the determinants of transcriptional repression have not been fully clarified, and it is unclear whether the same molecular mechanisms drive latency in different cell culture models.

494

495

496

497

498

499

500

Results: Here we compare data from five different *in vitro* models of latency based on primary human T cells or a T cell line. Cells were infected *in vitro* and separated into

501 fractions containing proviruses that were either expressed or silent/inducible, and integration
502 site populations sequenced from each. We compared the locations of 6,252 expressed
503 proviruses to those of 6,184 silent/inducible proviruses with respect to 140 forms of genomic
504 annotation, many analyzed over chromosomal intervals of multiple lengths. A regularized
505 logistic regression model linking proviral expression status to genomic features revealed no
506 predictors of latency that performed better than chance, though several genomic features
507 were significantly associated with proviral expression in individual models. Proviruses in the
508 same chromosomal region did tend to share the same expressed or silent/inducible status if
509 they were from the same cell culture model, but not if they were from different models.

510 Conclusions: The silent/inducible phenotype appears to be associated with chromosomal
511 position, but the molecular basis is not fully clarified and may differ among *in vitro* models
512 of latency.

513 2.2 Background

514 Highly active antiretroviral therapy (HAART) can suppress HIV-1 replication in infected pa-
515 tients, but the ability of HIV to persist as an inducible reservoir of latent proviruses^{137,387,388}
516 obstructs eradication of the virus and functional cure²⁷⁴. These latent proviruses are long
517 lived^{389,390} and relatively invisible to the immune system^{137,259}. The potential for even a
518 single virus to restart infection despite successful antiviral therapy means that it may be
519 necessary to eliminate all latent proviruses to eradicate HIV from an infected person.

520 After integration, a positive feedback loop of Tat transactivation appears to partition
521 proviral gene activity into either of two stable states^{391–393}—abundant Tat driving high
522 proviral expression or little Tat leading to quiescent latency. Similar to the positional effect
523 variegation observed in fruit fly chromosomal rearrangements^{394,395}, studies on cell clones
524 with single integrations show that differing integration sites can have large differences in
525 proviral expression^{396–398}. These data suggest that integration site location, along with the
526 cellular environment^{398–401}, influences the balance between latency and proviral expression.

527 Associations between latency and genomic features have also been reported in collections of
528 integration sites from cell culture models although the consistency of these effects across
529 model systems and their relationships to latency in patients remains uncertain. Lewinski
530 et al.⁴⁰² reported that proviruses integrated in gene deserts, alphoid repeats and highly
531 expressed genes are more likely to have low expression. Shan et al.⁴⁰³ reported an association
532 between latency and integration in the same transcriptional orientation as host genes. Pace
533 et al.⁴⁰⁴ found that silent and expressed provirus integration sites differed in the abundance
534 and expression levels of nearby genes, GC content, CpG islands and alphoid repeats. In
535 model systems with defined integration sites, Lenasi et al.⁴⁰⁵ reported decreased and Han
536 et al.⁴⁰⁶ reported increased viral transcription when the provirus is downstream of a highly
537 expressed host gene.

538 Cell-based models of latency are important for many aspects of HIV research, including
539 screening small molecules that can reverse latency and potentially allow eradication^{407,408}.
540 Location-driven differences in expression are preserved even after demethylation and histone
541 deacetylase treatment³⁹⁶, which suggests that integration location has the potential to
542 confound “shock and kill” anti-latency treatments^{409,410}. A greater understanding of the
543 effects of integration site location on latency could thus affect antiretroviral development.

544 To search for features of integration site associated with latency, we generated a set of
545 inducible and expressed integration sites using a primary central memory CD4⁺ T cell model
546 of latency^{411,412}, collected four previously reported integration site datasets and modeled
547 the effects of genomic features near the integration site on the expression status of these
548 proviruses. Although some genomic features associated with latency in individual models,
549 no feature was consistently associated with proviral expression across all five cell culture
550 models. However, closely neighboring proviruses within the same cellular model shared the
551 same latency status much more often than expected by chance suggesting that chromosomal
552 position of integration affects latency but that the mechanism remains unclear or differs
553 between cell culture models. Thus these data help inform the design of experiments in HIV

554 eradication research.

555 **2.3 Methods**

556 **2.3.1 Integration sites**

557 Naive CD4⁺ T cells were purified by negative selection from peripheral blood mononuclear
558 cells. The cells were activated with anti-CD3 and anti-CD28 (+TGF-beta, anti-IL-12, and
559 anti-IL-4) to generate “non-polarized” cells (the in vitro equivalent of central memory T
560 cells). Five days after isolation, cells were infected with an NL4-3-based virus with GFP in
561 place of Nef and the LAI envelope (X4) provided in trans at a concentration of 500 ng of
562 p24 as measured by ELISA per million cells. Based on previous experience with this model,
563 this amount of p24 should produce an MOI of approximately 0.15. Cells were cultured
564 in the presence of IL-2. Two days post-infection, cells were sorted for GFP+; this active
565 population expresses GFP even when treated with flavopiridol, although for this study they
566 were not treated. The inducible population was the set of GFP negative cells from the initial
567 sort that, 9 days post-infection, were activated with anti-CD3 and anti-CD28 and sorted for
568 GFP production.

569 Genomic DNA from the inducible and expressed populations was digested with MseI, ligated
570 to an adapter, and amplified by ligation-mediated PCR essentially as in Wu et al.⁴¹³ and
571 Mitchell et al.⁴¹⁴ except that the nested PCR primers included sequence for the Ion Torrent
572 P1 adapter and adapter A sequence with a 5 base barcode sequence specific to the inducible
573 or expressed conditions. Amplicons were sequenced using an Ion Torrent Personal Genome
574 Machine (PGM) according to manufacturer’s instructions using an Ion 316 chip and the Ion
575 PGM 200 Sequencing kit (Life Technologies). The sequence reads were sorted into samples
576 by barcode. All reads were required to match the expected 5' sequence with a Levenshtein
577 edit distance less than 3 from the expected barcode, 5' primer and HIV long terminal repeat
578 (LTR). The 5' primer and HIV sequence, along with the 3' primer if present, were trimmed
579 from the read. Sequences with less than 24 bases remaining or containing any eight base

580 window with an average quality less than 15 were discarded. Duplicate reads and reads
581 forming an exact substring of a longer read were removed.

582 **2.3.2 Analysis**

583 All statistical analysis was performed in R 2.15.2⁴¹⁵. The analyses are described in a
584 reproducible report (Appendix A.2). The annotated integration site data necessary to
585 perform the analyses and the compilable code to generate this reproducible report are
586 provided as supplemental information³⁸⁶. The new Central Memory CD4⁺ data set was
587 analyzed as in Berry et al.⁴¹⁶. The integration patterns appeared similar to previously
588 reported HIV integration site datasets⁴¹⁷.

589 **2.3.3 Previously published data**

590 We collected integration sites from three previously reported studies (Table 2.1), for a total
591 of four expressed versus silent/inducible pairs of samples. These studies used primary CD4⁺
592 T cells or Jurkat cells infected with HIV or HIV-derived constructs as cell culture models of
593 latency. Flow cytometry allowed cells expressing viral encoded proteins to be sorted from
594 non-expressing cells. In two of the studies, these non-expressing populations were stimulated
595 to ensure that the provirus could be aroused from latency. Specific differences in protocol
596 between the study sets are summarized below.

597 **Jurkat** Lewinski et al.⁴⁰² infected Jurkat cells with a VSV-G pseudotyped, GFP-expressing
598 pEV731 HIV construct (LTR-Tat-IRES-GFP)³⁹⁶ at an MOI of 0.1. The cells were
599 sorted into GFP+ and GFP- two to four days after infection. GFP+ cells were sorted
600 again two weeks after infection and cells that were again GFP+ were collected for
601 integration site sequencing. GFP- cells were sorted for GFP negativity twice more
602 then stimulated with TNF α . Cells that were GFP+ after stimulation were collected
603 for integration site sequencing. DNA was digested with MseI or a combination of NheI,
604 SpeI and XbaI, ligated to adapters for nested PCR, amplified and sequenced by Sanger
605 capillary electrophoresis.

606 **Bcl-2 transduced CD4⁺** Shan et al.⁴⁰³ transduced CD4⁺ T cells with Bcl-2, costimulated
607 with bound anti-CD3 and soluble anti-CD28 antibodies, interleukin-2 and T cell growth
608 factor and then infected with X4-pseudotyped GFP-expressing NL4-3- δ 6-drEGFP
609 construct⁴¹⁸ at an MOI of less than 0.1. DNA was extracted, digested with PstI and
610 circularized⁴¹⁹. HIV-human junctions were amplified by reverse PCR and sequenced
611 using Sanger capillary electrophoresis.

612 **Active CD4⁺ & Resting CD4⁺** Pace et al.⁴⁰⁴ spinoculated CD4⁺ T cells with HIV
613 NL4-3 at an MOI of 0.1. After 96 hours, the cells were stained for intracellular Gag
614 CD25, CD69 and HLA-DR and sorted into four subpopulations based on activation
615 state and Gag expression; activated Gag-, activated Gag+, resting Gag- and resting
616 Gag+. The ability of the viruses to reactivate was not tested although previous studies
617 have shown that the majority are likely inducible⁴²⁰. Genomic DNA was extracted and
618 digested with restriction enzymes MseI and Tsp509 and ligated to adapters. Proviral
619 LTR-host genome junctions were sequenced by 454 pyrosequencing after nested PCR.

620 All datasets were processed using the hiReadsProcessor R package⁴²¹. Adaptor trimmed
621 reads were aligned to UCSC freeze hg19 using BLAT⁴²². Genomic alignments were scored
622 and required to start within the first three bases of a read with 98% identity. Alignments for
623 a given read with a BLAT score less than the maximum score for that read were discarded.
624 Reads giving rise to multiple best scoring genomic alignments were excluded, while reads
625 with a single best hit were dereplicated and converged if within 5bp of each other. The
626 Bcl-2 transduced CD4⁺ sample was sequenced from U3 in the 5' HIV LTR while the other
627 samples were sequenced from U5 in the 3' LTR. To account for the 5 base duplication of
628 host DNA caused by HIV integration, the chromosomal coordinates of the Bcl-2 transduced
629 CD4⁺ sample were adjusted by ± 4 bases.

630 To allow for alignment difficulties in the analysis of genomic repeats, reads with multiple
631 best scoring alignments, along with the single best hit reads used above, were included in
632 the repeat analyses. If any best scoring alignment for a read fell within a repeat, then that

Title	Cell type	Virus	Time of harvest after infection	Sequencing	Generation of expressed vs. silent/inducible	Citation	Silent/inducible unique sites	Expressed unique sites
Jurkat	Jurkat cells	HIV vector pEV731 (LTR-Tat-IRES-GFP)	2 weeks	Sanger	TNF α , GFP expression	Lewinski et al. ⁴⁰²	463 inducible	643
Bcl-2 transduced CD4 $^{+}$	Primary CD4 $^{+}$ T cells (Bcl-2 transduced)	HIV NL4-3- δ 6-drEGFP (inactivated gag, vif, vpr, vpu, nef and env replaced by GFP)	3 days + 3-4 weeks + 3 days	Sanger	anti-CD3, anti-CD28 antibodies, GFP expression	Shan et al. ⁴⁰³	446 inducible	273
Active CD4 $^{+}$	Primary active CD4 $^{+}$ T cells	HIV NL4-3	3 days	454	high vs. low Gag	Pace et al. ⁴⁰⁴	1604 silent	1274
Resting CD4 $^{+}$	Primary resting CD4 $^{+}$ T cells	HIV NL4-3	3 days	454	high vs. low Gag	Pace et al. ⁴⁰⁴	1942 silent	784
Central Memory CD4 $^{+}$	Primary central memory CD4 $^{+}$ T cells	HIV NL4-3 Δ Nef GFP	2 days/9 days	Ion-Torrent	anti-CD3, anti-CD28 antibodies, GFP expression	This paper	1729 inducible	3278

Table 2.1: HIV-1 integration datasets from *in vitro* models of latency where the proviruses were determined to be silent/inducible or expressed

633 read was considered to map to that repeat.

634 2.3.4 Genomic features

635 A total of 140 whole genome features for CD4 $^{+}$ T-cells were gathered from data sources
 636 indicated in Table 2.2. For features encoded as peaks or hotspots, the log of the distance of
 637 each integration site to the nearest border was used for modeling. Integration sites from
 638 HIV 89.6 infection in primary CD4 $^{+}$ T cells⁴²³ were used to count nearby integrations and
 639 determine a \pm 20bp position weight matrix for integration targets. Illumina RNA-Seq from
 640 active CD4 $^{+}$ cells (Chapter 4) was used to estimate raw cellular expression and fragments
 641 per kilobase of transcript per million mapped reads for genes as calculated by Cufflinks³⁵⁵.
 642 For sequence-based data like RNA-Seq and ChIP-Seq, the number of reads aligned within
 643 a \pm 50, 500, 5,000 50,000 and 500,000 bp windows of each integration site were counted
 644 and log transformed. In addition, chromatin state classifications derived from a hidden

645 Markov model based on histone marks and a few binding factors⁴²⁴ were included as binary
646 variables. All data from previous genomic freezes were converted to hg19 using liftover⁴²⁵.

647 2.4 Results

648 The combination of integration site data newly reported here (set named “Central Memory
649 CD4⁺”) with previously published data (sets named “Jurkat”, “Bcl-2 transduced CD4⁺”,
650 “Active CD4⁺”, and “Resting CD4⁺”) provides a collection of 12,436 integration sites (Table
651 2.1) where the expression status of the provirus—silent/inducible or expressed—is known.
652 In three of the datasets, Jurkat, Central Memory CD4⁺ and Bcl-2 transduced CD4⁺, the
653 proviruses were sorted based on inducibility. In the Resting CD4⁺ and Active CD4⁺ datasets,
654 cells were sorted only based on proviral expression. Previous studies have shown that most
655 silent proviruses in this model system are inducible⁴²⁰.

656 2.4.1 Global model

657 If a genomic feature and latency are monotonically related then we should be able to detect
658 this relationship using Spearman rank correlation. In addition if a feature has a consistent
659 effect across models we should see a consistent pattern in the direction of correlation. A
660 simple first look for correlation between genomic features (Table 2.2) and latency status
661 yielded inconsistent results among the five samples with no variables having a significant
662 Spearman rank correlation across all, or even four out of five, of the samples (Figure 2.1).
663 This suggests that there is not a consistent simple monotonic relationship between the
664 genomic variable and latency, or that any such correlations are modest and not detectable
665 across all studies given the available statistical power. We return to some of the stronger
666 trends below.

667 To investigate whether a combination of variables may affect latency, we fit a lasso-regularized
668 logistic regression, as implemented in the R package glmnet⁴³⁴, to predict latency using
669 the genomic variables. The relationship between silent/inducible status and each genomic
670 variable was allowed to vary between models by including the interaction of genomic features

Group	Type	Source	Number	Types
T cell expression	RNA-Seq	Chapter 4	1	RNA
Jurkat expression	RNA-Seq	Encode ⁴²⁶	1	wgEncodeHudsonalphaRnaSeq
Integration sites	Locations	Berry et al. ⁴²³	1	sites
DNase sensitivity	DNA-Seq/peaks	Encode ⁴²⁶	1	wgEncodeOpenChromDnase
Methylation	DNA-Seq	427	1	Methyl
CpG	Locations	UCSC ⁴²⁸	1	cpgIslandExt
Sequence-based	Continuous	—	4	% GC, HIV PWM score, distance to centrosome, chromosomal position
Repeats	Locations	UCSC ⁴²⁸	16	DNA, LINE, Low_complexity, LTR, Other, RC, RNA, rRNA, Satellite, scRNA, Simple_repeat, SINE, snRNA, srpRNA, tRNA, alphoid
Histone features	ChIP-Seq/Peaks	Wang et al. ⁴²⁹	18	H2AK5ac, H2AK9ac, H2BK120ac, H2BK12ac, H2BK20ac, H2BK5ac, H3K14ac, H3K18ac, H3K23ac, H3K27ac, H3K36ac, H3K4ac, H3K9ac, H4K12ac, H4K16ac, H4K5ac, H4K8ac, H4K91ac
Histone features	ChIP-Seq/Peaks	Barski et al. ⁴³⁰	23	CTCF, H2AZ, H2BK5me1, H3K27me1, H3K27me2, H3K27me3, H3K36me1, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me1, H3K79me2, H3K79me3, H3K9me1, H3K9me2, H3K9me3, H3R2me1, H3R2me2, H4K20me1, H4K20me3, H4R3me2, PolII
Chromatin state	Binary	Ernst and Kellis ⁴²⁴	51	state ₁ ,state ₂ ,...,state ₅₁
HATs and HDACs	ChIP-Seq	Wang et al. ⁴³¹	11	Resting-HDAC1, Resting-HDAC2, Resting-HDAC3, Resting-HDAC6, Resting-p300, Resting-CBP, Resting-MOF, Resting-PCAF, Resting-Tip60, Active-HDAC6, Active-Tip60
Nucleosome	ChIP-Seq	Schones et al. ⁴³²	2	Resting-Nucleosomes, Active Nucleosomes
UCSC genes	Locations	Hsu et al. ⁴³³	4	in gene, in gene (same strand), gene count, distance to nearest gene, in exon, in intron

Table 2.2: Genomic data available for comparison to HIV integration sites

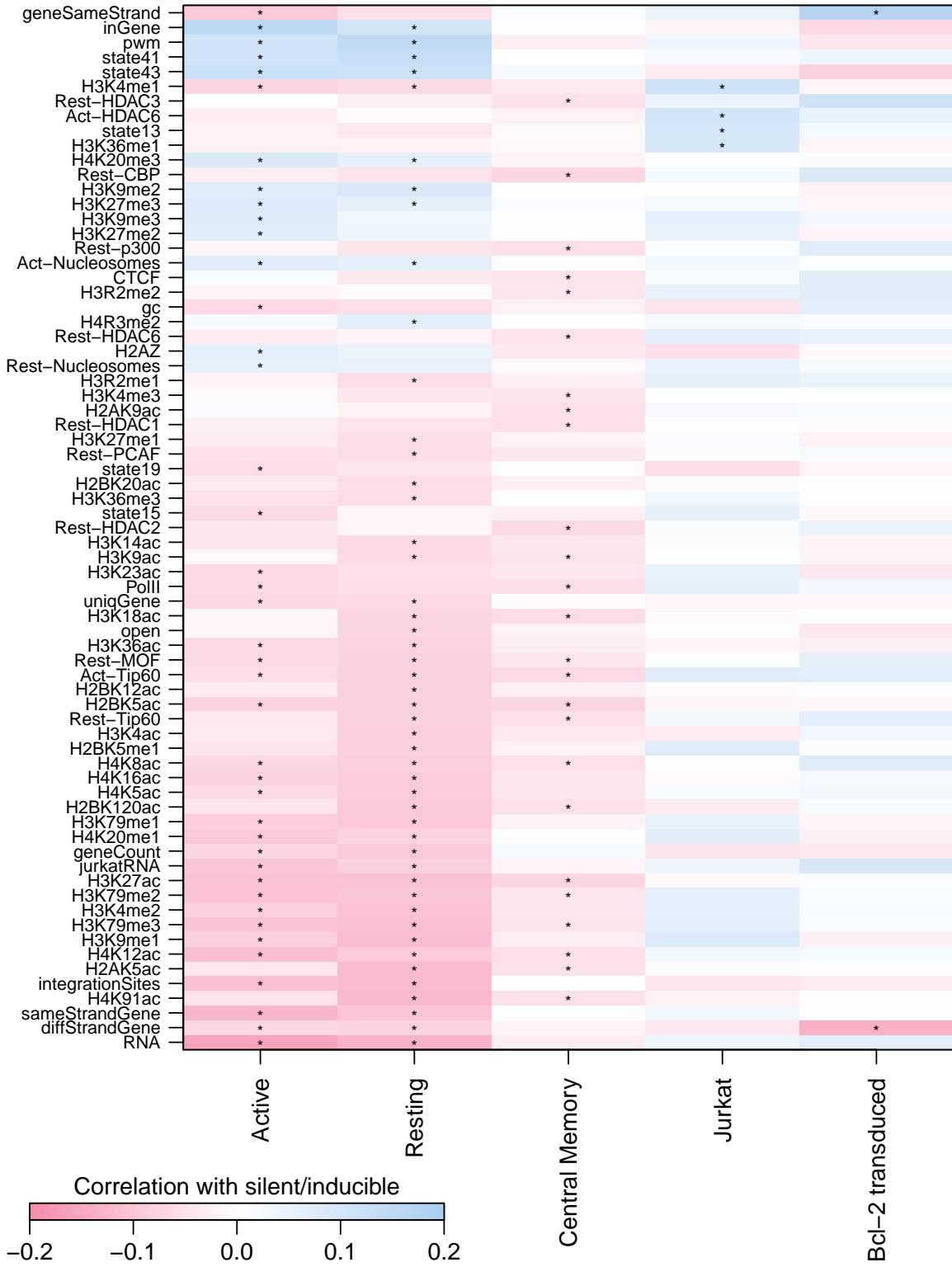


Figure 2.1: Spearman rank correlation between proviral expression status and genomic features. Only genomic features with at least one correlation with latency with a false discovery rate q -value < 0.01 (marked by asterisks) are shown.

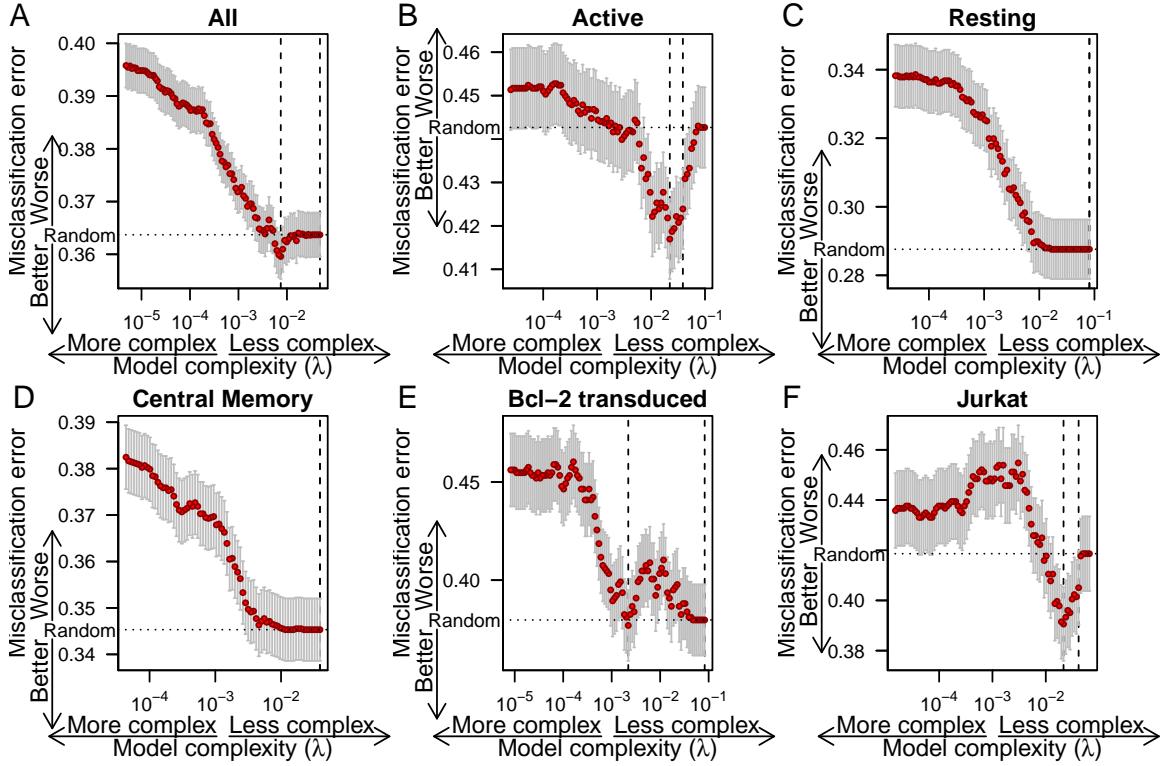


Figure 2.2: Misclassification error from cross validation for lasso regressions of silent/inducible status on genomic features as a function of λ , the regularization coefficient for the lasso regression, for all cell culture models combined and each individual cell culture model. The number of variables included and size of coefficients in the model increases to the left. Whiskers show the standard error of mean misclassification error. Dashed vertical lines indicate the minimum misclassification error and the simplest model within one standard error. Dotted horizontal line indicates the misclassification error expected from random guessing.

671 with dummy variables indicating cellular model. The λ smoothing parameter of the lasso
 672 regression was optimized by finding the λ with lowest classification error in 480-fold cross
 673 validation and finding the simplest model with misclassification error within one standard
 674 error.

675 The proportion of silent/inducible sites varied between the samples. To avoid the model
 676 overfitting on this source of variation, an indicator variable for each sample was included in
 677 the base model. The base model with no genomic variables was selected as the best model by
 678 cross validation (Figure 2.2A). This suggest that there is not a consistent linear relationship
 679 between an additive combination of genomic variables and latency across all models.

680 When each dataset was fit individually with leave-one-out cross validation, improvements in
681 cross-validated misclassification error were only observed in the Active CD4⁺ (5.8% decrease
682 in misclassification error, standard error: 2.1) and Jurkat (6.7% decrease in misclassification
683 error, standard error: 3.5) samples (Figure 2.2B-F). There was no overlap in variables
684 selected for the Active CD4⁺ and Jurkat samples.

685 Finding little global association between latency and genomic features, we investigated
686 whether predictors of latency reported previously by single studies were consistently associ-
687 ated with latency across studies.

688 **2.4.2 Cellular transcription**

689 Model systems with defined integration sites show upstream transcription can interfere with
690 viral transcription⁴³⁵ and that cellular transcription in the same orientation may interfere
691 with viral transcription⁴⁰⁵ or increase viral transcription⁴⁰⁶ and in opposite orientations
692 may decrease transcription⁴⁰⁶. In integration site studies, integration outside genes appears
693 to increase latency⁴⁰² but high transcription of nearby host cell genes may cause increased
694 latency^{402,403}. In addition, Tat or other viral proteins may affect cellular transcription^{319,436}.

695 To look at transcription and latency, we ran a logistic regression of silent/inducible status
696 on a quartic function of RNA expression, as determined by RNA-Seq reads within 5,000
697 bases in Jurkat cells for the Jurkat sample or CD4⁺ T cells for the remaining samples,
698 interacted with indicator variables encoding cell culture model. There appears to be little
699 agreement between samples (Figure 2.3). The Resting CD4⁺ and Active CD4⁺ datasets
700 show an enrichment in silent proviruses in regions with low gene expression. The other three
701 studies show the opposite or no relationship for low expression regions. The two samples
702 showing increased silence in areas of low expression (Resting CD4⁺ and Active CD4⁺) are
703 from a study that did not check whether inactive viruses could be activated. One possible
704 explanation is that regions with low gene transcription may harbor proviruses that are not
705 easily activated, though some other discrepancy between *in vitro* systems could also explain

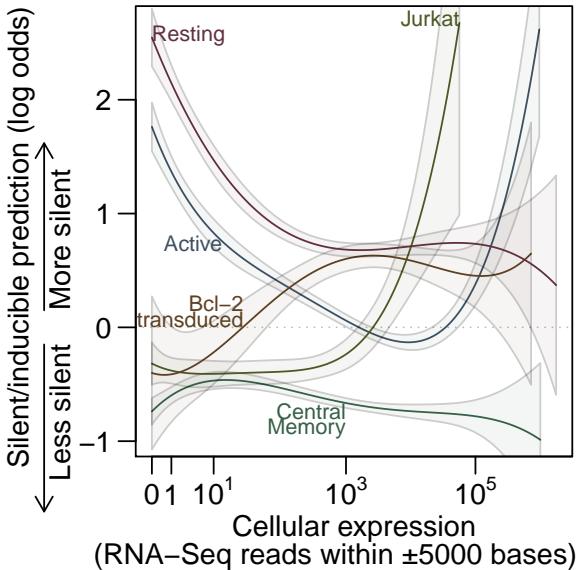


Figure 2.3: Predictions from a logistic regression of silent/inducible status on cellular RNA expression. High y-axis values are predicted to be silent/inducible. Dashed line shows where equal odds of silent/inducible and expressed are predicted. Solid lines show predictions from the regression for each sample and shaded regions indicate one standard error from the modeled predictions.

706 the difference. Both the Jurkat and Active CD4⁺ samples appear to increase in latency with
 707 increasing expression while the remaining three studies did not show a strong trend.

708 2.4.3 Orientation bias

709 Shan et al.⁴⁰³ reported that inducible proviruses were oriented in the same strand as the
 710 host cell genes into which they had integrated more often than chance. This orientation bias
 711 was still reproduced after our reprocessing of the Bcl-2 transduced CD4⁺ sample from Shan
 712 et al.⁴⁰³. However, the proportion of provirus oriented in the same strand as host genes did
 713 not differ significantly from 50% in the other samples (Figure 2.4). Perhaps orientation bias
 714 and transcriptional interference are especially sensitive to parameters of the model system.

715 2.4.4 Gene deserts

716 Lewinski et al.⁴⁰² reported increased latency in gene deserts. In the collected data, integration
 717 outside known genes was associated with latency (Fisher's exact test, $p < 10^{-6}$). This
 718 seemed to largely be driven by the Active CD4⁺ and Resting CD4⁺ samples with significant
 719 association found individually in only those two samples (both $p < 10^{-8}$) and no significant
 720 association observed in the other three samples (Figure 2.5A). Looking only at integration
 721 sites outside genes, silent sites in the Resting CD4⁺ sample had a mean distance to the

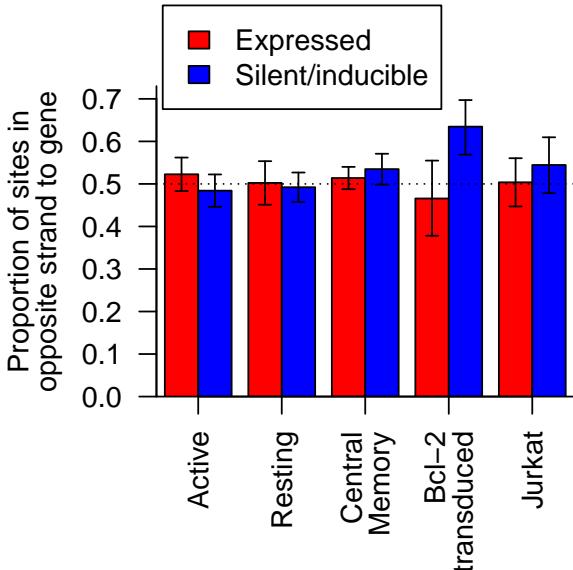


Figure 2.4: The proportion of provirus integrated in the opposite strand compared to cellular genes in silent/inducible (blue) and expressed (red) samples. Error bars show the 95% Clopper-Pearson binomial confidence interval.

⁷²² nearest gene 2.5 times greater than that of expressed sites (95% CI: 2.2–6.2 \times , $p < 10^{-6}$,
⁷²³ Welch two sample t-test on log transformed distance) (Figure 2.5B). The Active CD4 $^{+}$
⁷²⁴ sample had a small difference that did not survive Bonferroni correction.

⁷²⁵ Lewinski et al.⁴⁰² also reported decreased latency near CpG islands and reasoned this was
⁷²⁶ tied to the increased latency in gene deserts. In the Resting CD4 $^{+}$ sample, silent sites were
⁷²⁷ on average further from CpG islands than expressed sites (Bonferroni corrected Welch's two
⁷²⁸ sample T test, $p = 0.006$), but there was no significant relationship between silent/inducible
⁷²⁹ status and log distance to CpG island after Bonferroni correction if the integration site's
⁷³⁰ location inside or outside of a gene was accounted for first (analysis of deviance).

⁷³¹ 2.4.5 Alphoid repeats

⁷³² Alphoid repeats are repetitive DNA sequences found largely in the heterochromatin of
⁷³³ centromeres⁴³⁷. Integration near heterochromatic alphoid repeats has been reported to
⁷³⁴ associate with latency^{397,402,404}. Looking only at uniquely mapping sites, there was no
⁷³⁵ statistically significant association between latency and location inside an alphoid repeat in
⁷³⁶ pooled or individual samples (Fisher's exact test).

⁷³⁷ Since alphoid repeats are both problematic to assemble in genomes and difficult to map

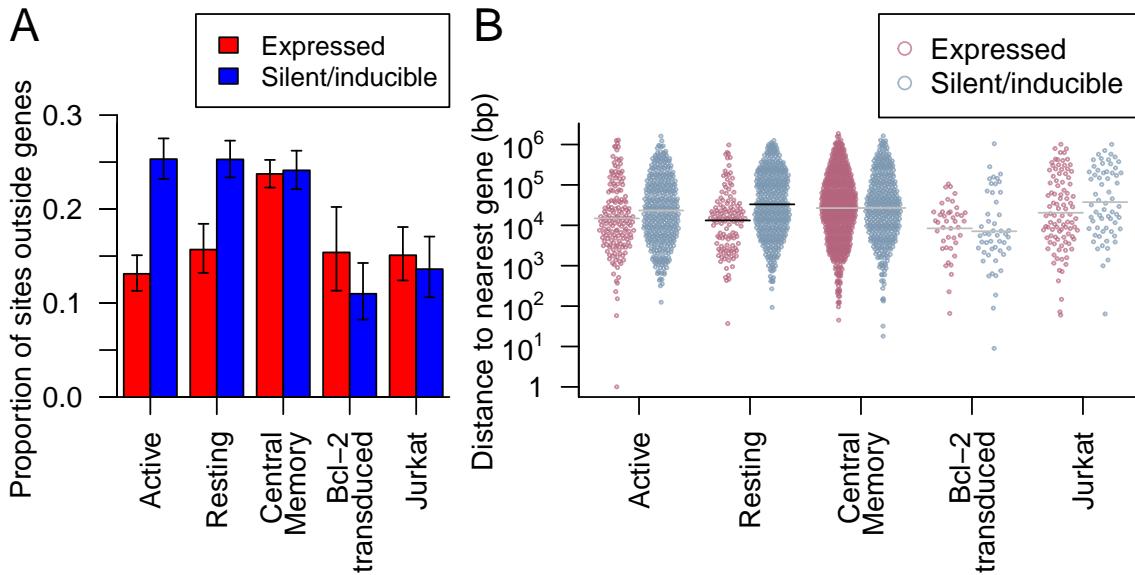


Figure 2.5: (A) The proportion of provirus integrated outside genes in silent/inducible (blue) and expressed (red) samples. Error bars show the 95% Clopper-Pearson binomial confidence interval. (B) The nearest distance to any gene for integration sites (points) outside genes in the five samples. Points are spread in proportion to kernel density estimates. Horizontal lines indicate sample means where there was a significant difference in means between silent/inducible and expressed provirus (black) or no significant difference (grey).

738 onto, we reasoned that some alphoid hits might be lost or miscounted in the filtering
 739 procedures of the standard workup. To counteract this, we treated each sequence read as an
 740 independent observation of a proviral integration and included sequence reads with more
 741 than one best scoring alignment. For multiply aligned reads, we considered the read to have
 742 been inside an alphoid repeat if any of its best scoring alignments fell within a repeat. We
 743 found 74 reads with potential alphoid mappings. Integration inside alphoid repeats was
 744 significantly associated with the expression status of a provirus in the Resting CD4⁺, Jurkat
 745 and Central Memory CD4⁺ datasets (Bonferroni corrected Fisher's exact test, all $p < 0.05$)
 746 and approached significance in the Active CD4⁺ dataset ($p = 0.053$) (Figure 2.6). The Bcl-2
 747 transduced CD4⁺ data did not contain any integration sites in alphoid repeats, probably due
 748 to 1) the relatively low number of integration sites in the dataset and 2) to the requirement
 749 for cleavage at two Pst1 restriction sites, which are not found in the consensus sequence of
 750 alphoid repeats⁴³⁸. Of the 1340 repeat types in the RepeatMasker database⁴³⁸, only alphoid
 751 repeats achieved a significant association with proviral expression in more than two datasets.

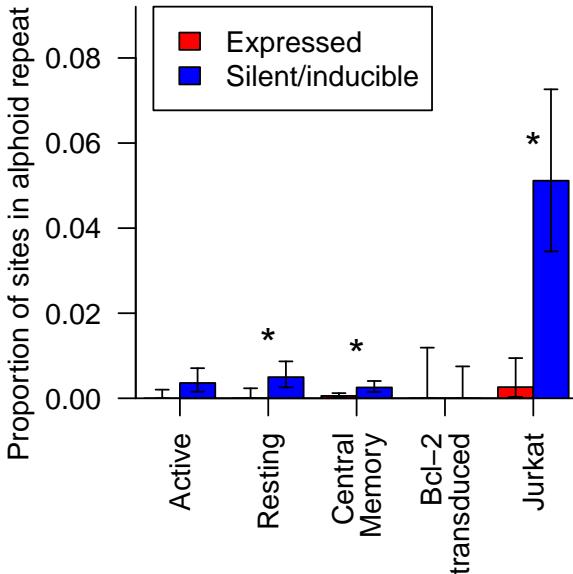


Figure 2.6: The proportion of integration sites with matches in alphoid repeats in silent/inducible (blue) and expressed (red) cells in five samples. Error bars show the 95% Clopper-Pearson binomial confidence interval. Asterisks indicate significant associations between integrations within an alphoid repeat and proviral expression status (Bonferroni corrected Fisher’s exact test $p < 0.05$).

752 2.4.6 Acetylation

753 Histone marks or chromatin remodeling, especially involving the key “Nuc-1” histone near
 754 the transcription start site in the viral LTR, appear to affect viral expression^{398,439,440}.
 755 Based on this effect, histone deacetylase inhibitors have been developed as potential HIV
 756 treatments and show some promise in disrupting latency⁴¹⁰. In these genome-wide datasets,
 757 we do not have information on the state of individual LTR nucleosomes. However, repressive
 758 chromatin does seem to spread to nearby locations if not blocked by insulators^{394,395} and
 759 the state of neighboring chromatin could affect proviral transcription independently of
 760 provirus-associated histones.

761 We found that the number of ChIP-seq reads near an integration site from several histone
 762 acetylation marks (Figure 2.1) were associated with efficient expression in the Active CD4⁺,
 763 Resting CD4⁺ and Central Memory CD4⁺ samples. H4K12ac had the strongest association
 764 (Bonferroni corrected Fisher’s method combination of Spearman’s ρ , $p < 10^{-25}$) with
 765 silence/latency (Figure 2.7A).

766 Although the appearance of several significantly associated acetylation marks might suggest
 767 acetylation exerts a considerable effect on the expression of a provirus, there are strong

768 correlations among these marks, so their effects may not be independent. To account for
769 the correlations between these variables, we performed a principal component analysis
770 (PCA) to convert the correlated acetylation marks into a series of uncorrelated principal
771 components that capture much of the variance within a few components. Here, the first
772 principal component explained 59% of the variance and the first ten components 84%.
773 Several of these principal components again displayed significant associations with latency
774 in the Active CD4⁺, Resting CD4⁺ and Central Memory CD4⁺ samples but no significant
775 correlations in the Bcl-2 transduced CD4⁺ or Jurkat samples (Figure 2.7B). A logistic
776 regression of expression status on the first ten principal components and sample did not
777 reduce misclassification error from a base model including only sample in 480-fold cross
778 validation (base model misclassification error: 36.4%, PCA model: 36.5%). This suggests
779 that acetylation of neighboring chromatin does not exert strong effects on latency in all
780 samples.

781 **2.4.7 Clustering**

782 We reasoned that if there was a strong relationship between latency and chromosomal
783 position, then integration sites that are near one another on the same chromosome should
784 share the same expression status more often than expected by chance. To test this, we
785 compared how often pairs of proviruses shared the same expression status in relation to
786 the distance between the two sites (Figure 2.8). Pairs of sites with little distance between
787 integration locations did share the same expression status more often than expected by
788 chance (e.g. neighbors closer than 100bp, Fisher exact test $p = 0.0002$). Breaking out the
789 data to separate between sample and within sample pairings showed that this matching was
790 limited to neighbors within the same experimental model (Figure 2.8), emphasizing that
791 chromosomal environment does appear to influence latency, but the factors involved differ
792 among experimental models of latency.

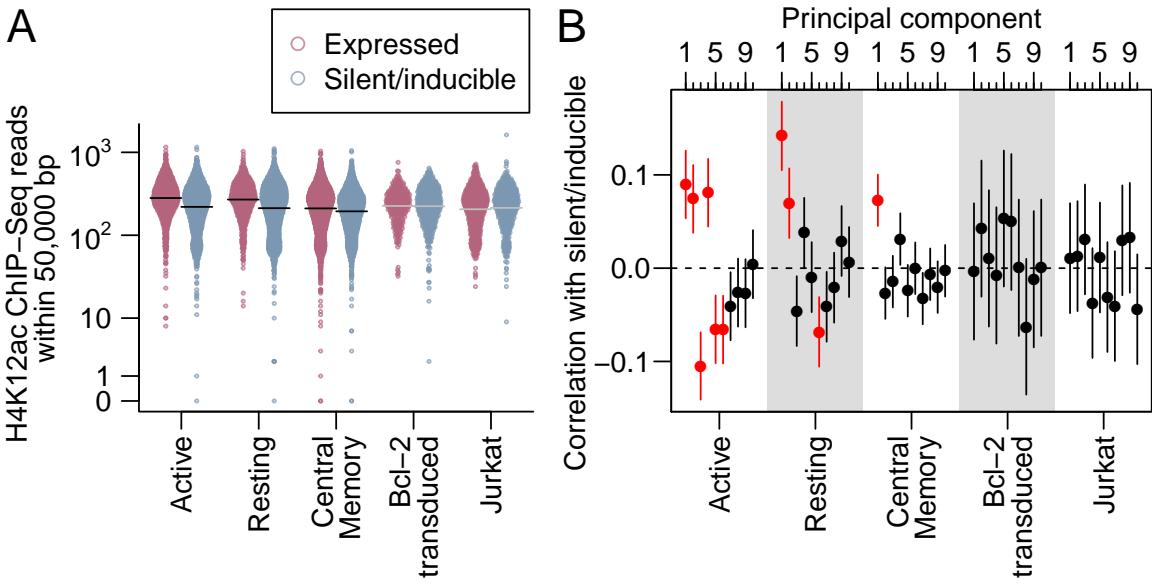


Figure 2.7: (A) The number of ChIP-seq reads for H4K12ac, the histone mark with the lowest Fisher's method p -value for correlation with latency, within 50,000 bases across the five samples. Integration sites (points) are spread in proportion to kernel density estimates. Horizontal lines indicate sample means where there was a significant difference (black) in means between silent/inducible and expressed provirus or no significant difference (grey). (B) The correlation (points) and its 95% confidence interval (vertical lines) between principal components of acetylation and silent/inducible status for each of the five samples. Red indicates correlations with a Bonferroni-corrected p -value < 0.05 .

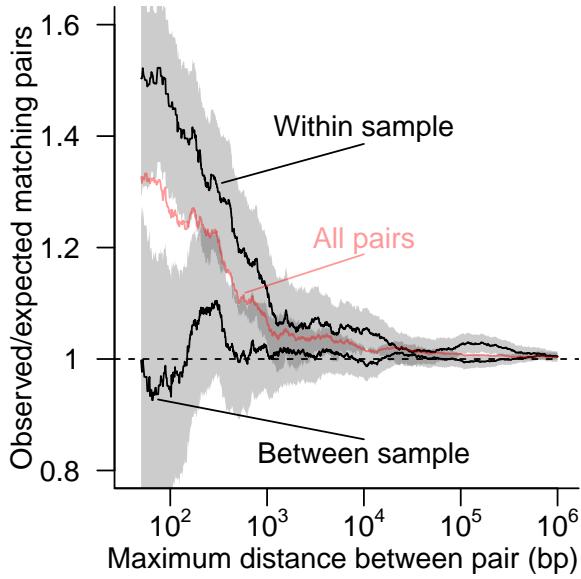


Figure 2.8: The ratio of the number of pairs of proviruses with matching expression status to the number of matches expected by random pairings given the frequency of silent/inducible proviruses. All possible pairs of proviruses integrated within a given distance of each other on the same chromosome (red line) were separated into two sets; one with both proviruses from within the same cell culture model and one with proviruses paired between two different cell culture models (black lines). The shaded region shows the 95% Clopper-Pearson binomial confidence interval for within and between sample pairings. The dashed horizontal line shows the ratio of 1 expected if there is no association between the expression status of neighboring proviruses.

793 2.5 Conclusions

794 Here we compared the latency status of HIV-1 proviruses in five model systems with the
795 genomic features surrounding their integration sites. Surprisingly, no relationships between
796 genomic features near the integration location and latency achieved significance in all models.

797 Proviruses from the same cellular model integrated in nearby positions did share the same
798 latency status much more often than predicted by chance, indicating the existence of local
799 features influencing latency, but these were not consistent among models. This suggests that
800 whatever features are affecting latency are highly local and model-specific, and that we may
801 not have access to all relevant chromosomal features e.g. 441–444.

802 In addition to differences in experimental conditions, methodological issues have the potential
803 to obscure patterns. Examples include multiply infected cells, inactivated viruses and
804 inaccurate assessment of HIV gene activity—each of these are discussed below.

805 A latent provirus integrated into the same cell as an expressed provirus will be erroneously
806 sorted as expressed, potentially confounding analysis. A low multiplicity of infection (MOI)
807 will help to avoid this problem, but there is still the potential for a significant proportion of
808 the cells studied to contain multiple integrations. This problem arises because although cells
809 with multiple integrations form a small proportion of total cells, most of the total are cells
810 lacking an integrated provirus and thus are excluded by experimental design. For example,
811 assuming integrations are Poisson distributed with an MOI of 0.1 (1 integration per 10 cells),
812 90.5% of cells will not contain a provirus, 9% of cells will contain one proviral integration
813 and 0.5% of cells will contain multiple integrations. The cells without an integration are
814 not amplified by HIV-targeted PCR leaving only 9.5% of the total cells. Of these cells
815 actually under study, 4.9% will contain multiple integrations. Thus the signal from expressed
816 proviruses may be muted by the presence of latent proviruses in the expressed population.

817 The replication cycle of HIV is error prone, and a significant proportion of virions contain
818 mutated genomes⁸⁶. In studies that do not check for inducibility, mutant proviruses

819 integrated in regions of the genome otherwise favorable to proviral expression can be sorted
820 into the latent pool due to mutational inactivation. This problem of inactivated provirus
821 is worse when latent provirus are rare and exacerbated further when looking at latency in
822 the cells of HIV patients due to selective enrichment of inactivated proviruses incapable
823 of spreading infection¹³⁷. Here, the effects of mutation are minimized in the datasets that
824 required inducible viral expression (Jurkat, Bcl-2 transduced CD4⁺, Central Memory CD4⁺)
825 but may be a confounder in the two datasets that were sorted based on lack of viral expression
826 only (Active CD4⁺, Resting CD4⁺).

827 Inaccurate staining or leaky markers may also result in misclassification of proviruses. False
828 positives and false negatives will result in incorrectly sorted latent and expressed integrations.
829 For example, if 5% of cells not containing Gag are labeled as Gag+ and there are an equal
830 amount of latent and expressed integration sites, then 4.8% of integrations labeled expressed
831 will actually be latent. If a category is rare, false staining has even greater potential to cause
832 error. For example, if only 5% of sites are latent and a Gag stain has a false negative rate
833 of 5%, then we would expect 48.7% of sites classified as latent to actually be mislabeled
834 expressed integrations.

835 Attempts to induce latent proviruses in patients have so far focused on using histone
836 deacetylase inhibitors, raising interest in associations with histone acetylation in these data.
837 An important caveat in results from these genome-wide data is that histone modification
838 near the integrated provirus may not be representative of modification within the provirus
839 at the key “Nuc-1” nucleosome of the transcription start site⁴⁴⁰, though local correlations in
840 chromatin states are well established from studies of position effect variegation^{394,395}. We
841 found that some histone acetylation marks were significantly associated with viral expression
842 in some but not all samples (Figures 2.1, 2.7). This lack of association may be due to a
843 lack of power in these studies, but the confidence intervals suggest that any correlations
844 between acetylations and latency are unlikely to be strong. These weak correlations raise
845 the possibility that there are populations of latent proviruses that are not associated with

846 acetylation and may not be inducible by histone deacetylase inhibitors.

847 This study highlights that the choice of model system can have a large effect on measurements
848 of latency. Further studies are needed to determine which *in vitro* models best reflect latency
849 *in vivo*. Different cell models may report genuinely different mechanisms of latency. While we
850 did see some relationship between histone acetylation and latency, paralleling a recent clinical
851 trial of SAHA⁴¹⁰, associations with histone acetylation did not explain a large fraction of
852 the difference between latent and expresssed proviruses in any of the five models. One
853 possible explanation is that there may be multiple mechanisms that maintain proviruses in a
854 latent state. To be successful, shock-and-kill treatments must induce and destroy all latent
855 proviruses to eliminate HIV from an infected individual, raising the question of whether
856 multiple simultaneous inducing treatments will be necessary.

857 2.6 Availability of supporting data

858 Sequence reads from the Central Memory CD4⁺ sample reported here, the Resting CD4⁺
859 and Active CD4⁺ data reported by Pace et al.⁴⁰⁴, the Bcl-2 transduced CD4⁺ data reported
860 by Shan et al.⁴⁰³ and reprocessed data originally reported by Lewinski et al.⁴⁰² are available
861 at the Sequence Read Archive under accession number SRP028573.

862 2.7 Author's contributions

863 SS-M led the computational analysis, with assistance from CCB and NM. MKL, DL and JG
864 analyzed integration sites using IonTorrent sequencing. MF, AB and VP prepared DNA
865 from latent and activated T cells using the Central Memory CD4⁺ model. LS, RFS, MJP,
866 LMA and UO'D contributed data and suggestions. SS-M, KEO and FDB planned the
867 overall study, and SS-M and FDB wrote the paper. All authors read and approved the final
868 manuscript.

869 **2.8 Acknowledgements**

870 We would like to thank Werner Witke for assistance with IonTorrent sequencing. This
871 work was supported in part by NIH grants R01 AI 052845-11 to FDB, R21AI 096993 and
872 K02AI078766 to UO'D, 5T32HG000046 to SS-M, AI087508 to VP and R01AI038201 to
873 JG, the Penn Genome Frontiers Institute, the University of Pennsylvania Center for AIDS
874 Research (CFAR) P30 AI 045008 and the University of California, San Diego, CFAR P30
875 AI036214.

876 **CHAPTER 3: Dynamic regulation of HIV-1 mRNA populations**
877 **analyzed by single-molecule enrichment and long-read**
878 **sequencing**

This chapter was originally published as:

KE Ocwieja, S Sherrill-Mix, R Mukherjee, R Custers-
Allen, P David, M Brown, S Wang, DR Link, J Olson
et al. 2012. Dynamic regulation of HIV-1 mRNA popu-
lations analyzed by single-molecule enrichment and long-
read sequencing. *Nucleic Acids Res*, 40:10345–10355. doi:
10.1093/nar/gks753

FD Bushman, K Travers, DR Link, E Schadt, KE Ocwieja and R Mukher-
jee conceived and designed the experiment. KE Ocwieja and R Custers-
Allen carried out sample preparation and experimental validation. P
David and J Olson performed single-molecule amplification. K Travers
and S Wang performed sequencing. KE Ocwieja, M Brown and I analyzed
the data. KE Ocwieja and I produced the figures. KE Ocwieja, FD
Bushman and I wrote the manuscript.

Supplementary data are available at [http://nar.oxfordjournals.org/
content/40/20/10345/suppl/DC1](http://nar.oxfordjournals.org/content/40/20/10345/suppl/DC1)

880 **3.1 Abstract**

881 Alternative RNA splicing greatly expands the repertoire of proteins encoded by genomes.
882 Next-generation sequencing (NGS) is attractive for studying alternative splicing because
883 of the efficiency and low cost per base, but short reads typical of NGS only report mRNA
884 fragments containing one or few splice junctions. Here, we used single-molecule amplification
885 and long-read sequencing to study the HIV-1 provirus, which is only 9700 bp in length, but
886 encodes nine major proteins via alternative splicing. Our data showed that the clinical isolate
887 HIV_{89.6} produces at least 109 different spliced RNAs, including a previously unappreciated
888 ~1 kb class of messages, two of which encode new proteins. HIV-1 message populations

889 differed between cell types, longitudinally during infection, and among T cells from different
890 human donors. These findings open a new window on a little studied aspect of HIV-1
891 replication, suggest therapeutic opportunities and provide advanced tools for the study of
892 alternative splicing.

893 3.2 Introduction

894 Alternative splicing greatly expands the information content of genomes by producing
895 multiple mRNAs from individual transcription units. Approximately 95% of human genes
896 with multiple exons encode RNA transcripts that are alternatively spliced, and mutations
897 that affect alternative splicing are associated with diseases ranging from cystic fibrosis to
898 chronic lymphoproliferative leukemia^{270,313,446–448}. Work to decipher an RNA ‘splicing code’
899 has revealed that multiple interactions between trans-acting factors and RNA elements
900 determine splicing patterns, though regulation is little understood for most genes³¹⁴.

901 The integrated HIV-1 provirus is ~9700 bp in length and has a single transcription start
902 site, but according to the published literature yields at least 47 different mRNAs encoding 9
903 proteins or polyproteins, making HIV an attractive model for studies of alternative splicing³⁰⁶.
904 HIV mRNAs fall into three classes: the unspliced RNA genome, which encodes Gag/Gag-Pol;
905 partially spliced transcripts, ~4 kb in length, encoding Vif, Vpr, a one-exon version of Tat,
906 and Env/Vpu; and completely spliced mRNAs of roughly 2 kb encoding Tat, Rev and Nef
907 (Figure 3.1A). Additional rare ‘cryptic’ splice donors (5’ splice sites) and acceptors (3’ splice
908 sites) contribute even more mRNAs^{308,449–453}. A complex array of positive and negative
909 cis-acting elements surrounding each splice site regulates the relative abundance of the
910 HIV-1 mRNAs, and disrupting the balance of message ratios impairs viral replication in
911 several models^{285,288,289,293,454–457}. Studies have suggested strain-specific splicing patterns
912 may exist^{306,458,459}. However, detailed studies of complete message populations have not
913 been reported for clinical isolates of HIV-1.

914 Several groups have demonstrated tissue- and differentiation-specific splicing of cellular

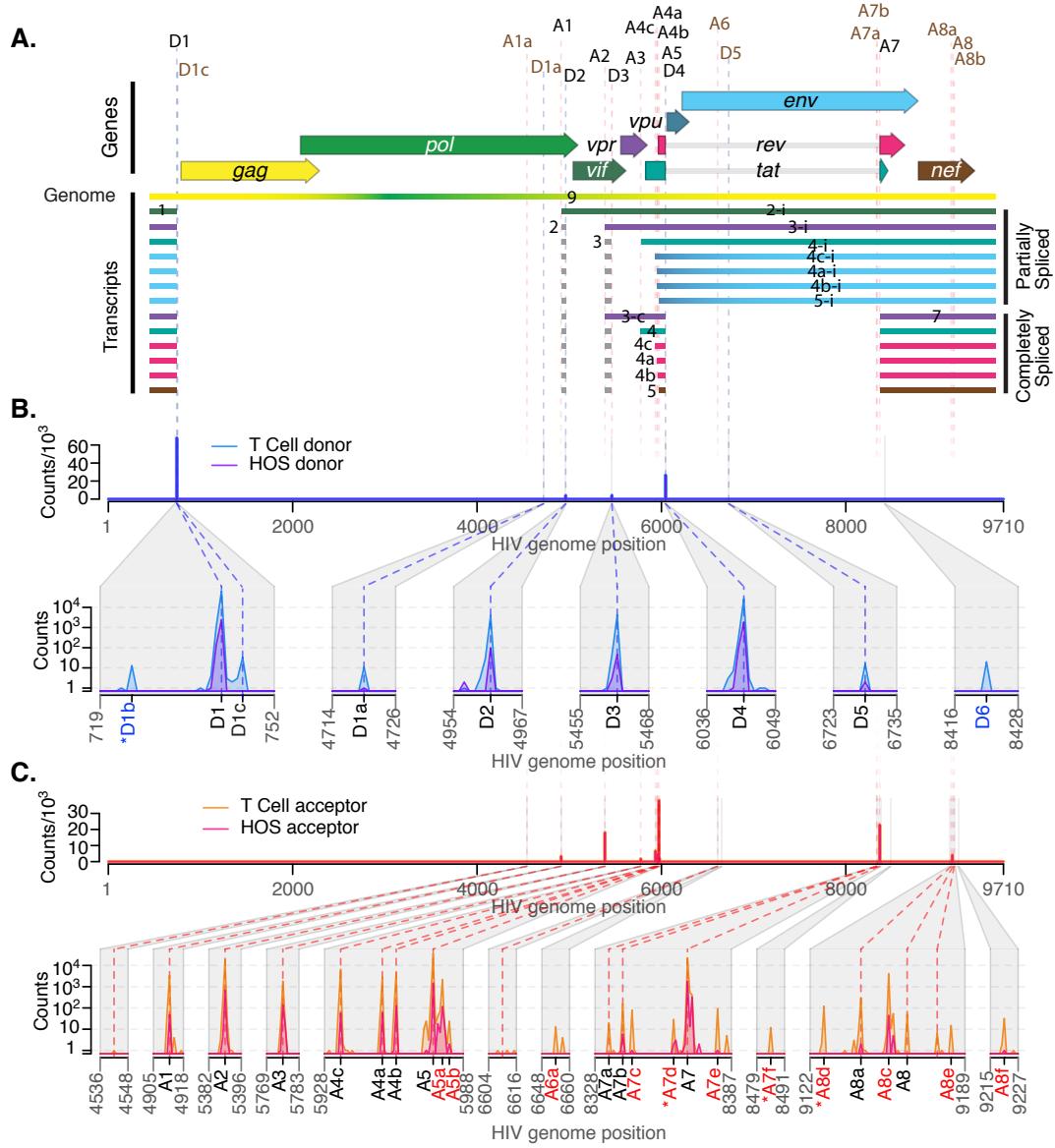


Figure 3.1: Mapping the splice donors and acceptors of HIV_{89.6}. PacBio sequence reads of HIV_{89.6} cDNA from infected HOS-CD4-CCR5 (HOS) and CD4⁺ T cells were aligned to the HIV_{89.6} genome shown in (A). Exons of the conserved HIV-1 transcripts are colored according to the encoded gene. Conserved (black) and published cryptic (brown) splice donors ('D') and acceptors ('A') are shown. Gaps in HIV-1 sequence alignments with at least one end located at a published or verified splice donor or acceptor were defined as introns. For each base of the HIV_{89.6} genome, the number of sequence reads in which that base occurred at the 5'-end (B) or 3'-end (C) of an intron is plotted for each cell type. Putative splice donors and acceptors were defined as loci that were found in at least 10 reads at the 5'- and 3'-ends of introns in sequence alignments from T-cell infections. Regions containing splice sites are enlarged for clarity. Asterisks indicate putative splice sites that are adjacent to dinucleotides other than the consensus GT and AG.

915 genes^{313,460,461}. Importantly for HIV, these include changes during T-cell activation^{462,463},
916 raising the question of how cell-specific splicing affects HIV replication. While most studies
917 of HIV-1 splicing have been conducted in cell lines using lab-adapted viral strains, limited
918 works in PBMCs from infected patients, monocytes and macrophages have suggested that
919 differences may indeed exist in relevant cell types^{281,449,458,464}. Moreover, human splicing
920 patterns differ between individuals, but such polymorphisms have not been investigated in
921 the context of HIV infection^{311,312}.

922 Here, we use deep sequencing to comprehensively characterize the transcriptome of an early
923 passage clinical isolate, HIV_{89.6}⁴⁶⁵, in primary CD4⁺ T cells from seven human donors
924 and in the human osteosarcoma (HOS) cell line. Many deep sequencing techniques provide
925 short reads, which rarely query more than a single exon-exon junction. To distinguish
926 the full structure of HIV-1 mRNAs, which can contain several splice junctions, we used
927 Pacific Biosciences (PacBio) sequencing technology, which yields read lengths up to 10 kb³²³.
928 We used RainDance Technologies single-molecule PCR enrichment to preserve ratios of
929 RNAs during preparation of sequencing templates. We identified previously published and
930 novel HIV-1 transcripts and determined that HIV_{89.6} encodes a minimum of 109 different
931 splice forms. These included a new size class of transcripts, some of which contain novel
932 open reading frames (ORFs) that encode new proteins. We also found significant variation
933 between cell types, over time during infection of HOS cells and among individuals. These
934 data reveal unanticipated complexity and dynamics in HIV-1 message populations, begin
935 to clarify a little studied dimension of HIV-1 replication and suggest possible targets for
936 therapeutic interventions.

937 **3.3 Materials and methods**

938 **3.3.1 Cell culture and viral infections**

939 HIV_{89.6} was generated by transfection and subsequent expansion in SupT1 cells. Primary
940 T cells were isolated by the University of Pennsylvania Center for AIDS Research Im-

941 munology core and confirmed to be homozygous for the wild-type CCR5 allele as shown
942 in Supplementary Table S1 and described in Supplementary Methods. HOS-CD4-CCR5
943 cells^{466,467} were obtained through the AIDS Research and Reference Reagent Program,
944 Division of AIDS, NIAID, NIH from Dr Nathaniel Landau. Single round infections in T
945 cells and HOS-CD4-CCR5 cells were performed using standard methods (see Supplementary
946 Methods).

947 **3.3.2 RNA and reverse transcription**

948 Total cellular RNA was purified using the Illustra RNA kit (GE Life Sciences, Fairfield, CT,
949 USA) from 5×10^6 cells per infection. Viral cDNA was made using a reverse transcription
950 primer complementary to a sequence in U3 (RTprime, Supplementary Table S2). We used
951 Superscript III reverse transcriptase (Invitrogen) in the presence of RNaseOUT (Invitrogen)
952 to conduct first-strand cDNA synthesis from equal amounts of total cellular RNA from each
953 HOS-CD4-CCR5 time point (15.2 μ g) and from each T-cell infection (3 μ g) according to the
954 manufacturer's instructions for gene-specific priming of long cDNAs, and then treated with
955 RNaseH (Invitrogen). We checked for full reverse transcription of the longest (unspliced)
956 viral cDNAs by PCR using primers that bind in the first major intron of HIV_{89.6} (keo003,
957 keo004, Supplementary Table S2, data not shown).

958 **3.3.3 Bulk RT-PCR and cloning**

959 Transcripts were amplified from cellular RNA using the Onestep RT-PCR kit (Qiagen)
960 with primer pairs keo056/keo057 and keo058/keo059 (Supplementary Table S2) with the
961 following amplification: 5 cycles of 30 s at 94°C, 12 s at 56°C, 40 s at 72°C; then 30 cycles
962 of 30 s at 94°C, 14 s at 56°C, 40 s at 72°C; and finally 10 min at 72°C. For verification of
963 dynamic changes, primers F1.2 and R1.2 were used with 35 cycles of 30 s at 94°C, 30 s at
964 56°C and 45 s at 72°C followed by 10 min at 72°C. Products were resolved on agarose gels
965 (Nusieve 3:1, Lonza for verification of dynamic changes, Invitrogen for cloning) stained with
966 ethidium-bromide (Sigma) for visualization, or SYBR Safe DNA gel stain (Invitrogen) for

967 cloning (keo056/keo057 amplified material). DNA was purified using Qiaquick gel extraction
968 kit (Qiagen) and cloned using the TOPO TA cloning kit (Invitrogen). Plasmid DNA was
969 prepared using Qiaprep Spin Miniprep kit (Qiagen). Inserts were identified and verified
970 using Sanger sequencing. The cDNAs for *tat*^{8c}, *tat* (1 and 2 exon), *ref*, *rev* and *nef*, and the
971 transcript with exon structure 1-5-8c were cloned into the expression vector pIRES2-AcGFP1
972 (Clonetech) as described in Supplementary Methods.

973 **3.3.4 Assays of protein activity and HIV replication**

974 Activity and HIV replication assays were performed as described in Supplementary Methods.
975 Tat activity expressed from each cDNA was measured in TZM-bl cells¹⁷⁸ (gift of Dr Robert
976 W. Doms). Rev activity was assayed in HEK-293T cells co-transfected with pCMVGagPol-
977 RRE-R, a reporter plasmid from which Gag and Pol are expressed in a Rev-dependent
978 manner (gift of David Rekosh)⁴⁶⁸. Intracellular and released supernatant p24 was measured
979 from cells transfected with expression constructs and infected with HIV_{89.6}.

980 **3.3.5 Western blotting**

981 HEK-293T cells were transfected with expression constructs and treated with MG132 (EMD
982 Chemicals) to inhibit the proteasome or DMSO (Supplementary Methods). Proteins were
983 detected by immunoblotting using a mouse antibody that recognizes the carboxy terminus
984 of HIV-1 Nef diluted 1:1000 in 5% milk (gift of Dr James Hoxie)⁴⁶⁹. Horseradish peroxidase
985 (HRP)-conjugated secondary rabbit-anti-mouse antibody (p0260, DAKO) was used for
986 detection with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).
987 Beta-tubulin was used as a loading control, detected by the HRP-conjugated antibody
988 (ab21058, Abcam).

989 **3.3.6 Single-molecule amplification**

990 Amplification was performed by RainDance Technologies using a protocol similar to that
991 previously reported (detailed description in Supplementary Methods)³²⁰. Amplification

992 was carried out in droplets to suppress competition between amplicons. PCR droplets
993 were generated on the RDT 1000 (RainDance Technologies) using the manufacturer's
994 recommended protocol. The custom primer libraries for this study contained 18 (HOS-CD4-
995 CCR5 cells) or 20 (primary T cells) PCR primer pairs designed to amplify different HIV
996 RNA isoforms (Supplementary Table S2).

997 **3.3.7 Single-molecule sequencing**

998 DNA amplification products from the RainDance PCR droplets were converted to SMRTbell
999 templates using the PacBio RS DNA Template Preparation Kit. Sequencing was performed by
1000 Pacific Biosciences using the PacBio SMRT sequencing technology as described³²³. Sequence
1001 information was acquired during real time as the immobilized DNA polymerase translocated
1002 along the template molecule. Prior to sequence acquisition, hairpin adapters were ligated to
1003 each DNA template end so that DNA polymerase could traverse DNA molecules multiple
1004 times during rolling circle replication (SMRTbell template sequencing⁴⁷⁰), allowing error
1005 control by calculating the consensus ('circular consensus sequence' or CCS). For raw reads,
1006 the average length was 2860 nt, and 10% were > 5000 nt. After condensing into consensus
1007 reads, the mean read length was 249.5 nt, due to the use of a shorter Pacific Biosciences
1008 sequencing protocol to accommodate the small size of many amplicons. Consensus reads of
1009 1% were > 1100 nt. Sequencing data were collected in 45-min movies.

1010 **3.3.8 Data analysis**

1011 Raw reads were processed to produce CCSs. Raw reads were also retained to help in primer
1012 identification and to avoid biasing against long reads. Reads were aligned against the human
1013 genome using Blat⁴²². Misprimed reads matching the RT primer, reads with a CCS length
1014 shorter than 40 nt or raw length shorter than 100 nt and reads matching the human genome
1015 were discarded. Filtered reads were aligned against the HIV_{89.6} reference genome. Potential
1016 novel donors and acceptors were found by filtering putative splice junctions in the Blat
1017 hits for a perfect sequence match 20 bases up- and downstream of the junction, ignoring

1018 homopolymer errors, and requiring that one end of the junction be a known splice site. Local
1019 maximums within a 5-nt span with > 9 such junctions were called as novel splice sites.

1020 Filter-passed reads were aligned against all expected fragments based on primers and known
1021 and novel junctions. Primers were identified in CCS reads by an edit distance ≤ 1 from
1022 the primer in the start or end of the read, in raw reads by an edit distance ≤ 5 from a
1023 concatenation of the primer, hairpin adapter and the reverse complement of the primer, and
1024 in both types of reads by a Blat hit spanning an entire expected fragment.

1025 Gaps in Blat hits were ignored if ≤ 10 bases long or in regions of likely poor read quality
1026 ≤ 20 bases long where an inferred insertion of unmatched bases in the read occurred at the
1027 same location as skipped bases in the reference. Any Blat hits with a gap > 10 nt remaining
1028 in the query read were discarded. If HIV sequence was repeated in a given read (likely due
1029 to PacBio circular sequencing), the alignments were collapsed into the union of the coverage.
1030 Gaps in the HIV sequence found in uninterrupted query sequence were called as tentative
1031 introns. Splice junctions were assigned to conserved or previously identified (published
1032 or in this work) splice sites and reads appearing to contain donors or acceptors further
1033 than 5 nt away from these sites were discarded. Reads with Blat hits outside the expected
1034 primer range were discarded from that primer grouping. The assigned primer pair, observed
1035 junctions and exonic sequence were used to assign each read to a given spliceform (specific
1036 transcript structure) or set of possible spliceforms. Partial sequences that did not extend
1037 through both primers were assigned to specific transcripts if the read contained enough
1038 information to rule out all other spliceforms or if all other possible spliceforms contained
1039 rare (< 1% usage) donors or acceptors (Supplementary Table S3). Otherwise, the read was
1040 called indeterminate.

1041 To calculate the ratios of transcripts within the partially spliced class, we counted the
1042 number of reads for each assigned spliceform amplified by primer pair 1.3 and divided by the
1043 total number of assigned partially spliced reads amplified with these primers (Supplementary
1044 Figure S1 and Supplementary Table S2). Assigned sequences amplified with primer pairs

1045 1.4 and 4.1 (full-length cDNAs, T cells only) were used to calculate ratios of transcripts
1046 within each of the two completely splice classes (~ 2 and ~ 1 kb). To compare ratios of ~ 2
1047 kb transcripts calculated within reads from primer pairs 1.4 and 4.1, we normalized ratios
1048 from pair 4.1 to the *nef* 2 transcript (containing exons 1, 5 and 7). Due to size biases
1049 inherent in the approach, we did not compare across size classes, and unspliced transcripts
1050 were not included in ratio analysis. For all ratio analysis, transcripts including cryptic or
1051 novel junctions were counted only if they appeared in at least five reads, otherwise they
1052 were excluded from the analysis and from the count of total assigned reads.

1053 To estimate the minimum total number of transcripts present, partial sequence reads were
1054 included. Each exon-exon junction occurring in at least five reads and not previously assigned
1055 to a particular transcript (Figure 3.2) was counted as evidence of an additional transcript
1056 (47 additional junctions were detected, see Supplementary Table S4). If two such junctions
1057 could conceivably occur in a single mRNA, we counted only one unless we could verify from
1058 sequence reads that they were amplified from separate cDNAs, resulting in 31 additional
1059 transcripts. The minimum transcript number calculated by a greedy algorithm treating
1060 introns as events in a scheduling problem agreed with the above calculation.

1061 Several groups have demonstrated tissue- and differentiation-specific splicing of cellular
1062 genes^{313,460,461}. Importantly for HIV, these include changes during T-cell activation^{462,463},
1063 raising the question of how cell-specific splicing affects HIV replication. While most studies
1064 of HIV-1 splicing have been conducted in cell lines using lab-adapted viral strains, limited
1065 works in PBMCs from infected patients, monocytes and macrophages have suggested that
1066 differences may indeed exist in relevant cell types^{281,449,458,464}. Moreover, human splicing
1067 patterns differ between individuals, but such polymorphisms have not been investigated in
1068 the context of HIV infection^{311,312}.

1069 For studies of transcript dynamics, reads from primer pairs 1.2, 1.3 and 1.4 containing
1070 junctions between D1 or any donor and each of five mutually exclusive acceptors, A3, A4c,
1071 A4a, A4b, A5 and A5a, were collected and their ratios calculated.

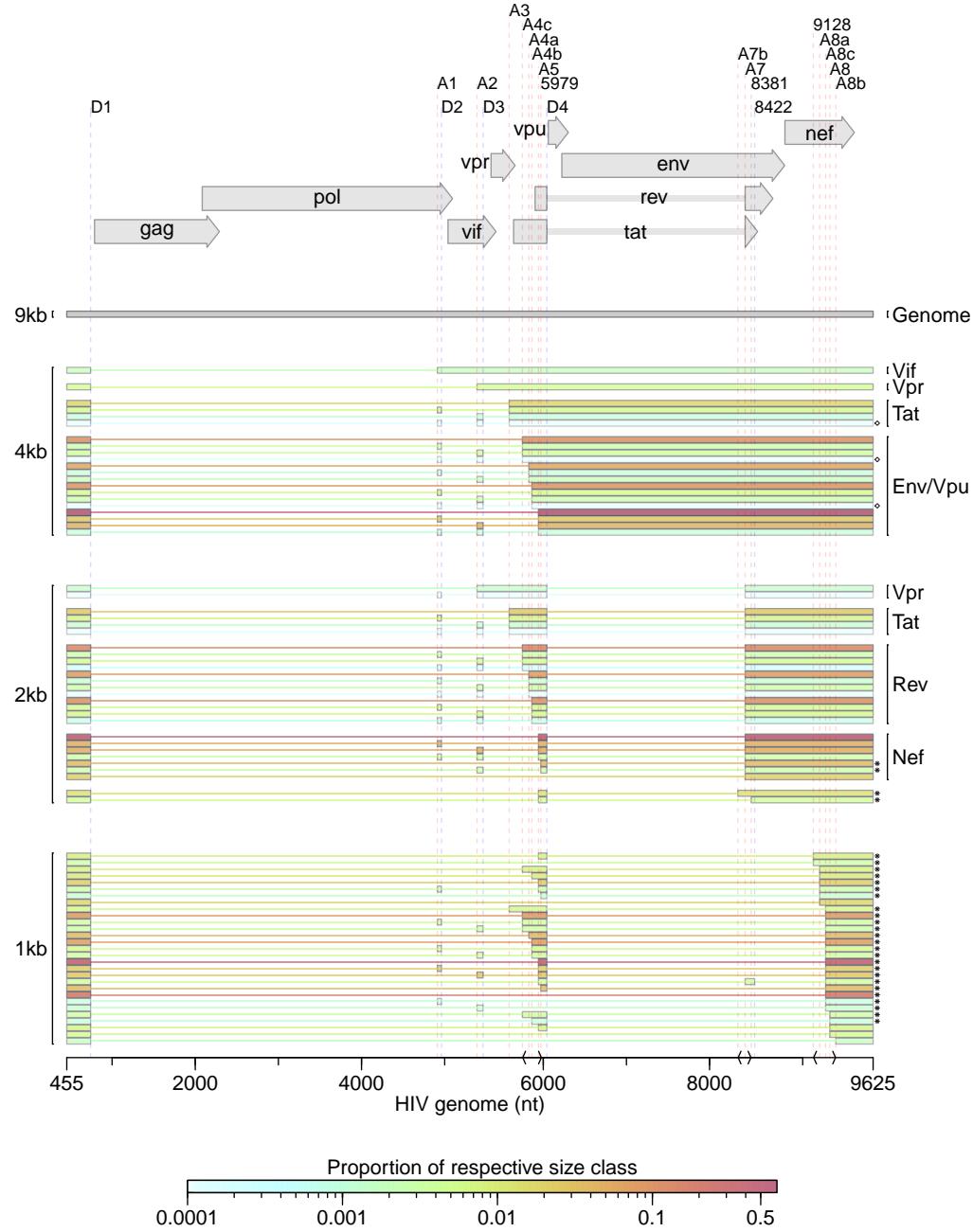


Figure 3.2: HIV_{89.6} transcripts in T cells for which the full message structure was determined are shown arranged by size class. Thick bars correspond to exons and thin lines to excised introns. For the well-conserved transcripts, encoded proteins are indicated. The relative abundance of each transcript within its size class is indicated by color. Asterisks denote transcripts that have not been reported previously to our knowledge. Of the 47 conserved HIV-1 transcripts, three were detected in fewer than five reads (indicated with ◊) and two messages were not detected and are not shown (one encoding Vpr and one encoding Env/Vpu). Depicted non-conserved transcripts (using novel or cryptic splice sites) were each detected in at least five independent sequence reads across samples from at least two different human T-cell donors.

1072 **3.3.9 Statistical analysis**

1073 Statistical modeling was performed using generalized linear modeling as described in Supple-
1074 mentary Report S2. All analyses were performed in R 2.14.0 (R Development Core)⁴¹⁵.

1075 **3.3.10 Data access**

1076 Sequence data is available in the SRA database with the following accession numbers:
1077 SRP014319.

1078 **3.4 Results**

1079 **3.4.1 Sequencing HIV-1 transcripts produced in primary T cells and HOS cells**

1080 In order to characterize HIV-1 transcript populations, we prepared viral cDNA from primary
1081 CD4⁺ T cells of seven different healthy human donors infected in vitro with HIV_{89.6}, an early
1082 passage dual-tropic clade-B clinical isolate (Supplementary Figure S1, human donor data
1083 in Supplementary Table S1)⁴⁶⁵. We also studied HIV messages produced in infected HOS
1084 cells engineered to express CD4 and CCR5 (HOS-CD4-CCR5) because these cells support
1085 efficient HIV replication and engineered variants are widely used in HIV research. HOS
1086 cells were harvested at 18, 24 and 48 hours post infection (hpi) to investigate longitudinal
1087 changes during infection, and for comparison to 48 h infected T cells.

1088 To preserve the relative proportions of template molecules while amplifying the cDNA, we
1089 used RainDance Technologies' single-molecule micro-droplet based PCR³²⁰. Droplet libraries
1090 containing multiple overlapping primer pairs were designed to query all message forms and
1091 allow later calculation of relative abundance (Supplementary Table S2 and Supplementary
1092 Figure S1). Each primer was unique so that sequences could be assigned to a specific
1093 primer pair, which helped reconstruct the origin of sequence reads and deduce message
1094 structures. Amplified DNA products were sequenced using Single Molecule Real-Time
1095 (SMRT) technology from Pacific Biosciences^{323,470}. We obtained 847 492 filtered reads of
1096 amplified HIV-1 transcripts in primary CD4⁺ T cells and 89 350 in HOS cells. The longest

1097 sequenced continuous stretch of HIV-1 cDNA was 2629 bp.

1098 **3.4.2 Splice donors and acceptors**

1099 We aligned PacBio reads containing HIV sequences to the HIV_{89.6} genome and identified
1100 candidate introns as recurring gaps in our sequences. Using this approach, we observed
1101 splicing at each of the widely conserved major splice donors and acceptors and several
1102 published cryptic sites (Figure 3.1A, hereafter referred to by their identifications shown in
1103 this figure, ‘D’ for donors, ‘A’ for acceptors).

1104 In addition, we identified 13 putative novel splice sites: 2 donors and 11 acceptors (Figure
1105 3.1 and Supplementary Table S3). In order to be selected as a bona fide splice site and
1106 remove artifacts possibly created by recombination during sample preparation, we required
1107 that the new acceptor or donor was observed spliced to previously reported splice donors or
1108 acceptors in > 10 sequence reads in CD4⁺ T cells. The most frequently used novel splice site
1109 was an acceptor that we have termed A8c because it lies near A8, A8a and A8b (discussed
1110 in detail below). Additional novel sites are further discussed in Supplementary Report S1.

1111 Most of the new splice sites adhered to consensus sequences for the standard spliceosome
1112 (Supplementary Table S3). However, there appeared to be one splice donor upstream of
1113 D1 with a cytidine in place of the usual uracil 2 nt downstream of the splice site. Similar
1114 ‘GC donors’ appear in 1% of known splice junctions in humans⁴⁷¹. Of the novel splice
1115 acceptors, three were preceded by dinucleotides other than the consensus AG. Alternative
1116 dinucleotides are used infrequently as splice acceptors^{472–475}; however, it is possible that our
1117 deep sequencing method allowed us to observe rare events.

1118 **3.4.3 Structures of spliced HIV_{89.6} RNAs**

1119 To quantify the populations of HIV-1 transcripts, we aligned all reads to the collection of
1120 47 well-established spliced HIV-1 transcripts and detected 45 of them (Figure 3.2). We
1121 additionally aligned reads to the HIV_{89.6} genome allowing all possible combinations of splice

1122 junctions—canonical, cryptic or novel—determined from the sequencing data (Figure 3.1),
1123 yielding an additional 32 complete transcripts, 19 of which were novel. The data also provide
1124 evidence for more novel splice junctions but in incomplete sequences, implying the existence
1125 of additional new transcripts (Supplementary Table S4 and Supplementary Report S1). The
1126 full data set taken together provides evidence for least 109 different HIV_{89.6} transcripts in
1127 primary T cells.

1128 Amplification primers that isolated the two main classes of spliced messages allowed us to
1129 determine the ratios of mRNAs in each (Figure 3.2 and Supplementary Table S5). Within
1130 the partially spliced class of transcripts, *env/vpu*, *tat* (1-exon), *vpr* and *vif* messages existed
1131 in an average ratio of 96:4:< 1:< 1 in CD4⁺ T cells. The ratio of *nef:rev:tat:vpr* within
1132 the ~2 kb transcript class was 64:33:3:< 1. Consistent with previous reports, the most
1133 abundant transcript in each class contained the splice junction from D1 to A5 (D1^A5)—an
1134 *env/vpu* transcript contributing 64% of the partially spliced class, and a completely spliced
1135 *nef* transcript contributing 47% of ~2 kb messages (Figure 3.2)^{306,476}. The relatively
1136 low abundance of transcripts encoding Tat suggests that Tat sufficiently stimulates HIV
1137 transcription elongation at low concentrations, or that the *tat* transcripts must be efficiently
1138 translated. Due to biases inherent in the reverse transcription step, we could only compare
1139 transcripts within each size class, and we note that our methods have not been validated
1140 for empirical quantification. However, the ratios were roughly confirmed using overlapping
1141 sequence reads obtained with alternate primer pairs and by end point RT-PCR analysis of
1142 HIV-1 RNAs (data not shown).

1143 Exons 2 and 3 are non-coding exons whose inclusion in transcripts other than *vif* and *vpr*
1144 has no known function. We found that they were included in other messages infrequently,
1145 each in ~7–8% of transcripts in the ~2 kb completely spliced class of transcripts and 5%
1146 of partially spliced transcripts accumulating in T cells. This is consistent with previous
1147 measurements in the partially spliced class but much lower than has been estimated for
1148 completely spliced transcripts in HeLa cells, suggesting cell-type-specific splicing patterns

1149 may influence inclusion of these exons³⁰⁶.

1150 **3.4.4 A novel ~1 kb class of completely spliced transcripts**

1151 Primers placed near the 5'- and 3'-ends of the HIV_{89.6} genome amplified a second class of
1152 completely spliced transcripts ~1 kb in length. In place of A7, these transcripts use a set of
1153 little studied splice acceptors located ~800 bp downstream within the 3'-TR. Two groups
1154 have previously observed splicing from D1 to acceptors A8, A8a and A8b in this region,
1155 yielding messages of this size class in patient samples; however, none of these could be
1156 translated to a protein of significant length^{449,453}. We determined the complete structure of
1157 29 members of the 1-kb class (Figure 3.2 and Supplementary Table S5). The most abundant
1158 messages observed in this class use the novel acceptor A8c to define their terminal exon. For
1159 HIV89.6, acceptor A8c was used nearly as frequently as A7, which gives us the 2-kb class
1160 of transcripts (Supplementary Table S3), and this was supported by end point RT-PCR
1161 analysis (data not shown).

1162 Acceptor A8c is not well conserved in HIV-1/SIVcpz (14%), although it is conserved in clade
1163 G viruses (> 95%) and most HIV-2/SIVsmm genomes (86%)⁴⁷⁷. This is due to the poor
1164 conservation of an adenine at the wobble base position of the 123rd codon (proline) of the
1165 Nef reading frame, which creates the AG dinucleotide generally required at splice acceptors.
1166 Since any base at this position would code for proline, there does not seem to be strong
1167 selection for a splice acceptor here. However, A8c is displaced from nearby well-conserved
1168 (> 90%) cryptic acceptors A8a and A8b by multiples of 3 bp (12 and 21 bp, respectively),
1169 so splicing to any of these three acceptors would create similar ORFs. All HIVs and SIVs
1170 maintain at least one of these three acceptors, suggesting possible function⁴⁷⁷. We confirmed
1171 that the 1 kb transcripts using A8a, A8b and A8c were present in infected HOS and T cells
1172 by end point RT-PCR using additional primer pairs and by Sanger sequencing of cloned
1173 transcripts (Figure 3.3A and B; data not shown).

1174 The 1-kb transcript containing exons 1, 4 and 8c (1-4-8c, where exon 8c begins at A8c

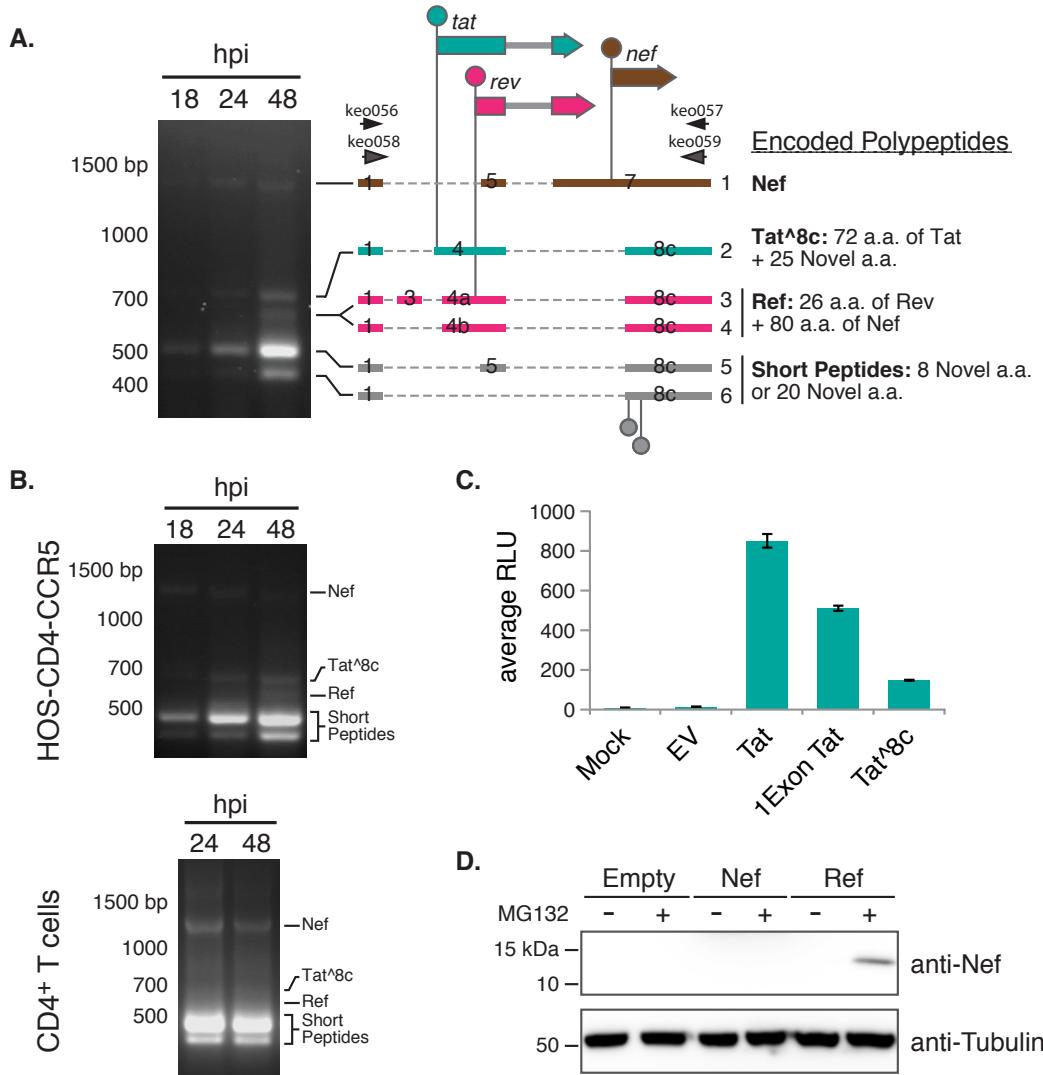


Figure 3.3: HIV_{89.6} transcripts were amplified by RT-PCR using RNA from infected HOS-CD4-CCR5 cells with primers keo056 and keo057. Major bands detected after gel electrophoresis were cloned from the 48 hpi sample and message structures determined by Sanger sequencing. Thick bars represent exons and dashed lines excised introns. Genes are shown above (not to scale) with start codons indicated by circles. Messages 1, 2, 4 and 5 were cloned into expression plasmids for activity assays. (B) Confirmation of presence of the ~1 kb message RNAs in HOS-CD4-CCR5 and primary CD4⁺ T cells (human donor 1, harvested 24 and 48 hpi). An independent primer pair (keo058 and keo059) was used to amplify transcripts by RT-PCR. (C) Tat activity was measured in Tzm-bl cells as Tat-dependent luciferase production after transient transfection with expression plasmids. (D) Western blot showing expression of protein of the predicted size for Ref (12.5 kb) in cells transfected with the Ref expression construct and treated with proteasome inhibitor MG132, detected by an antibody recognizing the carboxy-terminus of Nef. Expression plasmid encoding Nef was included to control for possible expression of partial Nef peptides or breakdown products from the Nef ORF.

and extends to the poly-adenylation site) encodes the first exon of Tat followed by 25 novel amino acids (termed Tat^{8c}). Tat^{8c} showed activity when overexpressed in cells containing a Tat reporter construct (Figure 3.3C, nucleotide and amino acid sequences in Supplementary Table S6). Transcripts with exon structures 1-4a/b/c-8c encode a novel fusion of the amino-terminal 26 amino acids of Rev and the carboxy-terminal 80 amino acids of Nef, hereafter referred to as Ref. We did not detect Rev activity on overexpression of the *ref* transcript, and Ref did not appear to interfere with the normal function of Rev or with HIV replication (Supplementary Figure S2). Ref was detectable by western blot using antibodies targeting the C terminus of Nef after inhibition of the proteasome, suggesting that the fusion is expressed but not stable (Figure 3.3D). Thus, Ref has the potential to encode a new epitope potentially relevant in immune detection of HIV. The transcripts with exon structures 1-5-8c and 1-8c encode at most a short peptide, and so are candidates for acting as regulatory RNAs.

3.4.5 Temporal dynamics of transcript populations

To assess longitudinal variation, we investigated HIV_{89.6} transcript populations during the course of a single round of infection in HOS-CD4-CCR5 cells. A sensitive method for comparison among conditions involves quantifying utilization of six mutually exclusive splice acceptors A3, A4c, A4a, A4b, A5 and a novel acceptor just downstream of A5 termed A5a. Splicing at these acceptors determines the relative levels of messages encoding Tat and Env/Vpu in the partially spliced class and messages encoding Tat, Rev and Nef in the completely spliced class.

We observed longitudinal changes in the levels of these messages in HOS cells over 12–48 h that were statistically significant ($p < 10^{-10}$; generalized linear model described in Supplementary Report S2). This pattern was especially evident in junctions involving donor 1 spliced to each of these acceptors (Figure 3.4A). Most dramatically, transcripts with splicing junctions between D1 and A3 (tat messages) increased with time ($p < 10^{-10}$), while D1^{8c}A4b junctions (used in *env/vpu* or *rev* messages) were used reciprocally less ($p < 10^{-10}$). Such

1202 kinetic changes affecting specific transcripts both with and without the Rev-response element
1203 cannot be explained by the accumulation of Rev, and they may reflect differential transcript
1204 stability or HIV-induced alterations to the host splicing machinery. Temporal changes in
1205 HOS cells were confirmed using end point RT-PCR and analysis after electrophoresis on
1206 ethidium-stained gels (Figure 3.4B).

1207 **3.4.6 Cell-type-specific splicing patterns**

1208 We also compared splicing between T cells and HOS cells and found significant cell type
1209 differences ($p < 10^{-10}$). For example, while transcripts with D1^A5 junctions were dominant
1210 in both cell types, messages using the D1^A4c splice junction (encoding Env/Vpu or Rev)
1211 made up the bulk of the remaining transcripts in T cells but were a minor species in
1212 HOS-CD4-CCR5 cells. Likewise, Tat messages (using A3), which were quite abundant in
1213 HOS cells at all time points, contributed relatively little to populations of transcripts in
1214 primary T cells harvested at 48 hpi (Figure 3.4A). We also used end point PCR and analysis
1215 on ethidium-bromide-stained gels to confirm that the relative ratios of transcripts containing
1216 junctions to A3, A4a, A4b and A4c were different in HOS and T cells (Figure 3.4B).

1217 **3.4.7 Human variation in HIV-1 splicing**

1218 Quantitative comparisons also revealed modest differences in splicing between primary CD4⁺
1219 T cells isolated from different human donors that were statistically significant ($p < 10^{-10}$)
1220 under a generalized linear model (Figure 3.4A). The magnitudes of predicted differences
1221 were small, all < 33% and most < 10%.

1222 **3.5 Discussion**

1223 Use of single-molecule enrichment and long-read single-molecule sequencing has made possible
1224 the most complete study to date of the composition of HIV-1 message populations, revealing
1225 several new layers of regulation. Studies of the low-passage HIV89.6 isolate in a relevant cell
1226 type showed numerous differences from studies of lab-adapted HIV strains in transformed

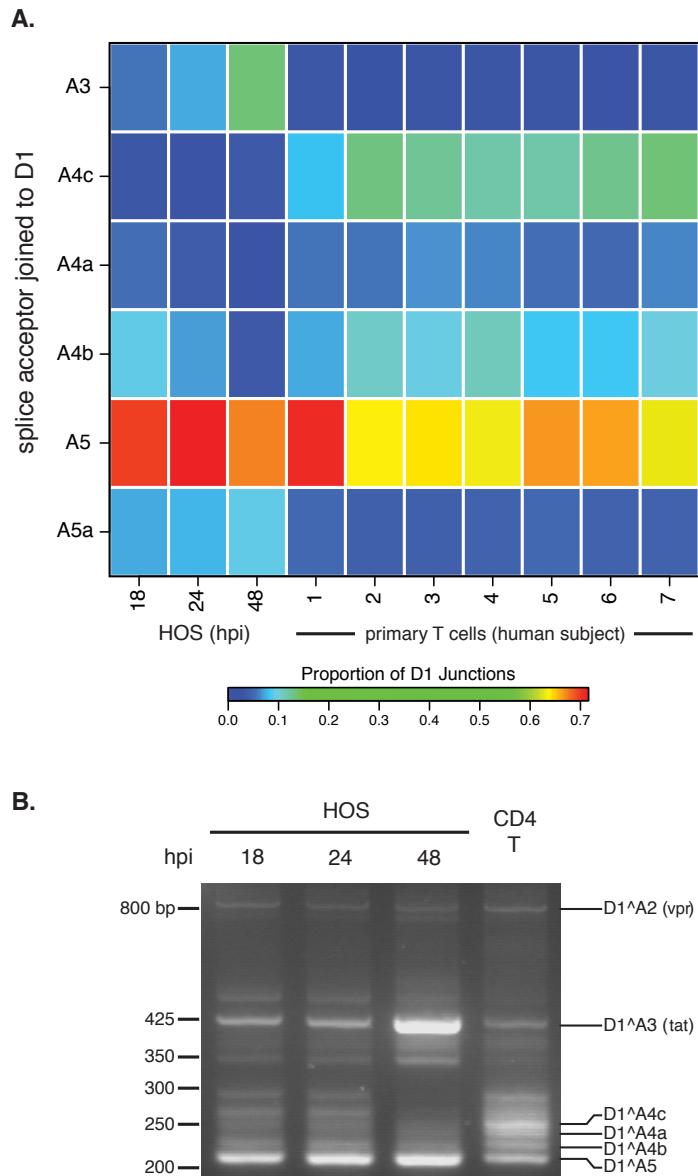


Figure 3.4: Temporal, cell type and donor variability in accumulation of HIV-1 messages. (A) In order to highlight changes in ratios of HIV-1 transcripts accumulating over time during infection and between HOS-CD4-CCR5 cells and primary T cells, we used PacBio read counts to calculate proportions of transcripts with splicing from the first major splice donor, D1, to each of the mutually exclusive acceptors: A3, A4c, A4a, A4c, A5 and the novel putative acceptor A5a. The heat map shows average data for T cell and HOS cell samples in columns with the color tiles indicating the proportion of D1 splicing to each of the mutually exclusive acceptors (rows), according to the color scale shown. (B) Reverse transcription and bulk PCR amplification of HIV_{89.6} transcripts from HOS cells and primary T cells from one human subject (subject 3) resolved by agarose gel electrophoresis and stained with ethidium bromide verified temporal and cell type changes shown in (A).

1227 cell lines, highlighting the importance of studying the most relevant models. These data
1228 also illustrate the limitations of gel-based assays for studying HIV-1 message population.
1229 Multiple different combinations of HIV-1 exons yield mRNAs of similar sizes that are easily
1230 confused in typical assays using gel electrophoresis. Thus, in many settings the more detailed
1231 information provided by single-molecule amplification and single-molecule DNA sequencing
1232 is more useful.

1233 Using these methods, we have detected significant variations between HIV message pop-
1234 ulations generated in T cells from different human donors. The differences were modest
1235 compared to those observed between cell types or time points, perhaps not surprisingly
1236 since any human polymorphisms strongly affecting mRNA processing might interfere with
1237 normal gene expression. However, because tight calibration of message levels is important to
1238 HIV-1, the observed differences in message ratios might affect HIV-1 acquisition or disease
1239 progression. The variation in observed transcripts could also be affected by different kinetics
1240 of infection in T cells from the different donors. In either case, these data suggest that human
1241 polymorphisms may exist that affect HIV-1 message populations in infected individuals,
1242 providing a new candidate mechanism connecting human genetic variation with measures of
1243 HIV disease.

1244 Sequences from the 89.6 viral strain revealed a class of small (~1 kb) completely spliced
1245 transcripts, most contributed by splicing to a new poorly conserved acceptor A8c. These
1246 encoded two new proteins, one of which had Tat activity, and we showed that another, a
1247 Rev-Nef fusion termed Ref, could be detected in cells. HIV_{89.6} is a particularly cytotoxic virus
1248 isolated from the CSF of a patient, and it forms unusually large syncitia in macrophages⁴⁶⁵.
1249 The abundance of 1-kb transcripts produced by this virus provides a possible explanation
1250 for its unique properties. In addition to the novel acceptor A8c, we have also identified 3
1251 putative novel splice donors and 11 putative novel acceptors, which require further studied
1252 to clarify possible functions.

1253 The wealth of new messages found here in HIV_{89.6} and in other HIV-1 isolates suggests there

may be ongoing evolution of novel splice sites and new ORFs. Because splice acceptors in HIV-1 are weak²⁸⁵, mutations creating sequences that even slightly resemble the 3' splice site consensus may be occasionally recruited as novel acceptors, creating new mRNAs. In fact, new splice signals may evolve with relative ease—it has been estimated that reasonable matches to the consensus for splice donors, acceptors and branch-point sites occur within random sequence every 290, 490 and 24 bp, respectively⁴⁷⁸, though sequence substitutions in HIV are usually also constrained by overlapping viral coding regions. We and others have observed appearance of novel exons within the major HIV-1 introns^{308,450,451}. Such long stretches of RNA relatively devoid of competing splice sites may be particularly poised to evolve new signals. On the other hand, most of the putative novel splice acceptors we observed clustered near previously identified acceptors in HIV-1, suggesting that conserved cis-acting splicing signals may recruit factors that act promiscuously on new nearby sequences. Clusters of splice sites might also provide redundancies that protect vital messages, as suggested previously^{479,480}. Frequent evolution of new splice sites may allow viruses to test out new combinations of exons, potentially yielding new RNAs and proteins, like those reported here. However, such novelty must compete with immune constraints—unstable novel polypeptides like Ref can be targeted to the proteasome and presented on MHC molecules as new epitopes for immune recognition.

HIV has likely evolved to produce calibrated message populations in T cells which seem to be altered with relative ease, as in infection in HOS cells, suggesting that therapeutic disruption of correct splicing may be feasible. A few studies have begun to explore small molecule therapy to disrupt HIV-1 splicing^{293,455}. Several factors could be responsible for the differences we observed between HOS and T cells, including hnRNP A/B and H, SC35, SF2/ASF and SRp40^{342,481}. Inhibition of SF2/ASF has already been shown to abrogate HIV-1 replication *in vitro*²⁹³. Thus the lability seen here for function of these factors suggests they may be attractive antiretroviral targets.

1280 **3.6 Acknowledgements**

1281 We would like to thank the University of Pennsylvania Center for AIDS Research (CFAR) for
1282 preparation of viral stocks and isolation of primary CD4⁺ T cells; James A. Hoxie, Ronald
1283 G. Collman, Jianxin You, Robert W. Doms, Paul Bates, David Rekosh and members of the
1284 Bushman laboratory for reagents, helpful discussion and technical expertise.

1285 **CHAPTER 4: Gene activity in primary T cells infected with HIV_{89.6}:**

1286 **intron retention and induction of distinctive genomic**

1287 **repeats**

This chapter is under review as:

S Sherrill-Mix, K Ocieja and F Bushman. Under Review.
Gene activity in primary T cells infected with HIV89.6: in-
tron retention and induction of distinctive genomic repeats.
Retrovirology

1288

KE Ocieja performed the infections and sequencing. I analyzed the data.
KE Ocieja, FD Bushman and I planned the overall study. I produced
the figures. FD Bushman and I wrote the paper.

1289 **4.1 Abstract**

1290 Background: HIV infection has been reported to alter cellular gene activity, but published
1291 studies have commonly assayed transformed cell lines and lab-adapted HIV strains, yielding
1292 inconsistent results. Here we carried out a deep RNA-Seq analysis of primary human T cells
1293 infected with the low passage HIV isolate HIV_{89.6}.

1294 Results: Seventeen percent of cellular genes showed altered activity 48 hours after infection.
1295 In a meta-analysis including four other studies, our data differed from studies of transcription
1296 after HIV infection of cell lines but showed more parallels with infections of primary cells.
1297 We found a global trend toward retention of introns after infection, suggestive of a novel
1298 cellular response to infection. HIV_{89.6} infection was also associated with activation of human
1299 endogenous retroviruses (HERVs) and several retrotransposons, of interest as possible novel
1300 antigens that could serve as vaccine targets. The most highly activated group of HERVs
1301 was a subset of the ERV-9, a group not reported previously to be induced by HIV. Analysis
1302 showed that activation was associated with a particular variant of an ERV-9 long terminal
1303 repeat that contains an indel near the U3-R border. These data also allowed quantification of
1304 >70 splice forms of the HIV_{89.6} RNA and specified the main types of chimeric HIV_{89.6}-host

1305 RNAs. Comparison to 147,281 integration site sequences from the same infected cells allowed
1306 quantification of authentic versus artifactual chimeric reads (0.1% of the total), showing
1307 that 5' read-in, splicing out of HIV_{89.6} from the D4 donor and 3' read-through were the most
1308 common HIV_{89.6}-host cell chimeric RNA forms.

1309 Conclusions: Analysis of RNA abundance after infection of primary T cells with the low
1310 passage HIV_{89.6} isolate disclosed multiple novel features of HIV-host interactions, notably
1311 intron retention and induction of transcription of distinctive retrotransposons and endogenous
1312 retroviruses.

1313 4.2 Background

1314 HIV replication requires integration of a cDNA copy of the viral RNA genome into cellular
1315 chromosomes, followed by transcription and splicing to yield viral mRNA. Alternative
1316 splicing allows the small 9.1 kb HIV genome to generate at least 108 mRNA transcripts
1317 encoding at least 9 proteins and polyproteins^{285,306,445,452,483,484}. During replication, HIV
1318 also reprograms cellular transcription and splicing. For example, the virus-encoded Vpr
1319 protein arrests the cell cycle^{225,227,228,230} and the viral Tat protein binds to P-TEFb and
1320 alters transcript at the HIV promoter and some cellular promoters^{485–490}.

1321 Multiple studies suggest that cells detect HIV infection and respond by inducing inter-
1322 feron-regulated, apoptotic and stress response pathways^{319,491–498}. Several studies have also
1323 suggested that HIV infection disrupts normal cellular splicing pathways^{281,498}. However,
1324 results have varied with many experimental parameters, including target cell type, HIV
1325 isolate and the duration of infection. Many of the published studies focused on infections
1326 with lab-adapted HIV strains in transformed cell lines^{317,319,358,491,498,499}, and so results
1327 may not be fully reflective of infections in patients.

1328 In this study, we sought to generate data more resembling HIV replication in patients
1329 by analyzing transcriptional responses after infection of primary T cells with HIV_{89.6}, a
1330 low passage patient isolate⁴⁶⁵. This represents a continuation of a long term effort to

understand HIV-host cell interactions at the transcriptional level that began with analysis of transcription by HIV_{89.6} in primary T cells using Pacific Biosciences long read single molecule sequencing⁴⁴⁵. Our strategy here was to analyze a single time after infection in depth, analyzing over 1 billion sequence reads from HIV_{89.6} infected and uninfected host cells. These data were then combined with 147,281 unique integration site sequences from the same infections and the Pacific Biosciences data on HIV_{89.6} transcription to 1) elucidate effects of HIV infection on host cell mRNA abundances and splicing, 2) characterize viral message structure in detail and 3) probe the nature of the chimeras formed between host cell and viral RNAs.

4.3 Methods

4.3.1 Cell culture and viral infections

HIV_{89.6} stocks were generated by the University of Pennsylvania Center for Aids Research. 293T cells were transfected with a plasmid encoding an HIV_{89.6} provirus, and harvested virus was passaged in SupT1 cells once. Viral stocks were quantified by measuring p24 antigen content. Primary CD4⁺ T cells were isolated by the University of Pennsylvania Center for AIDS research Immunology Core from apheresis product from a single healthy male donor (ND365) using the RosetteSep Human CD4⁺ T Cell Enrichment Cocktail (StemCell Technologies).

T cells were stimulated for 3 days at 0.5×10^6 cells per milliliter in R10 media (RPMI 1640 with GlutaMAX (Invitrogen) supplemented with 10% FBS (Sigma-Aldrich) with 100 units U/mL recombinant IL2 (Novartis) + 5 μ g/mL PHA-L (Sigma-Aldrich)). Cells were infected in triplicate and mock infections were performed in duplicate. For each infection, 6.6×10^6 cells were mixed with 1.32 μ g HIV_{89.6} in a total volume of 2.25 mL. Infection mixtures were split into three wells of a 6 well plate for spinoculation at 1200 g for 2 hr at 37°C. Cells were incubated an additional 2 hr at 37°C. Cells were then pooled into flasks and volume was increased to a total of 12 mL. Spreading infection was allowed to proceed 48 hr at 37°C,

1357 after which cells were harvested. 1×10^6 cells were harvested for flow cytometry, and 6×10^6
1358 cells were pelleted following two washes in PBS for nucleic acid extraction. Genomic DNA
1359 and total RNA were isolated from 6×10^6 T cells per infection using the AllPrep DNA/RNA
1360 Mini Kit (Qiagen) with Qiashredder columns (Qiagen) for homogenization according to the
1361 manufacturer's instructions. DNA was eluted in 140 μL elution buffer. RNA samples were
1362 treated with DNase prior to elution in 40 μL water.

1363 **4.3.2 Analysis of HIV_{89.6} integration sites in primary T cells**

1364 Integration site sequences were determined for DNA fractions from the above infections
1365 after ligation mediated PCR⁴²³. A total of 147,281 unique integration site sequences were
1366 determined. An analysis of integration site distributions for these samples was reported in
1367 Berry et al.⁴²³.

1368 **4.3.3 mRNA sequencing**

1369 Messenger RNA was isolated and amplified from purified total cellular RNA (3 μL or
1370 approximately 9 μg from each uninfected sample, 25 μL or approximately 3 μg from each
1371 infected sample) using the Illumina TruSeq RNA sample preparation kit according to
1372 manufacturer's protocol. SuperScript III (Invitrogen) was used for reverse transcription.
1373 Each sample was tagged with a separate barcode and sequenced on an Illumina HiSeq 2000
1374 using 100-bp paired-end chemistry.

1375 **4.3.4 Flow cytometry**

1376 To assess percent infected cells, 1×10^6 cells per infection were stained for flow cytometry.
1377 All staining incubations were at room temperature. Cells were first washed in PBS and
1378 then twice in FACS wash buffer (PBS, 2.5% FBS, 2 mM EDTA). Cells were fixed and
1379 permeabilized with CytoFix/CytoPerm (BD) for 20 minutes and washed with Perm-Wash
1380 Buffer (BD) before staining with anti-HIV-Gag-PE (Beckman Coulter) for 60 min. Finally
1381 cells were washed in FACS wash buffer and resuspended in 3% PFA. Samples were run

1382 on a LSRII (BD) and analyzed with FlowJo 8.8.6 (Treestar). Cells were gated as follows:
1383 lymphocytes (SSC-A by FSC-A), then singlets (FSC-A by FSC-H), then by Gag expression
1384 (FSC-A by Gag).

1385 **4.3.5 Analysis**

1386 Reads were aligned to the human genome using a combination of BLAT⁴²² and Bowtie⁵⁰⁰
1387 through the Rum pipeline⁵⁰¹. Estimates of fragments per kilobase of transcript per million
1388 mapped reads and changes in expression for cellular genes were calculated by Cufflinks³⁵⁵.
1389 Reads found to contain sequence similar to the HIV genome using a suffix tree algorithm were
1390 aligned against the HIV_{89.6} genome using BLAT⁴²². All statistical analyses were performed
1391 in R 3.1.2⁴¹⁵. RNA-Seq reads from Chang et al.³¹⁹ were downloaded from the Sequence
1392 Read Archive (SRP013224) and aligned using the Rum pipeline.

1393 Gene lists were obtained from the supplementary materials of four other studies of differential
1394 gene expression during HIV infection^{319,358,496,502}. We called genes differentially expressed
1395 in Li et al.⁵⁰² if they had a reported $p < 0.01$ or in Lefebvre et al.³⁵⁸, Chang et al.³¹⁹
1396 and Imbeault et al.⁴⁹⁶ if they had an adjusted $p < 0.05$. We called genes as differentially
1397 expressed in our own study if the adjusted $p < 0.01$. For the comparison of differentially
1398 expressed genes regardless of direction in figure 4.1 (below the diagonal), it was unclear
1399 exactly how many genes were studied in each study so we assumed a background of the
1400 14,192 genes (the number of genes which could be tested for significance in our data).

1401 We obtained transcriptional profiles comparing immune cell subsets from the Molecular
1402 Signatures Database⁵⁰³. MSigDB set names from the MSigDB used in Figure 4.2A were
1403 GSE10325 LUPUS CD4 TCELL VS LUPUS BCELL, GSE10325 CD4 TCELL VS MYELOID,
1404 GSE10325 CD4 TCELL VS BCELL, GSE10325 LUPUS CD4 TCELL VS LUPUS MYELOID,
1405 GSE3982 MEMORY CD4 TCELL VS TH1, GSE22886 CD4 TCELL VS BCELL NAIVE,
1406 GSE11057 CD4 CENT MEM VS PBMC, GSE11057 CD4 EFF MEM VS PBMC, GSE3982
1407 MEMORY CD4 TCELL VS TH2 and GSE11057 PBMC VS MEM CD4 TCELL and in

1408 Figure 4.2B were GSE36476 CTRL VS TSST ACT 72H MEMORY CD4 TCELL OLD,
1409 GSE10325 CD4 TCELL VS LUPUS CD4 TCELL, GSE22886 NAIVE CD4 TCELL VS 12H
1410 ACT TH1, GSE3982 CENT MEMORY CD4 TCELL VS TH1, GSE17974 CTRL VS ACT
1411 IL4 AND ANTI IL12 48H CD4 TCELL, GSE24634 IL4 VS CTRL TREATED NAIVE CD4
1412 TCELL DAY5, GSE24634 NAIVE CD4 TCELL VS DAY10 IL4 CONV TREG, GSE1460
1413 CD4 THYMOCYTE VS THYMIC STROMAL CELL and GSE1460 INTRATHYMIC T
1414 PROGENITOR VS NAIVE CD4 TCELL ADULT BLOOD.

1415 We downloaded the RepeatMasker track from the UCSC genome browser⁵⁰⁴ and used the
1416 SAMtools library⁵⁰⁵ to assign reads to the repeat regions. HERV-K age estimates were
1417 obtained from the supplementary materials of Subramanian et al.⁵⁰⁶.

1418 We used a Bayesian estimate of the ratio of expression in uninfected and HIV infected
1419 samples to account for sampling effort and differing expression in genomic regions. We
1420 modeled the observed counts as a binomial distribution with a flat beta prior ($\alpha = 1, \beta = 1$)
1421 separately for uninfected and infected samples. We then Monte Carlo sampled the two
1422 posterior distribution to estimate the posterior distribution of the ratio. For introns, the
1423 number of binomial successes was set to the number of reads mapped to the intron and the
1424 number of trials was the total number of reads observed in the genes overlapping that intron.
1425 For repeat regions, the number of binomial successes was set to the number of reads mapped
1426 to that region and the number of trials was the total number of reads mapped to the human
1427 genome.

1428 To estimate determinants of LTR12C expression, we fit a logistic regression for which
1429 LTR12C increased in expression with HIV_{89.6} infection (95% Bayesian credible interval
1430 >1) on to characteristics of the LTR12C regions. We extracted all the LTR12C regions
1431 from the human genome and determined the U3-R boundary using a ends free alignment of
1432 the previously reported U3-R border^{507–511} against the sequences. Regions less than 1,000
1433 bases long were discarded. Previous studies disagreed about the location of the LTR12C
1434 transcription start site and it appears that transcription may start in several places^{508,509}.

1435 We took the 5' most site that had agreement between studies (transcription starting with
1436 TGGCAACCC). We split the sequences into short, medium and long length classes based
1437 on an indel about 70 bases upstream from the transcription start site. For each length class,
1438 we generated a consensus sequence and counted the Levenshtein edit distance between the
1439 consensuses and each corresponding sequence. We also counted the number of NFY motifs
1440 (CCAAT or ATTGG), MZF1 motifs (GTGGGGA) and GATA2 motifs (GATA or TATC)
1441 in the entire U3 region or checked in any of the three motifs was present in the 150 bases
1442 upstream of the TSS. A final regression model was selected using stepwise regression with
1443 an AIC cutoff of 5. For display, the LTR12C sequences were aligned with MUSCLE⁵¹².

1444 The abundance of the HIV RNA size classes was estimated as described in Additional File
1445 5. These estimates were then multiplied by the within size class proportions estimated by
1446 Owieja et al.⁴⁴⁵ using PacBio sequencing of HIV_{89.6} to yield proportions over 78 measured
1447 HIV_{89.6} RNAs.

1448 4.4 Results

1449 4.4.1 Infections studied

1450 HIV_{89.6}, a clade B primary clinical isolate⁴⁶⁵, was used to infect primary CD4⁺ T cells from
1451 a single human donor in three replicate infections. For comparison, two additional replicates
1452 from the same donor were mock infected. Samples were harvested after 48 hours of infection,
1453 which allowed for widespread infection in the primary T cell cultures, though some cells may
1454 be infected secondarily by viruses produced in the first round. Thus cultures probably were
1455 not tightly synchronized but did have extensive representation of infected primary T cells.
1456 From these samples, we obtained 1,161,705,678 101-bp reads from primary CD4⁺ T cells
1457 from a single donor; 1,021,207,853 were mapped to the human genome and 24,783,844 to
1458 the HIV_{89.6} provirus (Table 4.1). Below we first discuss the influence of infection on cellular
1459 gene activity and RNA splicing, then analyze HIV RNAs and lastly analyze chimeras formed
1460 between HIV and cellular RNAs.

Sample	Infection rate (%)	Reads	Human reads	HIV reads	% HIV	% HIV in infected
Uninfected-1	—	232,450,106	212,391,460	—	—	—
Uninfected-2	—	235,048,212	203,760,783	—	—	—
Infected-1	37.5	234,378,088	199,871,662	10,219,315	4.86	13.0
Infected-2	26	226,078,422	198,436,507	7,322,556	3.56	13.7
Infected-3	21	233,750,850	205,747,441	7,241,973	3.40	16.2

Table 4.1: Samples used in this study, their infection rates and sequencing depth.

1461 4.4.2 Changes in gene activity in primary T cells upon infection with HIV^{89.6}

1462 Changes in host cell gene expression have been reported during HIV infection^{317–319,358,491–498}
 1463 and differences in expression have been observed associated with the stage⁵⁰² and progres-
 1464 sion⁵¹³ of disease. Here we observed significant changes in gene expression (false discovery
 1465 rate corrected $q < 0.01$) in 3,142 genes, 17.1% of expressed cellular genes (Additional file 1).
 1466 The genes with most extreme increases, all $>6\times$ fold higher, during HIV infection included
 1467 IFI44L, RSAD2, HMOX1, MX1, USP18, IGJ, OAS1, CMPK2, DDX60, IFI44, IFI6, IFNG
 1468 and CCL3. All of these have been reported to be involved in innate immunity⁵¹⁴ or are
 1469 interferon inducible⁵¹⁵, highlighting a strong innate immune response in the cells studied.
 1470 Genes with the largest decreases, all $>3\times$ fold lower, were GNG4, GPA33, IL6R, CCR8,
 1471 RORC, AFF2 and CCR2.

1472 Many gene ontology categories were significantly enriched for differentially expressed genes
 1473 (Additional file 2). Notably upregulated with infection were genes involved in apoptosis,
 1474 immune responses and cytokine production (all $q < 10^{-4}$) and down-regulated were genes
 1475 involved in viral gene expression, nonsense-mediated decay and translation elongation and
 1476 termination (all $q < 10^{-19}$). These changes suggest that the cells responded to HIV infection
 1477 with the induction of inflammatory, interferon regulated and apoptotic responses, patterns
 1478 posited from several previous studies^{319,358,491–497,499,516}. Several genes were activated that
 1479 were characteristic of other hematopoietic lineages, e.g. hemoglobin β , CD8, CD20 and
 1480 CD117, while several CD4 $^+$ T cell specific genes, e.g. CD4 and CD3, were downregulated,
 1481 potentially consistent with de-differentiation of infected and bystander cells. We return to

Cell type	HIV type	Differentially expressed genes (Up/Down)	Study
Primary CD4 ⁺ T	HIV _{89.6}	3393 (1756/1637)	This study
Primary CD4 ⁺ T	NL4-3 BAL-IRES-HSA	228 (182/46)	Imbeault et al. ⁴⁹⁶
Lymph node biopsies	Acute infection	448 (383/65)	Li et al. ⁵⁰²
SupT1	HIV _{LAI}	4997 (2666/2331)	Chang et al. ³¹⁹
SupT1	NL4-3Δenv-eGFP/VSV-G	579 (212/367)	Lefebvre et al. ³⁵⁸

Table 4.2: Data from this study and four others used for meta-analysis of human gene expression changes during HIV infection

1482 this point in the discussion.

1483 **4.4.3 Comparison of transcriptional profiles from HIV_{89.6} infection of primary
1484 T cells to data on HIV infection in other cell types**

1485 We sought to identify the transcriptional responses that were most conserved upon HIV
1486 infection and so collected and analyzed data from four other studies of transcription in
1487 HIV-infected cells (Table 4.2). These included two studies of infection of the SupT1 cell
1488 line^{319,358}, a study of primary CD4⁺ T cells⁴⁹⁶ and a study of lymphatic tissue in acutely
1489 viremic patients⁵⁰². Genes were scored as increased or decreased in activity after infection,
1490 and the amount of agreement was compared among the different studies.

1491 No gene was called as differentially expressed in all five studies. Eight genes were differentially
1492 expressed in the same direction in 4 out of 5 studies; AQP3 and EPHX2 were down-regulated
1493 with HIV infection and CD70, EGR1, FOS, ISG20, RGS16 and SAMD9L were up-regulated.
1494 A full listing is provided in Additional file 4. Several of the up-regulated genes are known to
1495 be interferon inducible, again emphasizing the role of innate immune pathways.

1496 For each pair of studies, we compared whether they agreed on the identities of differentially
1497 expressed genes and whether they agreed on the direction of change (Figure 4.1). The
1498 estimated alterations in gene activity showed notable differences in the responses to infection
1499 in primary cells versus the SupT1 cell line. The two SupT1 studies were significantly similar
1500 ($p < 10^{-15}$) to each other but were not significantly associated (Lefebvre et al.³⁵⁸, $p = 0.2$)
1501 or were negatively associated (Chang et al.³¹⁹, $p = 10^{-7}$) with data from lymphatic tissue

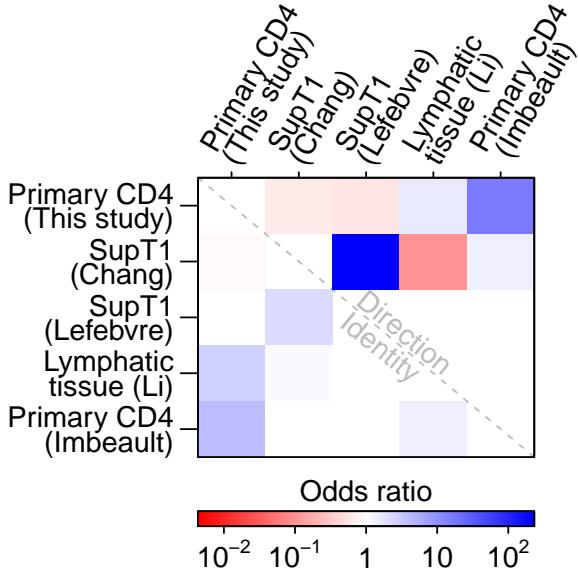


Figure 4.1: Comparisons among studies quantifying cellular gene expression after HIV infection. For each pair of studies, the association between up- and down-regulation calls was measured for genes identified by both studies as differentially expressed (above the diagonal). As another comparison, we also measured the agreement between studies for which genes were called differentially expressed regardless of direction (below the diagonal). The color scale shows the conservative (i.e. closest to 1) boundary of the confidence interval of the odds ratio with blue indicating a positive association and red a negative association between studies. For confidence intervals overlapping 1, the value was set to 1. Therefore all colored squares indicate significant associations.

in acute HIV patients. The primary T cell study reported here was significantly associated with the second study in primary cells ($p < 10^{-15}$) and with a study of lymphatic tissue from patients acutely infected with HIV ($p = 0.003$). Our primary T cell data was negatively associated with the SupT1 studies (both $p < 10^{-3}$). This documents significant differences in responses to HIV infection between infected primary cells and SupT1 cells and suggests that results of infections in primary cells more closely align with actual acute HIV infections in patients. SupT1 cells might be expected to respond to infection differently than primary cells since they have several nonsynonymous mutations in innate immunity genes⁵¹⁷, have blocks in immune signaling pathways⁵¹⁸ and fail to activate many interferon stimulated genes during HIV infection⁴⁹⁷.

4.4.4 Comparison of the HIV infected cell transcriptional profiles to additional experimental T cell profiles

To investigate the transcriptional changes in more depth, we compared the results of the five studies of HIV infection to transcriptional profiles comparing immune cell subsets available at the Molecular Signatures Database (MSigDB)⁵⁰³. The MSigDB reports genes that are

1517 increased or decreased in relative expression for each of 185 pairs of transcriptional profiles
1518 involving CD4⁺ T cells. We compared the lists of affected genes in each pair to genes altered
1519 in activity by HIV infection. Those pairs of studies with the most significant associations
1520 with HIV_{89.6} data are show in Figure 4.2A. For comparison, the associations with the four
1521 other HIV transcriptional profiling studies mentioned above are shown as well.

1522 The most significant associations for our data showed gene expression in HIV_{89.6}-infected
1523 cells moving away from typical T cell expression patterns and towards patterns more similar
1524 to B cells, myeloid cells and bulk peripheral blood mononuclear cells (all Fisher's $p < 10^{-15}$)
1525 (Figure 4.2A). These changes were also seen, although to a lesser extent, in the Imbeault
1526 et al.⁵¹⁹ study which also used primary CD4⁺ T cells.

1527 For comparison, we also extracted those profiles most strongly associated with the transcrip-
1528 tional data on lymphatic tissue of HIV patients⁵⁰². The profiles showed patterns similar to
1529 strongly stimulated T cells, autoimmune disease and to the Th1 T cell subset (all $p < 0.01$)
1530 (Figure 4.2B). Our data in primary CD4⁺ T cells paralleled the changes seen in lymphatic
1531 tissue. These transcriptional changes again highlights the strong immune response generated
1532 by HIV infection in primary cells.

1533 4.4.5 Intron retention

1534 Cells respond to infection by shutting down macromolecular synthesis at multiple levels⁵²⁰⁻⁵²⁴,
1535 so we investigated whether cells also showed perturbations in splicing efficiency after infection.
1536 As a probe, we created a database of cellular genomic regions annotated exclusively as exons
1537 or introns in all spliceforms in the UCSC gene database⁴³³ and quantified expression in
1538 these regions in infected and uninfected cells. We found a significant increase in intronic
1539 sequences relative to exonic sequence (Wilcoxon $p < 10^{-15}$) (Figure 4.3A). This increase
1540 in intronic sequence was reproducible between replicates in our study (Kendall's $\tau=0.42$,
1541 $p < 10^{-15}$) (Figure 4.3B). We reanalyzed RNA-Seq data from Chang et al.³¹⁹ and also
1542 documented intron retention which correlated with the changes seen in our data (Kendall's

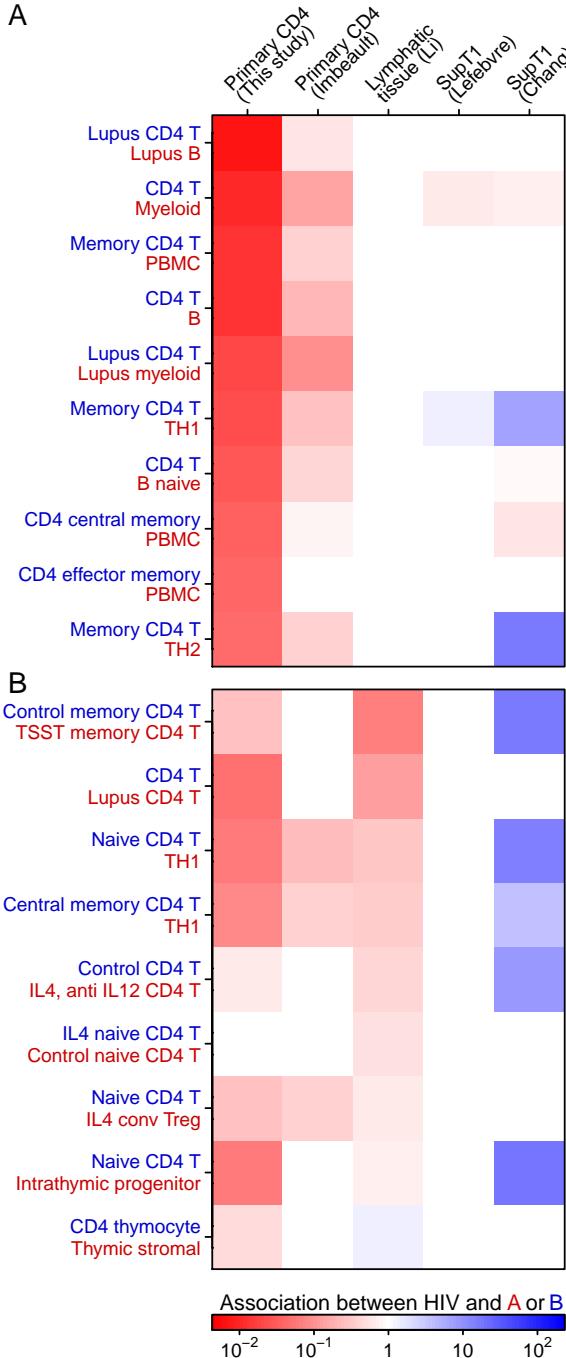


Figure 4.2: Comparisons of the effect of HIV infection on gene expression to studies comparing subsets of immune cells. The MSigDB database was used to extract 185 sets of differentially expressed genes from pairs of transcriptional profiling studies of immune cell subsets involving CD4⁺ T cells. For each pair of studies, we used Fisher's exact test to measure the association between up- and down-regulation calls for genes identified as differentially expressed in both our HIV study and the comparator immune subsets. A) The transcriptional profiles with strongest associations with changes observed in our study of HIV_{89.6} infection of primary T cells. Blue indicates a positive association between changes seen in HIV infected cells and the first immune subset (text colored blue) while red indicates a positive association with the second immune subset (text colored red). The color scale shows the conservative (i.e. closest to 1) boundary of the confidence interval of the odds ratio. For confidence intervals overlapping 1, the value was set to 1. Therefore all colored squares indicate significant associations. B) As in A, but showing the transcriptional profiles most strongly associated with changes observed in lymph node biopsies from acutely infected patients⁵⁰².

1543 $\tau=0.12$, $p < 10^{-15}$) (Figure 4.3C).

1544 A possible artifactual explanation for enrichment of intronic sequences could involve greater
1545 DNA contamination in the infected cells samples. That is, if the relative amount of DNA
1546 differed between treatments, the amount of apparent intronic sequences could also differ
1547 due to sequencing of contaminating DNA. To examine whether DNA contamination was
1548 abundant in our samples, we compiled a collection of 27 large gene desert regions, defined
1549 here as 1) regions outside the centrosome and first and last cytoband, 2) containing less than
1550 1% unknown sequence, 3) containing no genes annotated in UCSC genes⁴³³, 4) containing
1551 no repeats annotated in the repeatMasker database⁴³⁸ and 5) spanning more than 100
1552 kb. No reads were mapped to these 41 Mb of gene deserts in any sample, arguing against
1553 explanations based on DNA contamination. Thus these data indicate that intron retention
1554 was increased in these cell populations upon HIV infection, revealing a previously undisclosed
1555 aspect of the host cell transcriptional response to infection.

1556 Previous studies have reported changes in the expression and localization of splicing factors
1557 with HIV infection^{281,525,526}. In our data, HIV_{89.6} infection significantly altered the expression
1558 of genes involved in RNA splicing ($p = 2 \times 10^{-7}$) and nonsense-mediated decay ($p < 10^{-15}$).
1559 Genes related to nonsense-mediated decay genes showed a strong pattern of lowered RNA
1560 abundance, with 71 out of 118 annotated genes significantly lower in expression after infection.
1561 These patterns suggest potential mechanisms for the intron retention observed here.

1562 **4.4.6 Induction of transcription from HERVs and LINEs by HIV_{89.6} infection**

1563 HIV infection has been reported to induce expression of certain HERVs, particularly HERV-
1564 K⁵²⁷⁻⁵²⁹, and LINE and Alu transposable elements⁵³⁰, providing candidate markers of
1565 infection and possible vaccine targets. Thus we analyzed our data in primary T cells infected
1566 with HIV_{89.6} to investigate the expression of HERVs, LINEs and other repeated sequences.
1567 Figure 4.4A shows a comparison of the association between changes in expression with
1568 HIV_{89.6} infection and the various genomic repeat types over varying levels of differential

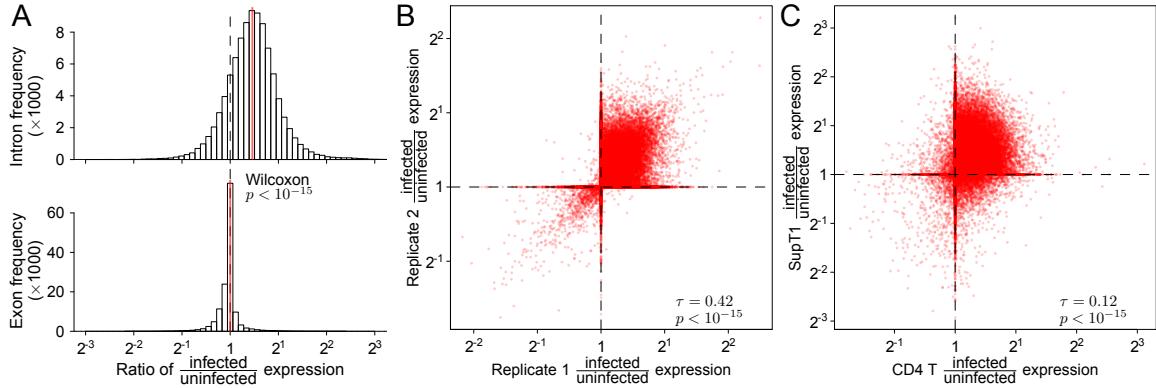


Figure 4.3: Changes in the abundance of intronic regions with HIV infection. Expression of intronic and exonic regions was quantified as the proportion of reads mapping within the intron/exon out of the total reads mapping to the transcription units overlapping that intron/exon. A) Comparison of the ratios of expression between infected and uninfected replicates in exclusively intronic or exonic regions of transcription units. B) Reproducibility of intron retention between replicates. Each point quantifies the change in expression with HIV infection for a specific intronic region. The x-axis shows changes in gene activity accompanying infection for one set of replicates (Infected-1 and Infected-2 vs. Uninfected-1) and the y-axis shows the same data for different replicates (Infected-3 vs. Uninfected-2). C) Reproducibility of intron retention between studies. The plot is arranged as in B but with all data from our study combined on the x-axis and corresponding data from Chang et al.³¹⁹ on the y-axis.

1569 expression. At high levels of expression, ERV-9 (odds ratio at $4\times$ expression: 152, 95%
 1570 CI: 82.5–259) and its long terminal repeat LTR12C (odds ratio at $4\times$ expression: 144, 95%
 1571 CI: 98.2–207) are the only repeats highly associated with upregulation during HIV infection.
 1572 Looking at genomic repeats with any significant increase, the expression of many recently
 1573 acquired genomic repeats, including L1HS, LTR5_Hs (a human specific LTR of HERV-K),
 1574 AluYa5, AluYg6 and SVA_D and SVA_F, were associated with HIV_{89.6} infection (Figure
 1575 4.4B).

1576 We saw a relationship between the age of genomic repeats and its likelihood of being induced
 1577 by HIV_{89.6} infection. The most highly enriched repeats were associated with relatively
 1578 recent hominid-specific repeat classes as annotated by the RepeatMasker database (repeat
 1579 classes with $p < 10^{-50}$ odds ratio: 31.6, 95% CI: 8.88–112). In HERV-K (HML-2), the
 1580 most recently active endogenous retrovirus in the human genome^{506,531,532}, we saw that
 1581 integrations unique to the human genome⁵⁰⁶ were more likely to be differentially expressed

1582 than older HERV-Ks (odds ratio: 5.38, 95% CI: 1.93–16.0).

1583 Previous RNA-Seq studies of cellular expression during HIV infection in transformed cell
1584 lines did not report increases in HERV mRNA^{319,358}. To investigate this difference, we
1585 downloaded and analyzed the RNA-Seq data from Chang et al.³¹⁹, which quantified gene
1586 activity in transformed SupT1 cells infected with a lab-adapted strain of HIV. We found a
1587 much higher level of HERV expression in their data in both HIV infected cells and uninfected
1588 controls than in primary cells (Figure 4.4C). We suspect that in SupT1 cells, as with many
1589 cancerous cells^{533–537}, the baseline expression of transposons and endogenous retroviruses is
1590 higher than in primary cells, masking further induction by HIV infection.

1591 We observed heterogeneous expression among ERV-9/LTR12C sequences and so investigated
1592 the primary sequence determinants. We observed that ERV-9/LTR12C has three variants of
1593 differing length in the U3 region just upstream of the transcription start site (Figure 4.5A),
1594 an important region for transcription initiation⁵⁰⁸. The U3 region of LTR12C also contains
1595 multiple motifs for transcription factors NFY, GATA2 and MZF1⁵¹¹. To clarify factors
1596 affecting expression levels, we counted the number of motifs matching these transcription
1597 factors, assigned each LTR12C to one of the length classes, counted the number of mutations
1598 away from the consensus for that length class and checked for integration in a transcription
1599 unit. We then carried out a regression analysis to test the effects of these variables on
1600 LTR12C differential expression. We found that HIV_{89.6} induced transcription was more
1601 likely with the fewer mutations away from consensus, the number of locations matching the
1602 NFY transcription factor binding motif (CCAAT) and LTRs containing the short length
1603 variant of the 3' U3 region. The presence of a MZF1 motif near the transcription start site
1604 decreased transcription (Figure 4.5B).

1605 4.4.7 HIV mRNA synthesis and splicing

1606 Over 24 million Illumina reads mapped to HIV_{89.6}, yielding an average coverage of over
1607 240,000-fold. Reads mapping to HIV_{89.6} comprised between 3.4–4.8% of mapped reads in

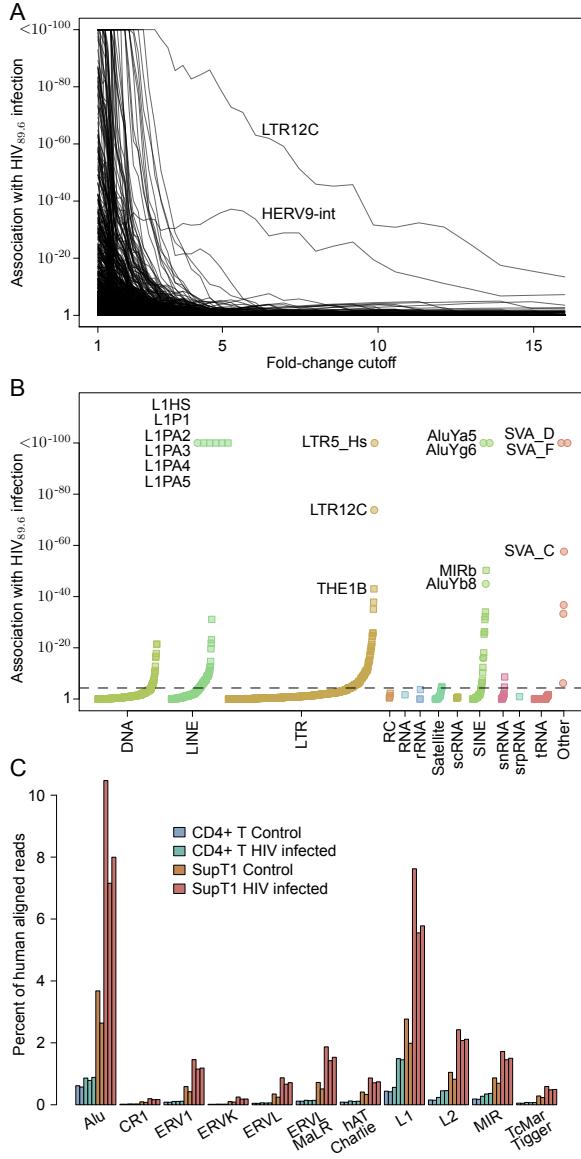


Figure 4.4: Repeat categories enriched upon infection with HIV. A) The association of repeat regions differentially expressed after HIV_{89.6} infection of primary T cells observed for varying thresholds of differential expression. The threshold used to call a gene differentially expressed based on the Bayesian posterior median was varied and Fisher's exact test was used to assess whether any genomic repeats had a significant association with this differential expression. Note that only ERV-9 (annotated as HERV9-int in the RepeatMasker database) and its corresponding long terminal repeat LTR12C were significantly associated with large changes in expression. B) Enrichment of repeat categories in regions differentially expressed (Bayesian 95% credible interval >1) between HIV-infected and control CD4⁺ T cells. The repeated sequences are ordered on the x-axis by the extent of induction within each class, the y-axis shows the p-value for upregulation after infection. The dashed line indicates a Bonferroni corrected p value of 0.05. (C) The proportion of human mapped reads that align within classes of genomic repeats for data from primary CD4⁺ T cells from this study and SupT1 cells from Chang et al.³¹⁹. A single read mapping multiple times to a given category was only counted once.

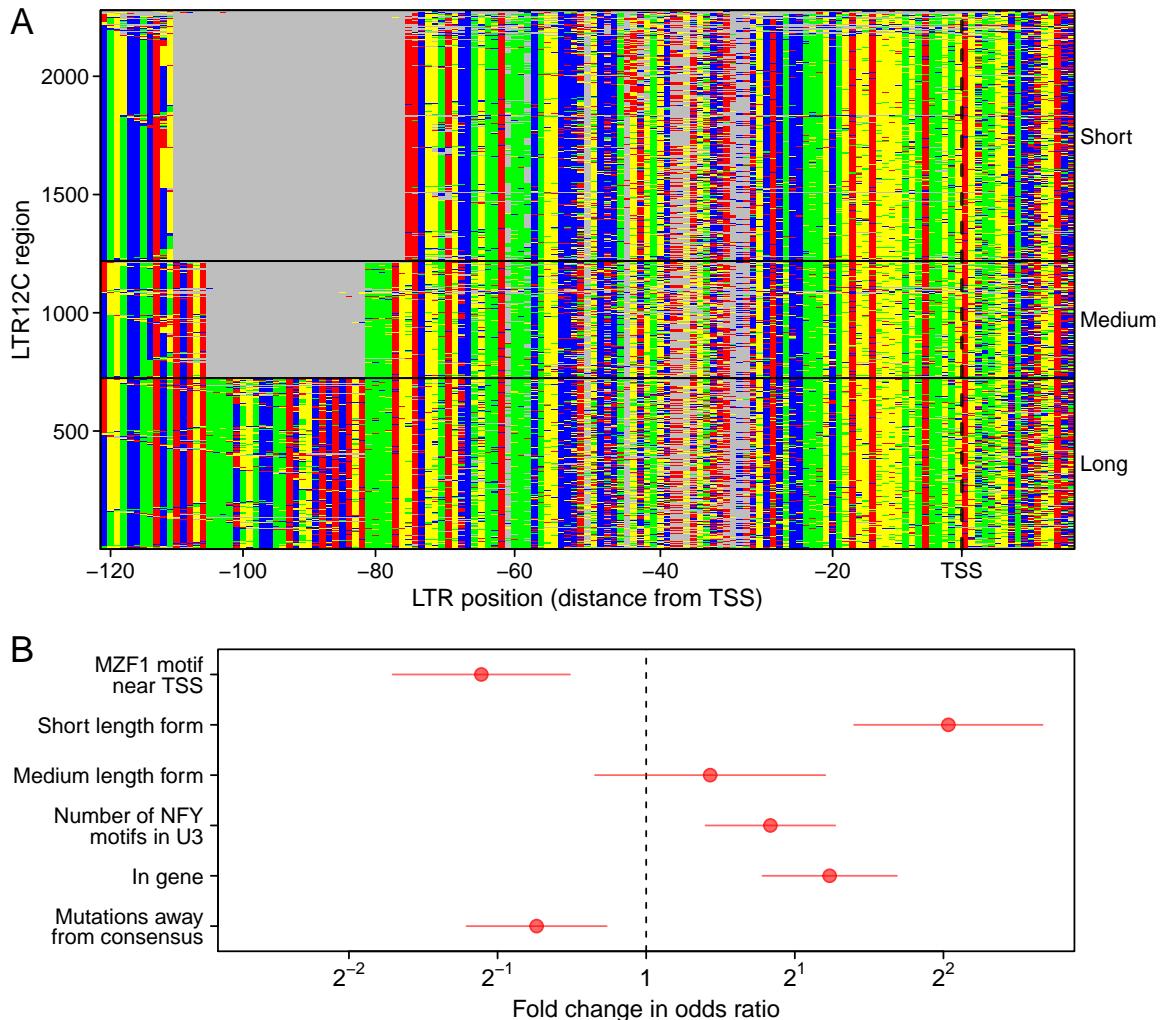


Figure 4.5: Characteristics of LTR12C sequences associated with induction upon infection of primary T cells with HIV_{89.6}. A) An alignment of the 3' end of the U3 region of repeats annotated as ERV-9 LTR12C. Each row is a LTR sequence and each column a base in that sequence colored by nucleotide identity. Three distinct classes are visible with a short, medium and long form. Mutations away from the consensus can also be seen. B) The coefficients (points) and ± 1.96 standard errors (horizontal lines) of a logistic regression comparing differential expression of LTR12C to the presence of MZF1 and NFY motifs, short/medium/long length alternate forms of the U3-R region, mutations away from the consensus for each length form and integration inside a transcription unit. The coefficient shown for mutations away from consensus is for a 10 mutation difference and the coefficient shown for NFY motifs is for a change of 5 additional motifs. All other coefficients are for binary values.

1608 the infected samples (Table 4.1). Assuming HIV-infected cells contain the same amount of
1609 mRNA as uninfected cells and adjusting for rates of infection ranging between 21–37.5%
1610 (Table 4.1), we estimate that HIV transcripts comprise between 13.0–16.2% of the total
1611 polyadenylated mRNA nucleotides in infected cells 48 hours after initial infection. This
1612 parallels previous estimates of around 10%⁵³⁸ at 48 hours postinfection, 38% at 24 hours³¹⁹
1613 or 30% after 72 hours⁴⁹¹.

1614 Over 47,257 single reads spanned previously reported HIV splice junctions, allowing a
1615 quantitative assessment of donor and acceptor utilization (Figure 4.7A). As expected from
1616 previous studies^{306,445}, the most abundant junctions were D1-A5 and D4-A7. We confirmed
1617 the use of unusual splice acceptors A8c and A5a, previously reported in HIV_{89.6}⁴⁴⁵. In
1618 our data, we also see a higher abundance of D1-A1 and D1-A2 splice junctions than might
1619 be expected^{306,445}, although previous studies reported proportional abundance within size
1620 classes, making comparisons between size classes uncertain.

1621 A 3' bias is apparent in our sequencing data (Additional file 5). This could be due to the
1622 poly-A capture step of the protocol where any break in the RNA would result in distal
1623 5' sequences being lost⁵³⁹. We used sequence reads from the large unspliced HIV intron
1624 1 to measure this bias using a regression of the log of the number of fragments with a
1625 5'-most end starting at a given position against the distance of that position from the
1626 viral polyadenylation site, yielding an estimated probability of breakage of 0.021% per base
1627 (Additional file 5). Given this rate of termination, there is only a 14% chance of reaching
1628 the 5' end of the 9171 nt unspliced HIV genome $((1 - 0.00021)^{9171})$.

1629 Ocwieja et al.⁴⁴⁵ determined the relative abundance of HIV_{89.6} of similarly sized transcripts
1630 using PacBio single molecule sequencing, but were not able to estimate the relative abundance
1631 of all transcripts due to a sequencing bias favoring shorter transcripts. For this reason,
1632 relative abundances could only be specified within message size classes (i.e. the 4 kb, 2 kb
1633 and unexpectedly a 1 kb size class as well) and the overall quantitative abundances were
1634 unknown. The RNA-Seq data reported here are unable to determine complete transcript

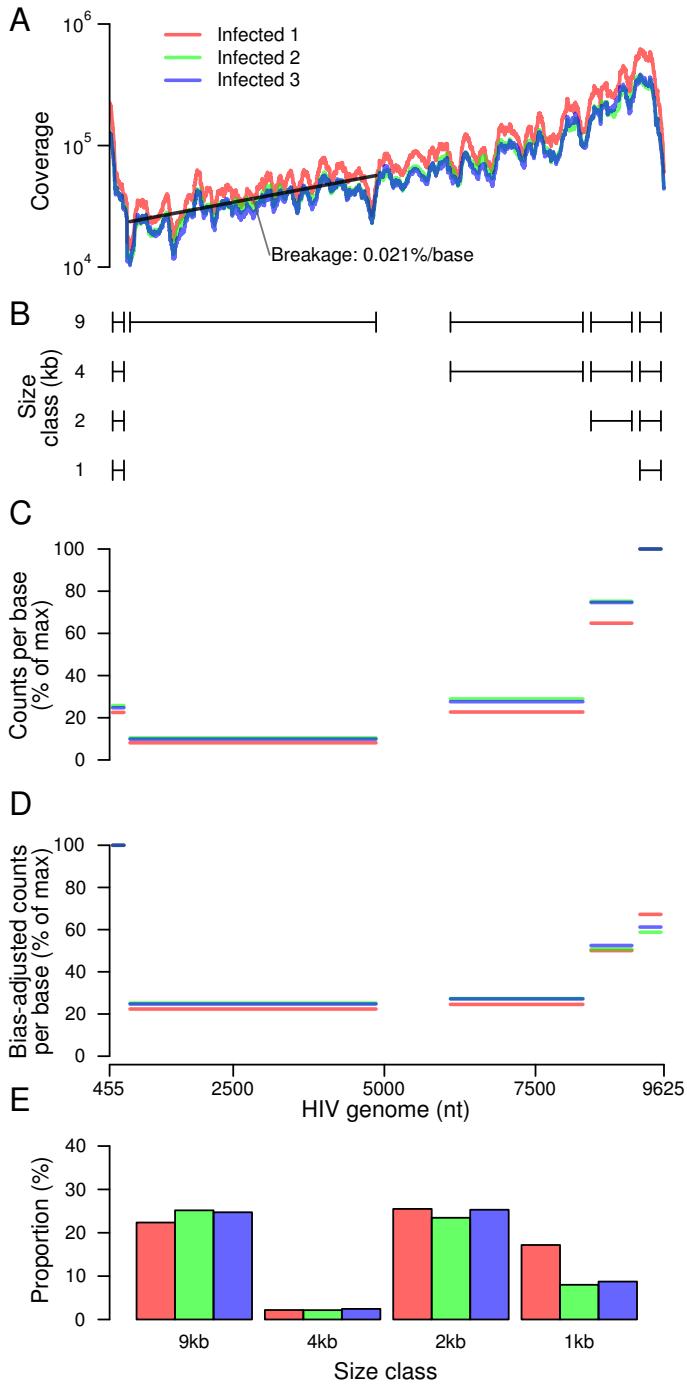


Figure 4.6: Estimating relative abundance of HIV_{89.6} message size classes using RNA-Seq data.

A) RNA-Seq coverage of the HIV_{89.6} genome for the replicates in this study. Each replicate is indicated by a different color. The HIV genome is shown on the x-axis and the number of reads that aligned to each position is shown on the y-axis. Black line indicates the 0.021% coverage decrease per base distance from the 3' end of the mRNA estimated from a least squares fit on the read counts in the first intron.

B) Diagram of the segments of the HIV_{89.6} RNA present in each of 9 kb, 4 kb, 2 kb and 1 kb size class.

C) The proportion of reads mapped to each of the segments of the HIV_{89.6} genome shown in B adjusted by the length of the segment. Each replicate is shown by a different color.

D) Corrected representation of RNA segments from the different size classes. Because cDNA synthesis was primed from the polyA tail, more 3' sequences are recovered preferentially. Using the bias estimate from A, we adjusted each genome segment by the inverse of the bias predicted based on its distance from the 3' end of the mRNA. Corrected proportions for the indicated RNA segments are shown colored by replicate.

E) The proportion of each size class was inferred using the estimates in D by calculating the difference between segments. Replicates are indicated by color.

1635 abundance because the short read length does not allow reconstruction of multiply spliced
1636 messages but do permit estimation of size class abundances after correcting for 3' bias
1637 (Additional file 5). Thus the PacBio data reported by Ocwieja et al.⁴⁴⁵ and the Illumina
1638 data reported here can be combined together to determine complete relative abundance of
1639 all HIV_{89.6} transcripts (Figure 4.7B).

1640 The most abundant HIV mRNAs were the unspliced HIV genome (37.6%), a transcript
1641 encoding Nef (D1-A5-D4-A7: 15.5%), two 1 kb size class transcripts (D1-A5-D4-A8c: 10.6%,
1642 D1-A8c: 4.9%) and two Rev-encoding transcripts (D1-A4c-D4-A7: 4.2%, D1-A4b-D4-A7:
1643 3.1%). The function of this large amount of 1 kb transcript is unknown. These two 1 kb
1644 transcripts do not appear to encode significant open reading frames although other 1 kb
1645 transcripts can encode a Rev-Nef fusion⁴⁴⁵.

1646 Using these abundances, we can estimate the number of HIV_{89.6} genomes in these primary T
1647 cells 48 hours after infection. To determine the proportion of the mRNA nucleotides from viral
1648 transcripts, we multiplied the estimated abundances by their transcript lengths. Unspliced
1649 genome transcripts appear to form 79% of the mRNA nucleotides from HIV_{89.6} transcripts.
1650 Assuming T cells contain at least 0.1 pg of mRNA then an infected cell should contain at
1651 least 0.011 pg of unspliced HIV transcript ($0.1\text{pg} \times 0.14 \frac{\text{HIV mRNA nt}}{\text{cell mRNA nt}} \times 0.79 \frac{\text{unspliced mRNA nt}}{\text{HIV mRNA nt}}$)
1652 or, assuming 9171 bases of RNA weigh about 5×10^{-6} pg, at least 2200 HIV genomes at 48
1653 hour post infection. This estimate roughly agrees with previous estimates of HIV production
1654 per cell^{538,540,541}.

1655 4.4.8 Human-HIV chimeric reads

1656 The suggestion that HIV integration may disrupt cellular cancer-associated genes and
1657 thereby promote cell proliferation^{542–545} has focused attention on the range of novel message
1658 types formed when HIV integrates within transcription units^{269,270,386,546,547}. Chimeric
1659 reads containing HIV and cellular sequence are also of clinical interest due to the potential
1660 of lentiviral vectors to trigger oncogenesis in gene therapy patients through insertional

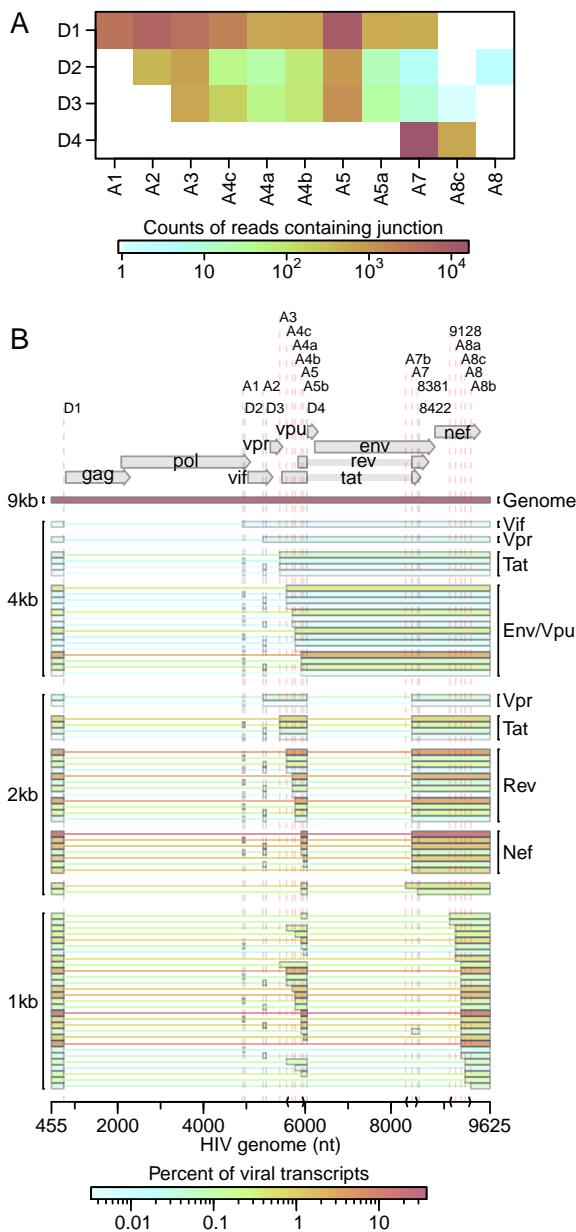


Figure 4.7: Transcription and splicing of the HIV_{89.6} RNA. A) Junctions between HIV splice donors and acceptors observed in the RNA-Seq data. Acceptors are shown as the columns and donors as the rows with the coloring indicating the frequency of each pairing. B) The relative abundance of all HIV_{89.6} transcripts as determined by a combination of PacBio sequencing⁴⁴⁵ and Illumina sequencing. Message structures were generated by targeted long read single molecule sequencing, which allowed association of multiple splice junctions in single sequence reads. The Illumina short read sequencing allowed normalization of message abundances between size classes. The inferred HIV message population is shown colored by relative abundance.

1661 mutagenesis^{548–551}.

1662 In our data, 80,045 reads contained sequences matching to both HIV and human genomic
1663 DNA, but a considerable complication arises because chimeras can be formed artifactually
1664 during the preparation of libraries for sequence analysis^{552–559}. Many of the chimeric
1665 sequences in our data contained junctions between the HIV and human sequence where the
1666 ends of the human and HIV sequence were similar and potentially complementary (Figure
1667 4.8A). This raises the concern that some of these chimeras could be products of in vitro
1668 recombinations during the reverse transcription, amplification and sequencing processes.
1669 Template switching between sequences with shared similarity is a well established property
1670 of retroviral reverse transcriptase enzymes used in RNA-Seq library preparation^{560–562}.
1671 Priming off incomplete transcripts during DNA synthesis is another potential source of
1672 chimeric transcripts^{552,553,563,564}. Failing to account for chimeras can hinder interpretation
1673 of deep sequencing data^{554–559}.

1674 Also consistent with artifactual chimera formation, 7,354 reads (9.2% of chimeric messages)
1675 contained HIV sequences joined to human mitochondrial sequences, yet HIV proviruses have
1676 not previously been found integrated in mitochondrial DNA²⁷⁰. To probe this further, we
1677 used ligation-mediated PCR to recover integration site junctions from the same infected cell
1678 populations analyzed by RNA seq, yielding 147,281 unique integration sites (Figure 4.8B)⁴²³.
1679 No integrations in mitochondrial DNA were detected. We conclude that chimeric HIV-
1680 mitochondrial sequence reads in the RNA-seq data represent artifacts of library construction
1681 and so used these chimeras as an assay to evaluate subsequent data filtering steps. We
1682 reasoned that reads without sequence similarity at junctions between human and HIV
1683 mapping were less likely to be artifacts caused by template switching. Filtering to only reads
1684 where no overlap and no unknown intervening sequence was present between human and HIV
1685 portions left 2181 junctions and reduced the proportion of reads containing mitochondrial
1686 DNA to 2.4%. Of the remaining HIV-human chimeric reads, the HIV portion of 605 sequences
1687 bordered the 3' or 5' end of HIV or an HIV splice donor or acceptor. Filtering to these

1688 more likely authentic junctions left only 2 (0.3%) chimeric reads containing mitochondrial
1689 sequence. This decrease in likely mitochondrial artifacts suggests that the filtering was
1690 effective. The high rate of mitochondrial chimeras in the unfiltered sequences raises the
1691 concern that artifacts may easily distort results in studies using similar amplification and
1692 sequencing techniques.

1693 Chimeric messages composed of HIV and cellular RNA sequences can be formed by cellular
1694 gene transcription reading into the integrated provirus, by HIV transcription reading out
1695 through the viral polyadenylation site or by splicing between human and viral splice sites.
1696 In our filtered data, the predominant forms appear to be derived from reading through the
1697 HIV polyadenylation signal into the surrounding DNA (78%), splicing out of the viral D4
1698 splice donor to join to human slice acceptors (17%) and reading into the HIV 5' LTR from
1699 human sequence (4.0%) (Figure 4.8C). No splice site other than D4 had more than two
1700 chimeric reads observed.

1701 The filtered chimeric reads had many traits consistent with biological chimera formation.
1702 The reads containing HIV D4 joined to human sequences had the characteristics expected of
1703 splicing—72.1% of the chimeric junctions mapped to known human acceptors and 96.1%
1704 mapped to a location immediately preceded by the AG consensus of human mRNA acceptors.
1705 The reads containing the 5' or 3' LTR border were almost exclusively (93%) found in
1706 transcription units, with odds of being in a gene 2.3-fold (95% CI: 1.6–3.2×) higher than
1707 integration sites from the same sample. The 5' or 3' chimeras were also more likely to be
1708 located in an exon than integration sites even after excluding any integration or chimera not
1709 located in a transcription unit (odds ratio: 2.1×, 95% CI: 1.6–2.6×).

1710 We next compared whether the human and viral segments of chimeric reads agreed or
1711 disagreed in orientation (i.e. strand transcribed) for reads with the human portion mapped
1712 within annotated transcription units. The sequencing technique used here does not preserve
1713 strand information, but we can check whether the strand of a sequence read agrees or
1714 disagrees with the annotated gene strand and compare this to the observed strand of the

1715 HIV portion of the read. We found a strong association between the orientation of the
1716 human and HIV portions of chimeric reads within 3' and 5' chimeras (odds ratio: 6.2 \times ,
1717 95% CI: 3.9–10.2 \times). This highly significant enrichment of HIV and human genes in the
1718 same orientation (Fisher's exact test $p < 10^{-15}$) might indicate that antisense HIV RNA
1719 is rapidly degraded by a response to double-stranded RNA or that polymerases oriented
1720 in opposing directions interfere with one another during elongation. Chimeras involving
1721 HIV splice donor D4 were even more highly enriched for matching orientations (odds ratio:
1722 52.5 \times , 95% CI: 12.1–307 \times) suggesting that pairing with human splice acceptors may add
1723 an additional constraint on the orientation of D4 chimeric reads.

1724 Based on these data, we can propose a lower bound on the relative abundance of chimeras. If
1725 we assume that our filtering removed nearly all artifacts so that we have few false positives,
1726 then our estimate should be lower than the true proportion of chimeras. In our data, only
1727 $\frac{604}{12,689,879} = 0.0048\%$ of reads containing sequence mapping to HIV also contained identifiable
1728 chimeric junctions. However, this is an underestimate because in an HIV-derived mRNA, any
1729 fragment of the sequence will be mappable to HIV, while for a chimeric sequence only a read
1730 spanning the HIV-human junction will allow identification of a chimera. If we assume that
1731 25 bases of sequence are necessary to map to human or HIV sequence, then, with the 100-bp
1732 reads used here, only read fragments starting between 75- and 25-bp downstream of the
1733 chimeric junction will be identifiable. If we assume the average chimeric mRNA sequences is
1734 at least 2 kb long, then a read from a chimeric sequence has at most a $\frac{50}{2000} = 2.5\%$ chance
1735 of containing a mappable junction. Thus, a lower bound for the proportion of HIV mRNA
1736 that also contain human-derived sequences is 0.2% ($\frac{0.0048\%}{2.5\%}$). Looking only at splicing from
1737 HIV donor D4, we saw 16,843 reads containing a junction from D4 to an HIV acceptor and
1738 104 reads from D4 to human sequence. Thus, in our data, 0.6% of D4 splice products form
1739 junctions with human acceptors instead of HIV acceptors.

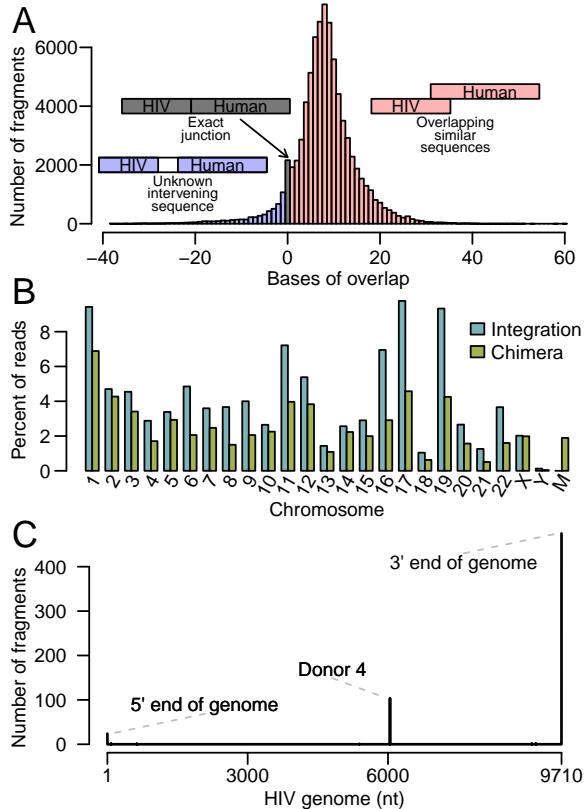


Figure 4.8: Analysis of chimeric RNA sequences containing both human and HIV sequences. A) The length of overlapping sequence (regions of complementarity potentially favoring chimera formation) matching both human and HIV at inferred chimeric junctions. The x-axis shows the length of the overlap and the y-axis shows the frequency of chimeric junctions with the indicated extent of overlap. B) Chromosomal distribution of uniquely mapping HIV integration sites from the same infections of primary T cells and comparison to uniquely mapping human sequences in chimeric reads observed in RNA-Seq. Note that the mitochondrial genome, denoted as M, has no authentic integration sites but does have extensive matches to chimeric junctions found in the RNA-Seq data. C) Counts of the location in the HIV genome of the HIV-human junctions in filtered chimeric reads.

1740 4.5 Discussion

1741 Here we used RNA-Seq to analyze mRNA accumulation and splicing in primary T cells
 1742 infected with the low passage isolate HIV_{89.6}. We did not carry out dense time series analysis,
 1743 compare different human cell donors or compare different perturbations of the infections—
 1744 instead, we focused on generating a dense data set at a single time point. We analyzed
 1745 replicate infected cell and control samples to allow discrimination of within-condition versus
 1746 between-condition variation and assessed differences using a series of bioinformatic approaches.
 1747 Many previous studies have used microarray technology or RNA-Seq to study gene activity
 1748 in HIV-infected cells^{317,319,358,491–499}, usually analyzing infections of transformed cell lines
 1749 or laboratory adapted strains of HIV-1. Here we present what is to our knowledge the
 1750 deepest RNA-Seq data set reported for infection in primary T cells using a low passage
 1751 HIV isolate (HIV_{89.6}). This data set was paired with a set of 147,281 unique integration
 1752 site sequences extracted from the same infections, which were critical to our ability to

1753 quality control chimeric reads. An advantage of studies using cell lines and laboratory
1754 adapted strains is that often a high percent of cell infection can be achieved, whereas in
1755 this study we achieved only ~30% infection. However, we report distinctive features of the
1756 transcriptional response not seen in studies of HIV infections in cell lines. Novel in this
1757 study are 1) identification of intron retention as a consequence of HIV infection, 2) the
1758 finding of activation of ERV-9/LTR12C after HIV infection, 3) generation of a quantitative
1759 account of the structures and abundances of over 70 HIV_{89.6} messages and 4) clarification of
1760 the predominant types of HIV-host transcriptional chimeras. These findings are discussed
1761 below.

1762 Broad changes in host cell mRNA abundances were evident after infection, with over 17% of
1763 expressed genes changing significantly in activity. Changes included expected response to
1764 viral infection, apoptosis and T cell activation. Although it is not possible here to separate
1765 the response of infected and bystander cells, this study highlights the drastic changes in
1766 cellular expression caused by HIV-1 infection. In a meta-analysis including four previously
1767 published studies, no gene was detected as differentially expressed in all five studies and
1768 only a handful of genes appeared in four out of five studies. Further analysis showed that
1769 expression changes appear to be cell type specific, raising concerns that studies using cell
1770 lines may not fully reflect host cell responses in *in vivo* infections.

1771 Unexpectedly, intronic sequences were more common in the RNA-Seq data from cells after
1772 HIV_{89.6} infection than in mock infected cells. The mechanism is unclear. It is possible
1773 that the splicing machinery is reduced in activity after 48 hours of infection, perhaps as a
1774 part of the antiviral response of infected and bystander cells. HIV infection does appear to
1775 alter expression and localization of some splicing factors^{281,526}. In addition, we saw a large
1776 reduction in the abundance of mRNA from nonsense-mediated decay related genes, perhaps
1777 indicating that RNA surveillance is loosened thus allowing more unspliced or aberrantly
1778 spliced transcripts. Alternatively, fully spliced mRNAs might be more rapidly degraded after
1779 infection, possibly by interferon-mediated induction of RNaseL⁵⁶⁵. A speculative possibility

1780 is that HIV_{89.6} encodes a factor that alters cellular splicing or promotes mRNA degradation
1781 to optimize splicing and translation of viral messages.

1782 Infection resulted in increased expression of specific cellular repeated sequences. HERVs, in
1783 particular HERV-K, have previously been observed to show increased RNA accumulation with
1784 HIV infection^{527–529,566} and possibly represent vaccine targets because of their production of
1785 distinctive proteins^{533,566–570}. Here, though we saw modest increases in HERV-K expression,
1786 ERV-9 had the greatest change in expression (33 LTR12C and 14 ERV-9 annotated regions
1787 with greater than 4× change in expression). Previous RNA-Seq studies of HIV infection in
1788 cell lines did not report increases in HERV expression^{319,358} but this difference is likely due
1789 to a much higher baseline expression of HERVs in transformed cell lines. We also observed
1790 increases in LINE and Alu element transcription, as has been reported previously⁵³⁰, and
1791 expression changes in ERV-9/LTR12C expression associated with transcription factor motifs
1792 and U3 variants.

1793 Many of the repeated sequence elements that were induced by HIV_{89.6} infection are relatively
1794 recently integrated in the human genome. The reason for this pattern is unclear. It may
1795 be that older elements have accumulated more mutations, resulting in an inactivation of
1796 transcriptional signals. Alternatively, perhaps the elements that are induced have been
1797 recruited for transcriptional control of cellular functions, so that their transcriptional activity
1798 is preserved evolutionarily^{510,571,572}.

1799 Comparison of results of sequencing HIV_{89.6} messages using long-read single molecule
1800 sequencing (Pacific Biosciences) and dense short read sequencing (Illumina data reported
1801 here) allowed a full quantitative accounting of more than 70 HIV_{89.6} splice forms. The full
1802 length unspliced HIV RNA comprised 37.6% of all messages, corresponding to about 2000
1803 genomes per cell. Notably abundant messages included those encoding Nef (D1-A5-D4-A7:
1804 15.5%) and two Rev-encoding transcripts (D1-A4c-D4-A7: 4.2%, D1-A4b-D4-A7: 3.1%).
1805 The full set of messages is summarized in Figure 4.7B. Our previous analysis revealed an
1806 unusually prominent 1 kb size class. HIV_{89.6} encodes a rare splice acceptor (A8c) within Nef

1807 responsible for formation of the short messages. Our data indicated that two members of the
1808 1-kb size class, D1-A5-D4-A8c and D1-A8c, accounted for 10.6% and 4.9% of all messages.
1809 The 1 kb size class as a whole accounted for fully 20% of messages. Most HIV/SIV variants
1810 appear to encode an acceptor near this position, suggesting a potential unknown function
1811 for these short spliced forms^{445,449,453}.

1812 After filtering, we detected a sizeable number of apparently authentic chimeras containing
1813 both HIV and cellular sequences, allowing comparison to examples of host-cell modification
1814 by integration. Mechanisms of insertional activation have been studied intensively in animal
1815 models of transformation and in adverse events in human gene therapy. One of the most
1816 common mechanisms involves insertion of a retroviral enhancer near a cellular promoter,
1817 so that the rate of initiation is increased and normal cellular messages are increased in
1818 abundance. However, another common mechanism involves formation of chimeric messages
1819 involving both cellular and viral/vector sequences. In HIV infection, examples of insertion
1820 in the Bach2 and MKL2 genes have been associated with long term persistence of particular
1821 cell clones^{542–545}. In these cells, proviruses were integrated within the cellular transcription
1822 unit, and the transcriptional direction of the integrated provirus was the same as that of
1823 Bach2 or MKL2. This would allow formation of a fusion of the 5' HIV sequences with 3'
1824 Bach2 sequences, potentially involving the most common events seen here (either 3' read out
1825 or splicing from HIV D4 to a cellular exon). However, a closely studied example of clonal
1826 expansion in a successful lentiviral vector gene therapy for beta-thalassemia was associated
1827 with expansion of a cell clone harboring an integrated vector within the transcription unit
1828 of HMGA2. In this case the message spliced into the vector and terminated, removing
1829 a negative regulatory sequence normally present in the 3' end HMGA2 message⁵⁴⁸. A
1830 targeted study in vitro of chimeric message formation by lentiviral vectors showed examples
1831 of multiple types of read-in and -out and splice-in and -out⁵⁵⁰, which may have been more
1832 frequent and more varied than for HIV^{89.6} proviruses studied here. The lack of splicing or
1833 reading into HIV in this study may be a reflection of the high rate HIV transcription in
1834 these infected cells—because HIV was so highly expressed, there would be more opportunities

1835 for polymerase to splice out of or read through the HIV genome than to read or splice in.
1836 The vast majority of HIV proviruses in expanded clones in well-suppressed patients now
1837 appear to be defective⁵⁴⁵—going forward, it will be of interest to investigate whether these
1838 HIV proviruses are damaged in ways that promote formation of chimeric transcripts.

1839 Lastly, we note that several features of the transcriptional response to HIV_{89.6} infection were
1840 suggestive of de-differentiation away from T cell specific expression patterns. The increase
1841 in expression of cellular HERVs and LINEs is characteristic of cells in early development.
1842 Specific HERVs and transposons, including ERV-9/LTR12C and HERV-K, have been
1843 implicated in regulating gene activity early in development^{510,571,573–576}. Several genes
1844 related to other hematopoietic cell types showed elevated RNA abundance after HIV_{89.6}
1845 infection. These data are of interest given the finding that patients undergoing long term
1846 ART can contain long lived T cell clones that may contribute to the latent reservoir^{545,577–580}.
1847 Possibly the transcriptional responses seen in infected primary T cells here are reflective
1848 of processes leading to formation of the long-lived latently-infected cells with stem-like
1849 properties.

1850 4.6 Conclusions

1851 Infections of primary T cells with a low passage HIV isolate show several distinctive features
1852 compared with previously published data using T cell lines and/or lab-adapted HIV strains.
1853 We found strong changes in expression in genes related to immune response and apoptosis
1854 similar to studies of HIV infection in patient samples and primary cells but different from
1855 studies performed in SupT1 cell lines. Notable changes after infection included intron
1856 retention and activation of recently integrated retrotransposons and endogenous retroviruses,
1857 in particular LTR12C/ERV-9. We also present complete absolute estimation of over 70
1858 messages from HIV_{89.6} and specify the major virus-host chimeras as read out from the 3'
1859 end of the provirus and splicing from viral splice donor 4 to cellular acceptors.

1860 **4.7 Availability of supporting data**

1861 RNA-Seq reads from this study are available at the Sequence Read Archive under accession
1862 number SRP055981. The integration site data is available at the Sequence Read Archive
1863 under accession number SRP057555.

1864 **4.8 Acknowledgements**

1865 We would like to thank the University of Pennsylvania Center for AIDS Research (P30
1866 AI045008) for preparation of viral stocks and isolation of primary CD4⁺ T cells; Ronald
1867 G. Collman and members of the Bushman laboratory for reagents, helpful discussion and
1868 technical expertise. This work was funded by NIH grant R01 AI052845, the HIV Immune
1869 Networks Team (HINT) consortium P01 AI090935 and NRSA computational genomics
1870 training grant T32 HG000046.

1871 **CHAPTER 5: A reverse transcription loop-mediated isothermal**
1872 **amplification assay optimized to detect multiple HIV**
1873 **subtypes**

This chapter was originally published as:

KE Ocwieja, S Sherrill-Mix, C Liu, J Song, H Bau and FD Bushman. 2015. A reverse transcription loop-mediated isothermal amplification assay optimized to detect multiple HIV subtypes. *PLoS One*, 10:e0117852. doi: 10.1371/journal.pone.0117852

KE Ocwieja and I were joint first authors. KE Ocwieja, C Liu, H Bau, FD Bushman and I conceived the experiments. KE Ocwieja and I designed the assay. KE Ocwieja, C Liu and J Song performed the experiments. KE Ocwieja, J Song and I analyzed the data. I produced the figures. KE Ocwieja, C Liu, H Bau, FD Bushman and I wrote the paper.

Supporting information are available at <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117852#sec011>

1875 **5.1 Abstract**

1876 Diagnostic methods for detecting and quantifying HIV RNA have been improving, but
1877 efficient methods for point-of-care analysis are still needed, particularly for applications in
1878 resource-limited settings. Detection based on reverse-transcription loop-mediated isothermal
1879 amplification (RT-LAMP) is particularly useful for this, because when combined with
1880 fluorescence-based DNA detection, RT-LAMP can be implemented with minimal equipment
1881 and expense. Assays have been developed to detect HIV RNA with RT-LAMP, but existing
1882 methods detect only a limited subset of HIV subtypes. Here we report a bioinformatic study
1883 to develop optimized primers, followed by empirical testing of 44 new primer designs. One
1884 primer set (ACeIN-26), targeting the HIV integrase coding region, consistently detected
1885 subtypes A, B, C, D, and G. The assay was sensitive to at least 5000 copies per reaction for

1886 subtypes A, B, C, D, and G, with Z-factors of above 0.69 (detection of the minor subtype F
1887 was found to be unreliable). There are already rapid and efficient assays available for detecting
1888 HIV infection in a binary yes/no format, but the rapid RT-LAMP assay described here has
1889 additional uses, including 1) tracking response to medication by comparing longitudinal
1890 values for a subject, 2) detecting of infection in neonates unimpeded by the presence of
1891 maternal antibody, and 3) detecting infection prior to seroconversion.

1892 **5.2 Introduction**

1893 Despite the introduction of efficient antiretroviral therapy, HIV infection and AIDS continue
1894 to cause a worldwide health crisis⁵⁸². Methods for detecting HIV infection have improved
1895 greatly with time⁵⁸³—today rapid assays are available that can detect HIV infection in a
1896 yes-no format using a home test kit that detects antibodies in saliva. Viral load assays that
1897 quantify viral RNA with quick turn-around time are widely available in the developed world.
1898 However, quantitative viral load assays are not commonly available with actionable time
1899 scales in much of the developing world. This motivates the development of new rapid and
1900 quantitative assays that can be used at the point of care with minimal infrastructure^{584,585}.

1901 One simple and quantitative detection method involves reverse transcription-based loop
1902 mediated isothermal amplification (RT-LAMP)⁵⁸⁶. In this method, a DNA copy of the viral
1903 RNA is generated by reverse transcriptase, and then isothermal amplification is carried out to
1904 increase the amount of total DNA. Primer binding sites are chosen so that a series of strand
1905 displacement steps allow continuous synthesis of DNA without requiring thermocycling.
1906 Reaction products can be detected by adding an intercalating dye to reaction mixtures
1907 that fluoresces only when bound to DNA, allowing quantification of product formation by
1908 measurement of fluorescence intensity. Such assays can potentially be packaged in simple
1909 self-contained devices and read out with no technology beyond a cell phone.

1910 RT-LAMP assays for HIV-1 have been developed previously and reported to show high
1911 sensitivity and specificity for subtype B, the most common HIV strain in the developed

world^{585,587,588}. Another recent study reported RT-LAMP primer set optimized for the detection of HIV variants circulating in China⁵⁸⁹, and another on confirmatory RT-LAMP for group M viruses⁵⁹⁰. Assays have also been developed for HIV-2⁵⁹¹. A complication arises in using available RT-LAMP assays due to the variation of HIV genomic sequences among the HIV subtypes^{592,593}, so that an RT-LAMP assay optimized for one viral subtype may not detect viral RNA of another subtype⁵⁹⁴. Tests presented below show that many RT-LAMP assays are efficient for detecting subtype B, for which they were designed, but often performed poorly on other subtypes. Subtype C infects the greatest number of people worldwide, including in Sub-Saharan Africa, where such RT-LAMP assays would be most valuable, motivating optimization for subtype C. Several additional non-B subtypes are also responsible for significant burdens of disease world-wide⁵⁹⁵.

Here we present the development of an RT-LAMP assay capable of detecting HIV-1 subtypes A, B, C, D, and G. We first carried out a bioinformatic analysis to identify regions conserved in all the HIV subtypes. We then tested 44 different combinations of RT-LAMP primers targeting this region in over 700 individual assays, allowing identification of a primer set (ACeIN-26) that was suitable for detecting these subtypes. We propose that the optimized RT-LAMP assay may be useful for quantifying HIV RNA copy numbers in point-of-care applications in the developing world, where multiple different subtypes may be encountered.

5.3 Methods

5.3.1 Viral strains used in this study

Viral strains tested included HIV-1 92/UG/029 (Uganda) (subtype A, NIH AIDS Reagent program reagent number 1650), HIV-1 THRO (subtype B, plasmid derived, University of Pennsylvania CFAR)⁵⁹⁶, CH269 (subtype C, plasmid derived, University of Pennsylvania CFAR)⁵⁹⁶, UG0242 (subtype D, University of Pennsylvania CFAR), 93BRO20 (subtype F, University of Pennsylvania CFAR), HIV-1 G3 (subtype G, NIH AIDS Reagent program reagent number 3187)⁵⁹⁷.

1938 Viral stocks were prepared by transfection and infection. Culture supernatants were cleared
1939 of cellular debris by centrifugation at 1500g for 10 min. The supernatant containing virus
1940 was then treated with 100 U DNase (Roche) per 450 uL virus for 15 min at 30°C. RNA was
1941 isolated using the QiaAmp Viral RNA mini kit (Qiagen GmbH, Hilden, Germany). RNA
1942 was eluted in 80 uL of the provided elution buffer and stored at -80°C.

1943 Concentration of viral RNA copies was calculated from p24 capsid antigen capture assay
1944 results provided by the University of Pennsylvania CFAR or the NIH AIDS-reagent program.
1945 In calculating viral RNA copy numbers, we assumed that all p24 was incorporated in virions,
1946 all RNA was recovered completely from stocks, 2 genomes were present per virion, 2000 p24
1947 molecules per viral particle, and the molecular weight of HIV-1 p24 was 25.6 kDa.

1948 5.3.2 Assays

1949 RT-LAMP reaction mixtures (15 μ L) contained 0.2 μ M each of primers F3 and B3 (if a
1950 primer set used multiple B3 primers, mixture contained 0.2 μ M of each); 1.6 μ M each of FIP
1951 and BIP primers (if a primer set had multiple FIP primes, reaction mixture contained 0.8
1952 μ M of each FIP primer); and 0.8 μ M each of LoopF and LoopB primers; 7.5 μ L OptiGene
1953 Isothermal Mastermix ISO-100nd (Optigene, UK), ROX reference dye (0.15 μ L from a 50X
1954 stock), EvaGreen dye (0.4 μ L from a 20X stock; Biotium, Hayward, CA); HIV RNA in 4.7
1955 μ L; AMV reverse transcriptase (10U/ μ L) 0.1 μ L and water to 15 μ L In most cases where
1956 two primer sets were combined, the total primer concentration within the reaction was
1957 doubled such that the above individual primer molarities were maintained. For the mixture
1958 ACeIN-26+F-IN (S2 Table, line 46), the total primer concentration was not doubled—the
1959 F-IN primer set comprised 25% of the total primer concentration, and the ACeIN-26 primer
1960 set comprised 75% of the total primer concentration with the ratios of primers listed above
1961 preserved. This mixture was combined 1:1 with the ACe-PR primer set (S2 Table, line 47)
1962 such that total primer concentration in the final mixture was doubled.

1963 Amplification was measured using the 7500-Fast Real Time PCR system from Applied

1964 Biosystems with the following settings: 1 minute at 62°C; 60 cycles of 30 seconds at 62°C
1965 and 30 seconds at 63°C. Data was collected every minute. Product structure was assessed
1966 using dissociation curves which showed denaturation at 83°C. Products from selected
1967 amplification reactions were analyzed by agarose gel electrophoresis and showed a ladder of
1968 low molecular weight products (data not shown).

1969 Product synthesis was quantified as the cycle of threshold for 10% amplification. Z-factors⁵⁹⁸
1970 were calculated from tests of 24 replicates using the ACeIN26 primer set in assays with viral
1971 RNA of each subtype. No detection after 60 min was given a value of 61 min in the Z-factor
1972 calculation.

1973 5.4 Results

1974 5.4.1 Testing published RT-LAMP primer sets against multiple HIV subtypes

1975 We first assessed the performance of existing RT-LAMP assays on RNA samples from
1976 multiple HIV subtypes. We obtained viral stocks from HIV subtypes A, B, C, D, F, and
1977 G, estimated the numbers of virions per ml, and extracted RNA. RNAs were mixed with
1978 RT-LAMP reagents which included the six RT-LAMP primers, designated F3, B3, FIP, BIP,
1979 LF and LB⁵⁸⁶. Reactions also contained reverse transcriptase, DNA polymerase, nucleotides
1980 and the intercalating fluorescent EvaGreen dye, which yields a fluorescent signal upon DNA
1981 binding. DNA synthesis was quantified as the increase in fluorescence intensity over time,
1982 which yielded a typical curve describing exponential growth with saturation (examples are
1983 shown below). Results are expressed as threshold times (T_t) for achieving 10% of maximum
1984 fluorescence intensity at the HIV RNA template copy number tested.

1985 In initial tests, published primer sets targeting the HIV-1 subtype B coding regions for
1986 capsid (CA), protease (PR), and reverse transcriptase (RT) (named B-CA, B-PR and B-RT)
1987 were assayed in reactions with RNAs from four of the subtypes. Results with each primer set
1988 tested are shown in Figure 5.1 in heat map format, where each tile summarizes the results of
1989 tests of 5000 RNA copies. Primers and their groupings into sets are summarized in S1 and

1990 S2 Tables, average assay results are in S3 Table, and raw assay data is in S4 Table. Assays
1991 (Figure 5.1, top) with the B-CA, B-PR and B-RT primer sets detected subtypes B and D
1992 at 5000 RNA copies with threshold times less than 20 min. However, assays with B-CA
1993 and B-RT detected subtypes C and F with threshold times > 50 min, indicating inefficient
1994 amplification and the potential for poor separation between signal and noise. B-PR did
1995 not detect subtype C at all. In an effort to improve the breadth of detection, we first tried
1996 mixing the B-PR primers, which detected clade F (albeit with limited efficiency) with the
1997 B-CA and B-RT primers (Figure 5.1 and S3 and S4 Tables). In neither case did this provide
1998 coverage of all four clades tested. We thus did not test these primer sets on RNAs from the
1999 remaining subtypes and instead sought to develop primer sets targeting different regions of
2000 the HIV genome.

2001 **5.4.2 Primer design strategy**

2002 To design primers that detected multiple HIV subtypes efficiently, we analyzed alignments
2003 of HIV genomes (downloaded from the Los Alamos National Laboratory site⁵⁹²) for regions
2004 with similarity across most viruses, revealing that a segment of the pol gene encoding
2005 IN was particularly conserved (Figure 5.2A). A total of six primers are required for each
2006 RT-LAMP assay⁵⁸⁶. We used the EIKEN primer design tool to identify an initial primer set
2007 targeting this region. In further analysis, positions in the alignments were identified within
2008 primer landing sites that commonly contained multiple different bases. Primer positions
2009 were manually adjusted to avoid these bases when possible, and when necessary mixtures
2010 were formulated containing each of these commonly occurring bases (S1 and S2 Tables).
2011 An extensive series of variants targeting the IN coding region was tested empirically in
2012 assays containing RNAs from multiple subtypes (5000 RNA copies per reaction, over 700
2013 total assays; S3 and S4 Tables). Based on initial results, primers were further modified by
2014 adjusting the primer position or addition of locked nucleic acids as described below.

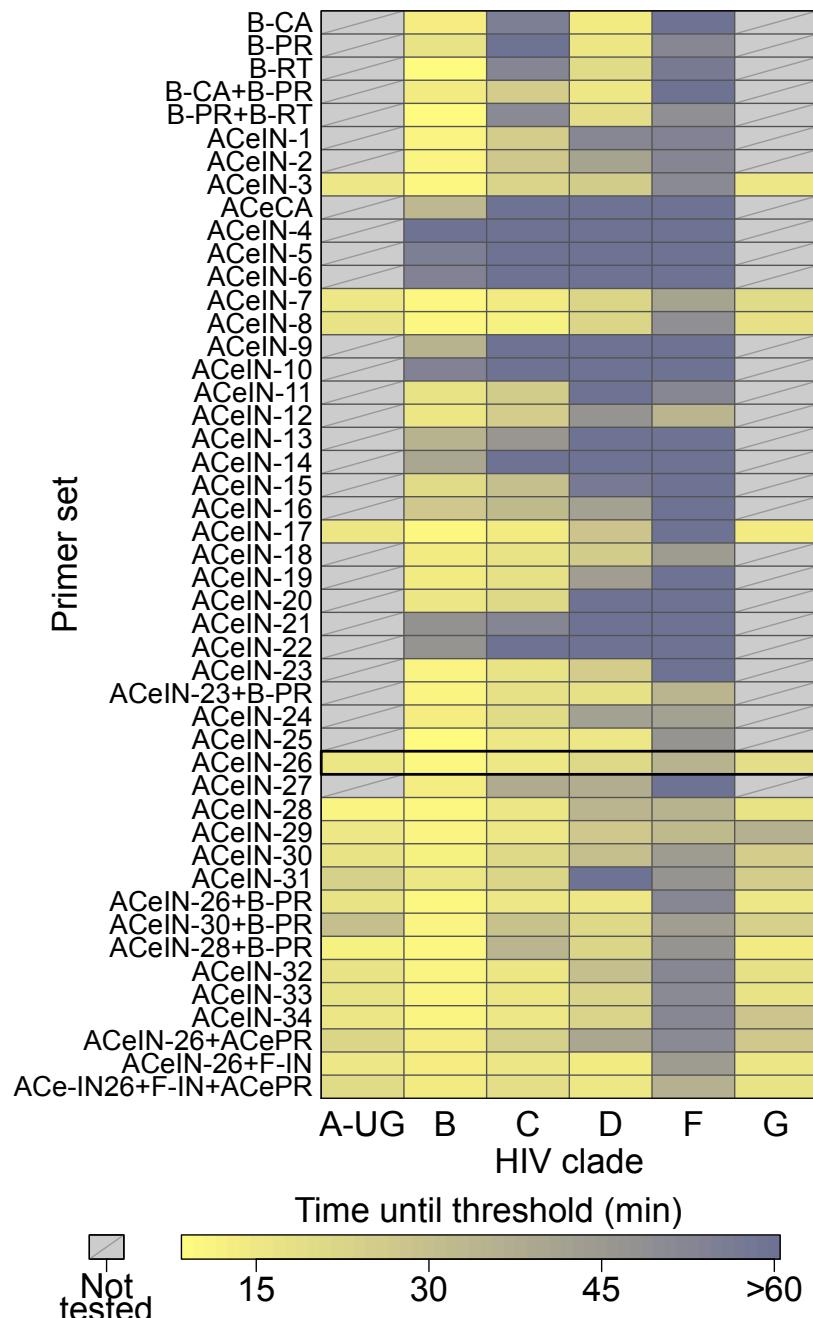


Figure 5.1: Summary of amplification results for all the RT-LAMP primer sets tested in this study. The data is shown as a heat map, with more intense yellow coloring indicating shorter amplification times (key at bottom). Primer sets tested are named along the left of the figure. Primer sequences, and their organization into LAMP primer sets, are catalogued in S1 and S2 Tables. The raw data and averaged data are collected in S3 and S4 Tables. ACeIN-26 primer set (highlighted) had one of the best performances across the subtypes and a relatively simple primer design.

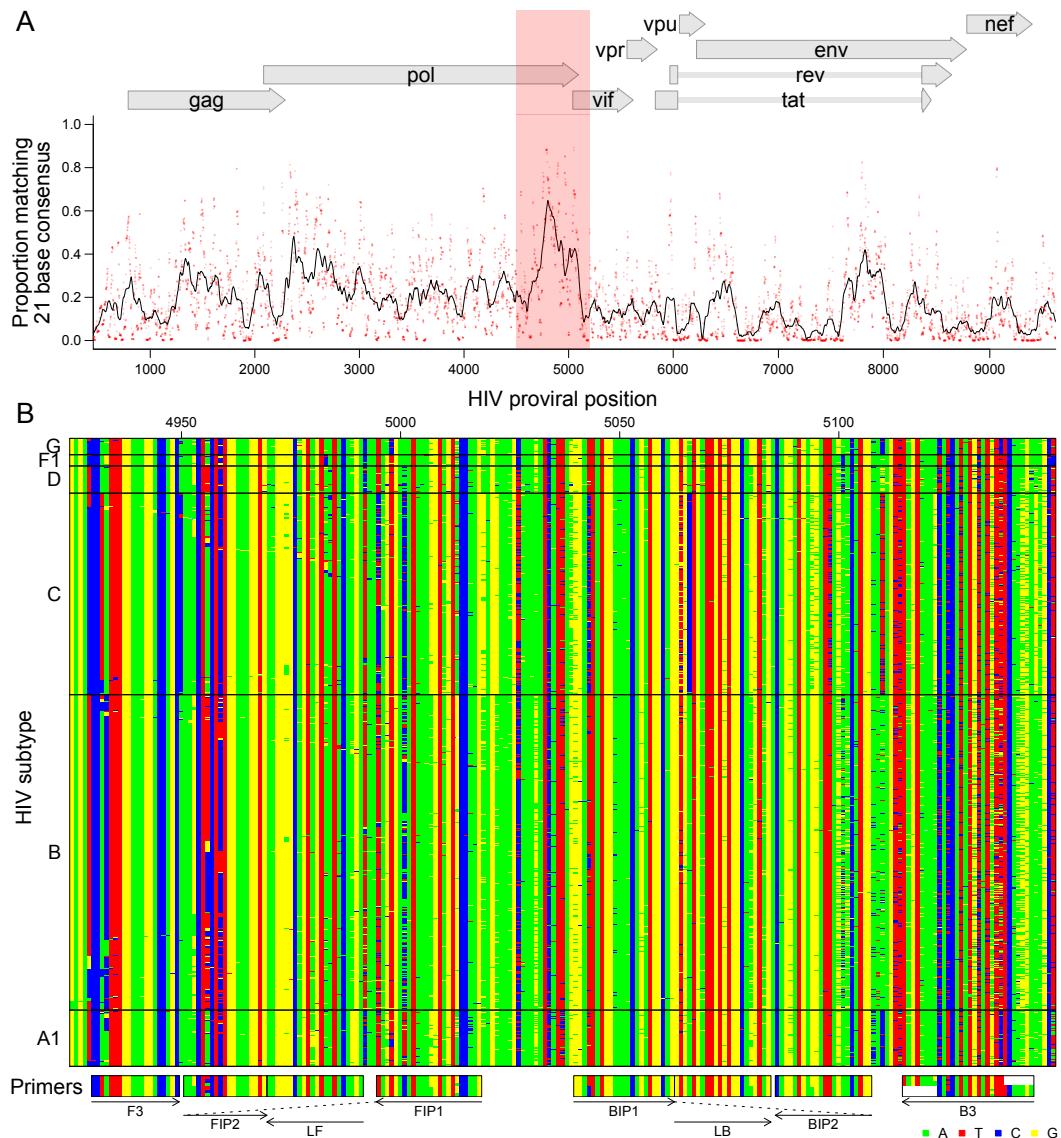


Figure 5.2: Bioinformatic analysis to design subtype-agnostic RT-LAMP primers. A) Conservation of sequence in HIV. HIV genomes ($n = 1340$) from the Los Alamos National Laboratory collection were aligned and conservation calculated. The x-axis shows the coordinate on the HIV genome, the y-axis shows the proportion of sequences matching the consensus for each 21 base segment of the genome (red points). The black line shows a 101 base sliding average over these proportions. The vertical red shading shows the region targeted for LAMP primer design that was used as input into the EIKEN primer design tool. Numbering is relative to the HIV_{89.6} sequence. B) Aligned genomes, showing the locations of the ACeIN-26 primers. Sequences are shown with DNA bases color-coded as shown at the lower right. Each row indicates an HIV sequence and each column a base in that sequence. Horizontal lines separate the HIV subtypes (labeled at left). Arrows indicate the strand targeted by each primer. Primers targeting the negative strand of the virus are shown as reverse compliments for ease of viewing.

2015 5.5 Testing different primer designs

2016 Our first design, ACeIN-1 (“ACe” for “All Clade” and “IN” for “integrase”), targeted the
2017 HIV IN coding region and contained multiple bases at selected sites to broaden detection
2018 (Figure 5.1). ACeIN-2 and-3 have primers (B3) with slightly different landing sites. Tests
2019 showed that the mixture of primers allowed amplification with a shorter threshold time than
2020 did either alone (Figure 5.1).

2021 We also tried to design a new primer set to the CA coding region (Figure 5.1, ACeCA)
2022 but found that the set only amplified clade B, and not efficiently. Thus this design was
2023 abandoned.

2024 ACeIN-3 through-6 were altered by inserting a polyT sequence between the two different
2025 sections of FIP and BIP in various combinations, a modification introduced with the goal of
2026 improving primer folding, but these designs performed quite poorly (Figure 5.1).

2027 Because the FIP primer appeared to bind the region with most variability among clades, we
2028 tried variations that bound to several nearby regions. These were tried with and without
2029 the polyT containing BIP and FIP primers in various combinations (Figure 5.1, ACeIN-7
2030 through-22). We also tried mixing all of the variations of FIP together (ACeIN-23; S2 Table).
2031 The ACeIN-23 primer set was tried as a mixture with the B-PR set to try to capture clade
2032 F, yielding a relatively effective primer set (Figure 5.1, ACeIN-23+B-PR).

2033 In an effort to increase affinity, an additional G/C pair was added to F3 and tested with
2034 various other IN primers (Figure 5.1, ACeIN-24 through-31). Testing showed improvement,
2035 with ACeIN-26 showing particularly robust amplification.

2036 In a second effort to increase primer affinities, we substituted locked nucleic acids (LNAs) for
2037 selected bases that were particularly highly conserved among subtypes (Figure 5.1, ACeIN-30,
2038 -31, -32, -33, and-34). Some improvement was shown over the non-LNA containing bases.
2039 However, the ACeIN-26 primer set was as effective as or better than any LNA containing

2040 primer sets.

2041 In further tests, the ACeIN-26, -28 and-30 primers were tested combined with the ACePR
2042 primer set (a slightly modified version of the B-PR primer set, S2 Table, row 2, designed
2043 to accommodate a wider selection of HIV-1 subtypes) but no improvement was seen and
2044 efficiency may even have fallen for some subtypes. We also designed a primer set that
2045 matched exactly to the targeted sequences found in the problematic subtype F, and mixed
2046 this set with the ACeIN-26 primers. However, no improvement was seen (Figure 5.1, mixtures
2047 with F-IN set). Mixing the ACeIN-26 primers with both the ACePR and F-specific primers
2048 did yield effective primer sets (Figure 5.1, ACeIN26+F-IN and ACeIN26+F-IN+ACePR).
2049 However, amplification efficiency was not greatly improved over the ACeIN-26 primer set, so
2050 we proceeded with the simpler ACeIN-26 primer set (Figure 5.2B) in further studies.

2051 **5.5.1 Performance of the optimized RT-LAMP assay**

2052 The ACeIN-26 RT-LAMP primer set was next tested to determine the minimum concentration
2053 of RNA detectable under the reaction conditions studied (Figure 5.3). RNA template amounts
2054 were titrated and time to detection quantified. Tests showed detection after less than 20
2055 min of incubation for 50 copies of subtypes A or B, detection after less than 30 min for 5000
2056 copies for C, D, and G, and detection after less than 20 min for 50,000 copies for F.

2057 For clinical implementation the reliability of an assay is critical. This is commonly sum-
2058 marized as a Z-factor⁵⁹⁸, which takes into account both the separation in means between
2059 positive and negative samples and the variance in measurement of each. An assay with
2060 a Z-factor above 0.5 is judged to be an excellent assay. Z-factors for detection of each of
2061 the subtypes at 5000 RNA copies per reaction were > 0.50 for subtypes A, B, C, D, and
2062 G, respectively (Figure 5.4, n = 24 replicates per test). Detection of subtype F at 5000
2063 copies per reaction was sporadic, showing a much lower Z-factor. Therefore our ACeIN-26
2064 RT-LAMP primer set appears well suited to detect 5000 copies of subtypes A, B, C, D and
2065 G.

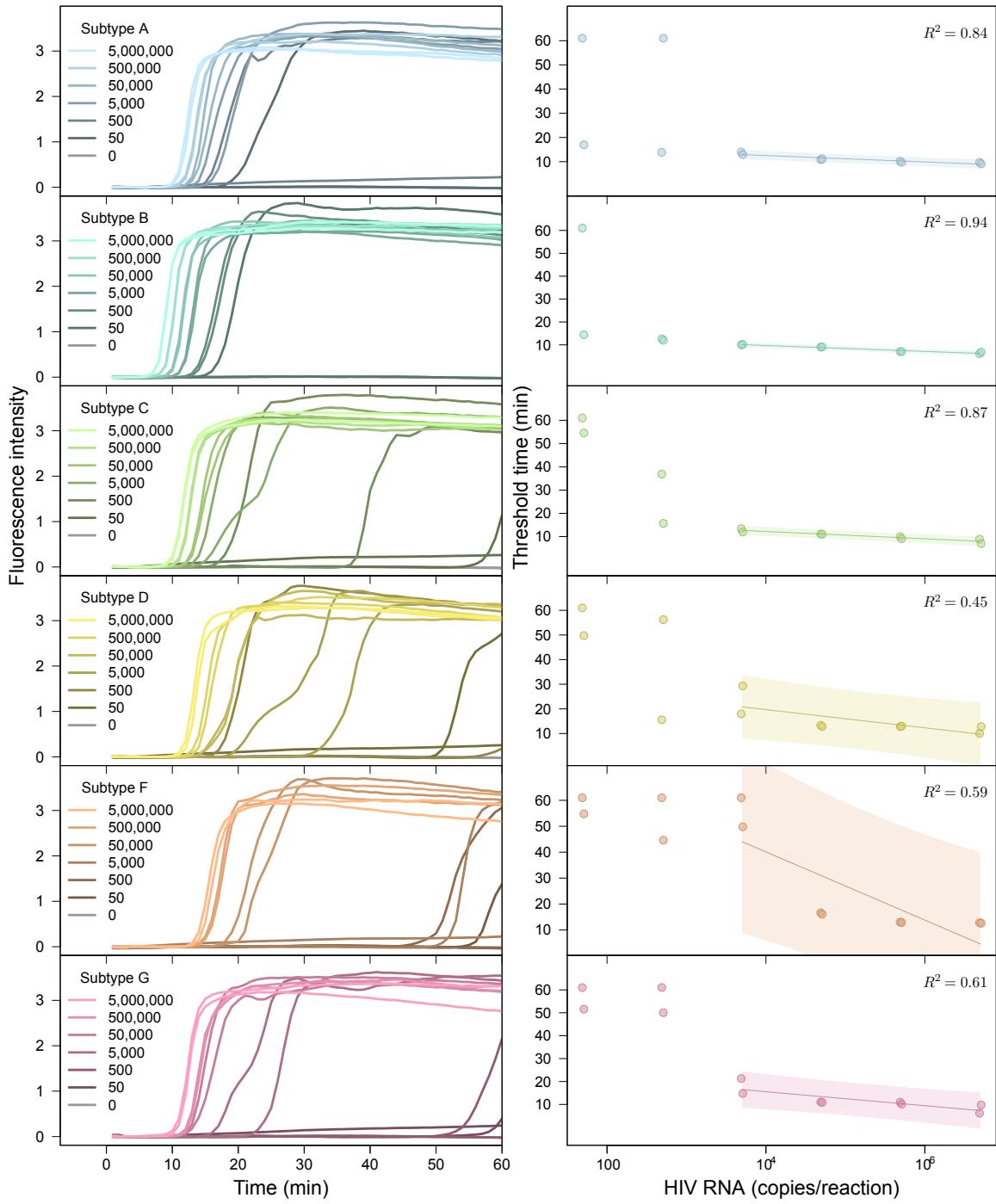


Figure 5.3: Performance of the AceIN-26 primer set with different starting RNA concentrations. Tests of each subtype are shown as rows. In each lettered panel, the left shows the raw accumulation of fluorescence signal (y-axis) as a function of time (x-axis); the right panel shows the threshold time (y-axis) as a function of log RNA copy number (x-axis) added to the reaction. In the right hand panels, values were dithered where two points overlapped to allow visualization of both.

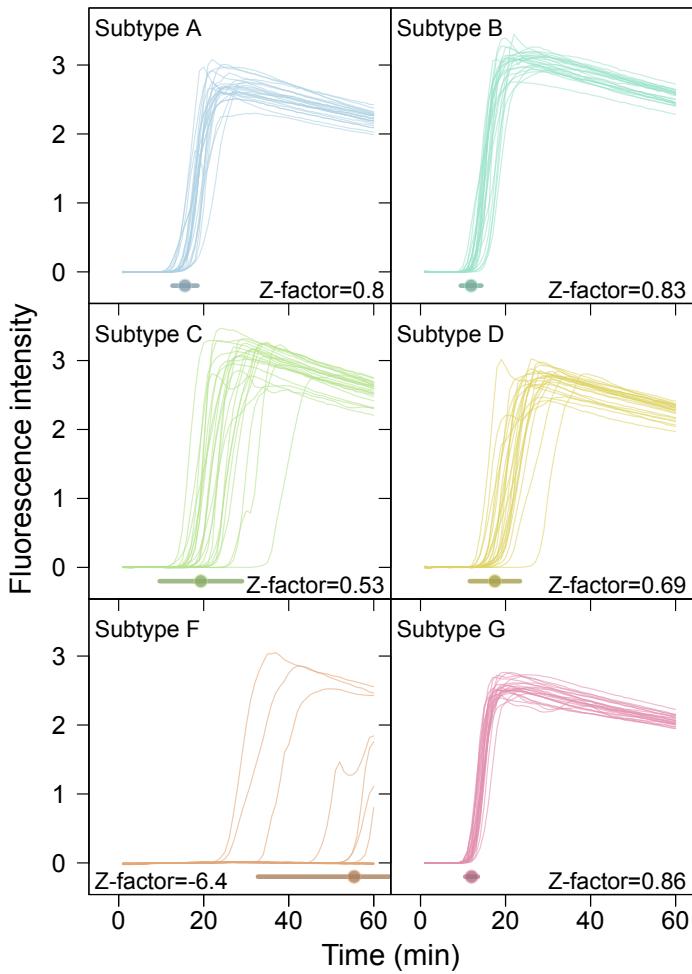


Figure 5.4: Examples of time course assays, displaying replicate tests of RT-LAMP primer set ACeIN-26 tested over six HIV subtypes, used in Z-factor calculations. A total of 5000 RNA copies were tested in each 15 μ L reaction. Time is shown on the x-axis, Fluorescence intensity on the y-axis. Replicates are distinguished using an arbitrary color code. Z-factor values and standard deviations are shown on each panel.

2066 **5.6 Discussion**

2067 Here we present an RT-LAMP assay optimized to identify multiple HIV subtypes. Infections
2068 with subtype B predominate in most parts of the developed world, but elsewhere other
2069 subtypes are more common⁵⁹⁵. Thus nucleic acid-based assays for use in the developing
2070 world need to query HIV subtypes more broadly. Previously reported RT-LAMP assays,
2071 while effective at detecting subtype B, commonly showed poor ability to detect at least some
2072 of the HIV subtypes, including C, which is common in the developing world (Figure 5.1).
2073 Here we first carried out an initial bioinformatic survey to identify regions conserved across
2074 all HIV subtypes that could serve as binding sites for RT-LAMP primers. We then tested
2075 primer sets targeting these regions empirically for efficiency. Testing 44 different primer
2076 sets revealed that assays containing ACeIN-26 were effective in detecting 5000 copies of
2077 RNA from subtypes A, B, C, D, and G within 30 minutes of incubation. For these five
2078 subtypes, the times of incubation to reach the threshold times were not too different, which
2079 simplifies interpretation when the subtype in the sample is unknown. Regardless of the
2080 efficiency, these assays can be applied to longitudinal studies of changes in viral load within
2081 an individual. We propose that RT-LAMP assays based on the ACeIN-26 primer set can be
2082 useful world-wide for assaying HIV-1 viral loads in infected patients.

2083 There are several limitations to our study. Subtypes A, B, C, D, and G were detected
2084 efficiently and showed Z-factors above 0.5, but subtype F was detected reliably only with
2085 higher template amounts, probably due to more extensive mismatches with the ACeIN-26
2086 primer set. Subtype F is estimated, however, to comprise only 0.59% of all infections
2087 globally⁵⁹⁵, though it is common in some regions. For many of the common circulating
2088 recombinant forms, such as AE and BC, the target site for ACeIN-26 is from a subtype
2089 known to be efficiently detected, though in some cases the efficiency of detection is not easy
2090 to predict and will need to be tested. We did not test subtypes beyond A, B, C, D, F and
2091 G, and we did not attempt to assess multiple different variants within each subtype. Thus,
2092 while we do know that our RT-LAMP assays are more widely applicable than many of those

2093 reported previously, we do not know whether they are able to detect all strains efficiently. In
2094 addition, although we carried out more than 700 assays in this study, there remain multiple
2095 parameters that could be optimized further, such as primer concentrations, salt type and
2096 concentration, temperature, and divalent metal concentrations, so there are likely further
2097 opportunities for improvement. Also, possible effects of RNA quality on assay performance
2098 were not tested rigorously.

2099 A particularly important parameter for further optimization is primer sequence. Several
2100 groups have recently published primer sets optimized for broad detection of different HIV
2101 lineages^{589,590}, offering opportunities for creating sophisticated primer blends with increased
2102 breadth of detection. However, in developing such mixtures, it will be important to monitor
2103 for possible complicating interactions of primers with each other. As an example of ongoing
2104 development of mixtures, we found that addition of another primer to the ACeIN-26 set
2105 that was matched to a common subtype C lineage allowed improved detection of subtype C
2106 variants (S1 Report). In order to improve detection of subtype F, which was suboptimal with
2107 ACeIN-26, additional primer sets could be mixed to specifically target subtype F, though
2108 the ones we tried so far did not work well. It will be useful to explore the performance of
2109 broader primer mixtures in future work.

2110 Today rapid assays are available that can report infection efficiently, for example by detecting
2111 anti-HIV antibodies in oral samples—however, the nucleic acid-based method presented here
2112 has additional potential uses. We envision combining the RT-LAMP assay with simple
2113 point-of-care devices for purifying blood plasma⁵⁸⁴ and quantitative analysis of accumulation
2114 of fluorescent signals⁵⁹⁹. In one implementation of the technology, cell phones could be used
2115 to capture and analyze results, thereby minimizing equipment costs. Point-of-care devices are
2116 available facilitating the concentration of viral RNA from blood plasma or saliva⁵⁹⁹ to allow
2117 the detection of the 1000 RNA copy threshold that the WHO defines as virological treatment
2118 failure (World Health Organization, Consolidated ARV guidelines, June 2013). Together,
2119 these methods will allow assessment of parameters beyond just the presence/absence of

2120 infection. Quantitative RT-LAMP assays should allow tracking of responses to medication,
2121 detection in neonates (where immunological tests are confounded by presence of maternal
2122 antibody), and early detection before seroconversion.

2123 **5.7 Acknowledgments**

2124 We are grateful to members of the Bushman and Bau laboratory for help and suggestions.

CHAPTER 6: Conclusions and future directions

2126 In this dissertation, we described studies characterizing HIV-1 latency, expression and
2127 alternative splicing and host cell response to infection. We then developed point-of-care
2128 methods for the detection of infection and quantification of viral load. These projects suggest
2129 many avenues for continuing research.

2130 **6.1 Latency and integration location**

2131 In Chapter 2, we showed that the chromosomal location of integration affects proviral latency
2132 but the mechanisms appear to differ between cell culture models. Similarly a recent study
2133 of nine cell culture models found that no single model reliably predicted the performance of
2134 activating compounds in *ex vivo* tests of latently infected cells from HIV patients⁶⁰⁰. This
2135 suggests that either some cell culture models do not accurately reflect latency in patients or
2136 that there are diverse subsets of cells with differing mechanisms of latency within patients.

2137 Cell culture models are currently used to screen potentially therapeutic compounds^{136,600}. If
2138 some cell culture models are not representative of *in vivo* conditions then potential treatments
2139 may be discarded or marked for development erroneously. Further comparisons between
2140 additional cell culture models and additional replicates of existing models might allow
2141 discrimination between batch/lab effects and reveal patterns between models. Comparison
2142 with cells extracted from patients or infected lab animals might offer a gold standard
2143 comparison although it is difficult to obtain large amounts of cells and difficult to distinguish
2144 defective provirus from latent provirus in such populations.

2145 Various treatments are now being considered for the reactivation of latent provirus⁶⁰⁰. To
2146 further understand the mechanisms of these treatments, it would be informative to compare
2147 the features of latent provirus induced by a given treatment to latent viable provirus
2148 remaining uninduced. Repeated cell sorting and integration site sequencing might provide
2149 insight on mechanism. For example, one could first sort out cells with active provirus, then

2150 treat with the potential latency modulator and sort out cells with newly active provirus and
2151 then treat with a strong inducer or alternative stimuli and sort out cell with newly activated
2152 provirus. This would give subsets of cells where latent proviruses had been activated by
2153 treatment and cells with provirus which were not activated by treatment but still inducible.
2154 Synergies between treatments could be assessed and the location of integration sites could
2155 be determined and used to locate patterns of genomic features correlated with induction for
2156 each treatment.

2157 Current efforts at “shock and kill” therapy, inducing latent virus to activate and then
2158 eliminating infected cells, focus on histone deacetylase inhibitors. If there are diverse
2159 mechanisms of latency within patients then much of the latent reservoir may remain
2160 unactivated by single-target therapies. Clinical trials with histone deacetylase inhibitors
2161 have shown some small increases in viral RNA but little decrease in the latent reservoir of
2162 HIV^{410,601–603}. It appears that the majority of viable latent provirus from patient cells are
2163 not reactivated by current therapies¹³². These results are particularly worrisome because a
2164 functional cure for HIV will likely require a greater than 10,000-fold reduction of the latent
2165 reservoir⁶⁰⁴.

2166 In Chapter 2, we used publicly available genomic data. Perhaps there is some chromosomal
2167 feature with a strong association with latency but the data is not currently available or
2168 varies greatly between cell populations. More varieties of annotations are rapidly becoming
2169 available^{605–609}. Decreasing sequencing costs^{610–612} may also make it feasible to measure
2170 more epigenetic features in the exact cell population of interest. Repeating analyses similar
2171 to Chapter 2, perhaps by simply rerunning the reproducible report in Appendix A.2 with
2172 new data, would allow any new features to be monitored for correlations with latency.

2173 **6.2 HIV-1 alternative splicing**

2174 In Chapters 3 and 4, we showed that HIV RNA spliceforms are more diverse than previously
2175 appreciated and estimated the abundances of viral spliceforms. We also showed that splicing

2176 at some splice sites vary between host subjects, between cell types and over the course of
2177 infection. Further characterization of viral splicing would be beneficial to the study and
2178 treatment of HIV-1 especially as there were some technical limitations to our research that
2179 might be improved upon using current techniques.

2180 We studied HIV splicing using droplet PCR³²⁰ and a set of customized primer in Chapter
2181 3 and bulk sequencing of cellular mRNA in Chapter 4. Sequencing biases and difficulties
2182 determining full length transcripts from short reads hindered characterization of HIV
2183 sequencing. One alternative to these techniques is the targeted capture and enrichment^{613,614}
2184 of HIV-specific sequences. Using probes targeted to conserved regions of HIV, similar to
2185 finding conserved regions for primers as in Chapter 5, would allow for enrichment of viral
2186 reads without the biases induced by primer-based PCR while still allowing for efficient use
2187 of sequencing effort.

2188 The research in Chapter 3 was also limited by a short read bias in the PacBio sequencing.
2189 PacBio sequencing has improved⁶¹⁵ and additional long read sequencers have been devel-
2190 oped^{616–618}. In addition, Illumina MiSeq sequencers can now produce 25 million paired 300
2191 bp reads in a single run^{619,620} and better spliceform estimation methods are being devel-
2192 oped^{621,622}. These improved sequencing techniques might allow for more straightforward
2193 analysis of new samples and verification of our previous results.

2194 RNA transcribed antisense to the canonically expressed strand of HIV have been ob-
2195 served^{358,623–628}. These transcripts may be translated to proteins^{629,630} that trigger immune
2196 response in infected individuals^{292,628,629}. Our sequencing techniques were designed only for
2197 the HIV positive strand (Chapter 3) or did not preserve strand information (Chapter 4).
2198 Strand-specific sequencing^{631,632} of multiple HIV strains under varying cellular conditions
2199 would clarify the identity of these transcripts.

2200 Cryptic polypeptides encoding epitopes recognized through major histocompatibility complex
2201 type I also appear to be generated from alternative reading frames in the sense strand of

2202 the virus^{633,634}. Ribosome profiling^{635–637} of infected cells might reveal whether transcripts
2203 generated through alternative splicing or antisense expression are likely to be translated.
2204 These cryptic transcripts could offer new opportunities in vaccine design^{292,628,638,639} but
2205 first their abundance, identity and conservation across strains of HIV must be ascertained.

2206 We observed that splicing varies over the course of infection, between human subjects and
2207 between cell types. Further sampling could reveal additional patterns in these splicing
2208 changes.

2209 Long-lived reservoir of HIV infected cells exist in both macrophages^{253,640} and resting
2210 central memory CD4 T cells^{137,259,387,641,642}. It may be difficult to obtain enough viral
2211 RNA from resting CD4 cells⁶⁴¹ but macrophages provide an interesting target. Splicing
2212 changes due to differing abundances of splice factors have been reported in macrophages²⁸¹.
2213 Characterization of splicing in these important reservoirs might aid in the understanding of
2214 latency.

2215 We quantified the splicing of a single clinical isolate and showed unexpected diversity. Most
2216 previous studies of HIV splicing have been performed with lab-adapted strains³⁰⁶. Additional
2217 studies could determine if the high number of transcripts seen here is an anomaly and whether
2218 additional cryptic splice sites and novel proteins or epitopes exist. In addition, an important
2219 subset of HIV are the founder viruses transmitted between hosts^{643,644}. These viruses are
2220 not well studied and perhaps their splicing and gene expression differ from the rest of the
2221 viral swarm of infected patients. Comparisons to splicing in other retroviral taxa might
2222 highlight evolution and adaptation in this viral lineage.

2223 Disruption of RNA processing can drastically reduce viral replication^{342,645–648}. Small
2224 molecules that inhibit cellular SR splicing proteins and disrupt viral splicing show promise
2225 as antiretroviral therapies^{277,293,649,650}. Characterization of splicing in cells treated with
2226 splicing inhibitors could reveal potential escape pathways and optimal combinations of drug
2227 therapies.

2228 6.3 Host expression during HIV infection

2229 In Chapter 4, we saw many changes in host expression and splicing in HIV infected cells
2230 including intron retention and strong changes in apoptotic and innate immunity genes.
2231 We focused on generating a dense data set at a single time point and subject to allow
2232 discrimination of within-condition versus between-condition variation. Further sampling
2233 using more human subjects and time points, improved sequencing techniques, alternative
2234 culturing and extraction and more viral strains would clarify and extend these patterns.

2235 In our primary cell infections, only about 25% of cells were infected with HIV. This makes
2236 it difficult to distinguish between the responses of bystander and infected cells. In addition,
2237 changes in expression due to cellular response to infection are confounded with changes
2238 due to hijacking of cellular controls by the virus. For example, bystander cell death has
2239 been suggested as a major driver of HIV pathogenesis^{651,652} but our data do not make it
2240 clear whether bystander or infected cells are undergoing apoptosis. Cell pull-down with a
2241 labelled HIV strain⁵¹⁹ or an anti-Env antibody⁶⁵³ or flow cytometry with a labelled antibody
2242 targeting HIV antigen^{404,654} might allow the separation of bystander and infected cells.

2243 Additionally, abortive infections can drive cell death^{652,655} so our populations might be a
2244 mix of three responses; cells responding to a progressive infection, cells responding to an
2245 aborted infection and cells responding to neighbor cell infections. A useful control might be
2246 to infect cells with integrase-deficient virions to guarantee that all infections are aborted.
2247 This would provide a good measure of innate immune response and the effect of abortive
2248 infections undiluted by productive HIV infection and help to deconvolute the patterns seen
2249 in mixed populations.

2250 HIV infection appeared to increase the abundance of intronic sequences. We observed a
2251 significant decrease of nonsense-mediated decay-related genes so perhaps these transcripts
2252 escape degradation due to decreased cellular RNA surveillance. Alternatively, HIV Vpr
2253 protein has been reported to disrupt nuclear integrity and allow mixing of nuclear and

2254 cytoplasmic components²²⁹. These sequences might represent incompletely spliced mRNA
2255 that escaped into the cytoplasm before processing. Infection with a Vpr-deficient HIV virus
2256 and separate isolation of RNA from nuclear and cytoplasmic compartments^{656–658} would
2257 test these hypotheses.

2258 We saw that chimeric sequences were almost entirely derived from read-in or -out from
2259 viral long terminal repeats or splicing from the viral splice donor D4 to human acceptors.
2260 With this knowledge, we could use targeted amplification of these three sites, analogous to
2261 integration site sequencing^{269,270,423}, on cellular cDNA to get a much deeper and cleaner
2262 sampling of chimera formation. Comparison of these data to deeply sequenced integration
2263 site data from the same samples might reveal associations between integration location and
2264 chimera formation.

2265 MicroRNA are small RNAs that block translation through base pairing with comple-
2266 mentary mRNA^{659–661}. Viral derived microRNA, perhaps in part from Dicer processing
2267 of the structured trans-activation response element of HIV^{625,662–664}, may suppress HIV
2268 expression^{191,665,666} and inhibit apoptosis⁶⁶⁴ but the presence of such microRNA is controver-
2269 sial^{193,667}. HIV may suppress silencing by microRNA^{190–192} but this is also controversial¹⁹³.
2270 Cellular microRNA may have antiviral effects^{668,669} or be exploited by HIV to enhance
2271 replication^{670–674} or promote latency^{675,676} but there seems to be disagreement on which
2272 microRNA are involved among different studies⁶⁷⁷. High-throughput genome-wide assays of
2273 small RNA^{358,516} from primary cells infected with patient isolates would help clarify these
2274 debates.

2275 6.4 LAMP PCR and lab-on-a-chip

2276 In Chapter 5, we report a loop-mediated isothermal amplification system using primers
2277 optimized to detect most subtypes of HIV-1. An alternative to a single broadly targeted
2278 primer set would be to design separate primer sets targeted specifically to each subtype so
2279 that a positive amplification would then be able to discriminate viral subtype. Different viral

2280 subtypes can have different rates of disease progression^{678–681}, transmission dynamics^{682–684}
2281 and response to treatment^{685–687}. Simple low-cost devices with multiple reactions chambers
2282 could be used to both identify viral subtype, estimate viral load^{688,689} and allow more
2283 informed treatment decisions.

2284 A LAMP chip with subtype-specific primers would also allow the detection of intersubtype
2285 superinfections. Superinfection of a single individual with multiple distinct strains of HIV is
2286 common in high risk individuals^{579,690–693} and the general population⁶⁹⁴. Superinfection with
2287 a phenotypically different strain of HIV can lead to disease progression^{695–700} or drug resis-
2288 tance⁷⁰¹. Superinfection also allows recombination between divergent strains^{690,696,697,699,702}
2289 and this rapid exchange of genetic information can lead to more fit recombinant strains and
2290 worsen the global epidemic^{57,61,697,703,704}. LAMP detection of superinfection could allow
2291 early intervention and suppression in superinfected individuals.

2292 The techniques described in Chapter 5 also allow for rapid development of detection assays
2293 for novel pathogens. For example, in a recent outbreak in West Africa, Zaire ebolavirus
2294 has infected over 26,000 confirmed, probable and suspected cases and caused over 11,000
2295 reported deaths^{705–707}. Early detection and quarantine are essential to the control of this
2296 epidemic⁷⁰⁸. Amplification of Ebolavirus nucleic acid through polymerase chain reaction is
2297 the best diagnostic test currently available but the necessary resources are often not available
2298 in these resource-poor regions^{709,710}. Antigen-based tests are quicker and available at the
2299 point-of-care but are not as accurate or sensitive as polymerase chain reaction tests and are
2300 still in limited supply⁷¹⁰. Loop-mediated isothermal amplification offers the potential for
2301 rapid, sensitive and efficient detection of Ebolavirus RNA but available LAMP primers⁷¹¹ do
2302 not match the current outbreak strain. Using sequences from the recent outbreak^{705,712} and
2303 the methods described in Chapter 5, we designed primers to match all known Zaire ebolavirus
2304 (Figure 6.1). These primer combined with simple lab-on-a-chip devices for purifying blood
2305 plasma⁵⁸⁴ and imaging fluorescent signals^{599,688} could allow rapid point-of-care detection of
2306 Ebolavirus.

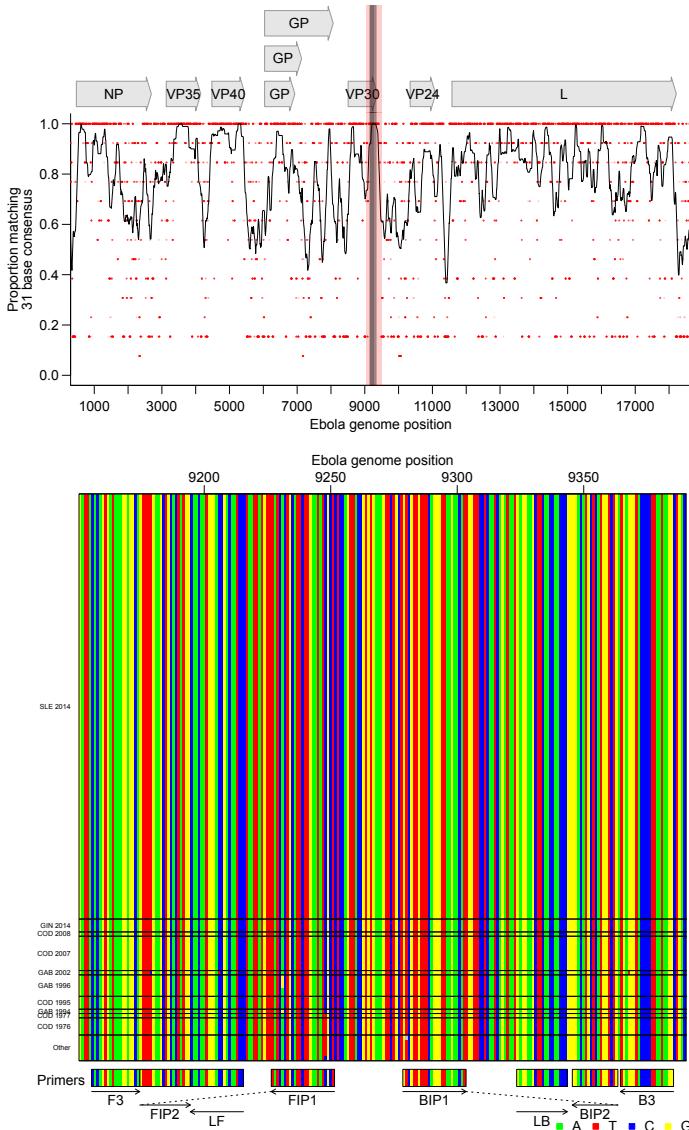


Figure 6.1: Bioinformatic analysis to design Ebolavirus RT-LAMP primers. A) Conservation of sequence in Ebolavirus. Ebolavirus genomes ($n = 131$) from Genbank and sequences from the recent Zaire Ebolavirus outbreak⁷⁰⁵ were aligned and conservation calculated. The x-axis shows the coordinate on the Ebola genome, the y-axis shows the proportion of sequences matching the consensus for each 21 base segment of the genome (red points). The black line shows a 101 base sliding average over these proportions. The vertical red shading shows the region targeted for LAMP primer design that was used as input into the EIKEN primer design tool and grey shading indicates the area covered by the optimized primer set. Numbering is relative to the Ebola Mayinga sequence. B) Aligned genomes, showing the locations of the LAMP primers. Sequences in the grey-shaded region in A are shown, with DNA bases color-coded as shown at the lower right. Each row indicates an Ebolavirus sequence and each column a base in that sequence. Horizontal lines separate Ebolavirus outbreaks (SLE: Seirra Leone, GIN: Guinea, COD: DR Congo, GAB: Gabon). Arrows indicate the strand targeted by each primer. Primers targeting the negative strand of the virus are shown as reverse compliments for ease of viewing.

2307 **6.4.1 Conclusions**

2308 These studies contribute to the study and treatment of HIV-1 by revealing aspects of latency,
2309 expression and host response. They highlight the importance of primary cell models and
2310 the effects that host cell can have on viral processes. With rapidly increasing sequencing
2311 throughput, studies like those presented here offer the opportunity for a deeper and broader
2312 understanding of HIV-1 biology and host response and further development of diagnostics
2313 and therapeutics.

APPENDIX A.1 : Generalized linear models of changes in use of mutually exclusive HIV-1 splice acceptors

Reads splicing from D1 to one of five mutually exclusive acceptors, D3, D4c, D4a, D4b, D5, and D5a, in three primers, 1.2, 1.3 and 1.4, were collected. Since these data are based on counts, we modeled them as Poisson distributed with an extra variance term allowing for additional variance using a quasi-Poisson generalized linear model with log link. We accounted for differences in sequencing effort by including the total number of D1 to mutually exclusive acceptors reads in each primer-sample as an offset. Differences in the read counts a) over time,b) between human donor and c) cell type were analyzed separately. A term was included for each acceptor and its interaction with the variable of interest. The models included primer and replicate terms and their individual interactions with acceptor to account for any confounding factors.

A.1.1 HOS vs T Cells

R command:

```
glm(count~offset(log(total)) + acceptor:primer + acceptor:  
    isHos  
    + acceptor, data = mutEx[mutEx$time == 48,],  
    family = 'quasipoisson')
```

Difference between HOS and T cells may be confounded by run differences between early sequencing and later sequencing. Verification by agarose gel (Figure 3.4b) suggest that these differences are likely biological.

Variable	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL	395	138 330				
acceptor	5	133 985	390	4345	9004	$<2.2 \times 10^{-16}$
acceptor:primer	12	751	378	3594	21.03	$<2.2 \times 10^{-16}$
acceptor:isHos	6	2466	372	1127	138.1	$<2.2 \times 10^{-16}$

So after accounting for primer-acceptor bias, the difference between HOS and T cells is significant.

The interesting terms in the model are:

Variable	Estimate	Std. Error	t value	Pr(> t)
acceptorA3:isHosTRUE	1.4717	0.065 86	22.35	$<2.2 \times 10^{-16}$
acceptorA4a:isHosTRUE	-0.9449	0.1246	-7.583	2.73×10^{-13}
acceptorA4b:isHosTRUE	-0.9285	0.1059	-8.767	$<2.2 \times 10^{-16}$
acceptorA4c:isHosTRUE	-1.228	0.1066	-11.51	$<2.2 \times 10^{-16}$
acceptorA5:isHosTRUE	0.090 82	0.026 08	3.483	0.000 555
acceptorA5a:isHosTRUE	0.6308	0.079 40	7.945	2.33×10^{-14}

So it appears A3 is up; A4c, A4a and A4b are down; A5 is up a little and A5a up in HOS.

A.1.2 HOS Over Time

R command:

```
glm(value~offset(log(total)) + acceptor + acceptor:primer
+ acceptor:time, data=mutEx[mutEx$isHos ,],
family = 'quasipoisson')
```

Looking only within HOS, we see a significant linear effect of time:

Variable	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL	53	17962				
acceptor	5	17710	48	252.2	6698	$<2.2 \times 10^{-16}$
acceptor:primer	12	18.0	36	234.2	2.834	0.01018
acceptor:time	6	217.8	30	16.4	68.65	3.57×10^{-16}

We are assuming that a particular acceptor will have the same change in all three primers here.

The interesting terms are:

Variable	Estimate	Std. Error	t value	Pr(> t)
acceptorA3:time	0.02477	0.001778	13.93	1.22×10^{-14}
acceptorA4a:time	-0.01621	0.002812	-5.765	2.69×10^{-6}
acceptorA4b:time	-0.02526	0.002271	-11.12	3.62×10^{-12}
acceptorA4c:time	0.015867	0.003050	5.202	1.32×10^{-5}
acceptorA5:time	-0.001918	0.0006313	-3.038	0.0049
acceptorA5a:time	0.004919	0.001969	2.499	0.0182

So A3, A4c and A5a increase over time and A4a, A4b and A5 decrease over time. All of these coefficients are with a log link and linear and so multiplicative. That means that for example A3 will increase 2.5%/hour ($\exp(.0247)$) or equivalently 81% (1.025^{24}) over 24hours.

A.1.3 Between Human Comparison

R command:

```
glm(value~offset(log(total)) + acceptor + acceptor:run
+ acceptor:primer + acceptor:subject ,
data=mutEx[!mutEx$ishos,], family = 'quasipoisson')
```

In humans, we added a term to account for any potential run bias between the three replicates. Subject refers to the seven human blood donors from which T cells were collected:

Variable	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL	377	128 430				
acceptor	5	126 446	372	1985	19 598	$<2.2 \times 10^{-16}$
acceptor:run	12	136	360	1849	8.792	1.77×10^{-14}
acceptor:primer	12	850	348	998	54.91	$<2.2 \times 10^{-16}$
acceptor:subject	36	597	312	401	12.86	$<2.2 \times 10^{-16}$

So after accounting for any run and primer bias, subject ID has a statistically significant effect on our observed counts. If we compare everything to subject 7, the interesting terms are:

Variable	Estimate	Std. Error	t value	Pr(> t)
acceptorA3:subject6	-0.001 399	0.072 86	-0.019	0.9847
acceptorA4a:subject6	-0.112 90	0.049 44	-2.284	0.023 07
acceptorA4b:subject6	-0.054 33	0.040 38	-1.345	0.1795
acceptorA4c:subject6	0.028 29	0.033 60	0.842	0.4005
acceptorA5:subject6	0.016 83	0.016 00	1.051	0.2939
acceptorA5a:subject6	-0.030 85	0.060 92	-0.506	0.6129
acceptorA3:subject5	-0.077 67	0.074 23	-1.046	0.2962
acceptorA4a:subject5	-0.1144	0.049 82	-2.296	0.0223
acceptorA4b:subject5	-0.0684	0.040 90	-1.672	0.0956
acceptorA4c:subject5	-0.085 85	0.034 75	-2.471	0.0140
acceptorA5:subject5	0.038 88	0.016 16	2.406	0.0167
acceptorA5a:subject5	0.078 77	0.060 38	1.304	0.1930
acceptorA3:subject4	-0.1849	0.095 78	-1.931	0.0544
acceptorA4a:subject4	0.071 86	0.057 91	1.241	0.2156
acceptorA4b:subject4	0.126 20	0.047 14	2.677	0.0078
acceptorA4c:subject4	-0.100 21	0.043 03	-2.329	0.0205
acceptorA5:subject4	-0.001 16	0.019 69	-0.059	0.9531
acceptorA5a:subject4	0.023 46	0.073 53	0.319	0.7499
acceptorA3:subject3	-0.003 51	0.086 65	-0.041	0.9677
acceptorA4a:subject3	0.071 07	0.055 64	1.277	0.2024
acceptorA4b:subject3	0.006 46	0.046 99	0.138	0.8907
acceptorA4c:subject3	-0.063 34	0.040 76	-1.554	0.1212
acceptorA5:subject3	0.010 52	0.018 87	0.557	0.5776
acceptorA5a:subject3	-0.070 95	0.072 85	-0.974	0.3309
acceptorA3:subject2	-0.2329	0.091 76	-2.539	0.0116
acceptorA4a:subject2	0.024 05	0.056 43	0.426	0.6702
acceptorA4b:subject2	0.1107	0.045 35	2.441	0.0152
acceptorA4c:subject2	0.021 76	0.039 52	0.551	0.5823
acceptorA5:subject2	-0.003 760	0.018 69	-0.201	0.8407
acceptorA5a:subject2	-0.1608	0.073 51	-2.187	0.0295
acceptorA3:subject1	0.095 36	0.065 56	1.454	0.1468
acceptorA4a:subject1	0.029 32	0.044 31	0.662	0.5087
acceptorA4b:subject1	-0.2144	0.038 43	-5.578	5.28×10^{-8}
acceptorA4c:subject1	-0.3974	0.033 85	-11.74	$<2.2 \times 10^{-16}$
acceptorA5:subject1	0.091 44	0.014 70	6.221	1.58×10^{-9}
acceptorA5a:subject1	0.027 47	0.055 94	0.491	0.6238

So there were small but significant effects between subjects especially between subject 1 and subjects 2–7. A potential confounder is that T cells were collected from apheresis product in

subject 1 and from whole blood in subjects 2–7 although why this would affect later assays is unknown.

APPENDIX A.2 : Reproducible report of HIV integration sites and latency analysis

A.2.1 Supplementary data

Additional File 2 is a gzipped csv file that includes a row for each uniquely mapped provirus and its surrounding genomic annotations. The csv file should have 12436 rows (excluding header) with 6252 expressed and 6184 latent proviruses.

```
integrationData <- read.csv("AdditionalFile2.csv.gz",
  stringsAsFactors = FALSE)

nrow(integrationData)

## [1] 12436

table(integrationData$isLatent)

##
##  FALSE   TRUE
##  6252   6184
```

A.2.2 Lasso regression

The lasso regressions take a while to run so I've turned down the number of cross validations here (set `eval=FALSE` below to completely skip this step). Leave one out and 480-fold cross validation were used in the paper but processing may take a few days without parallel processing. Lasso regression requires the R `glmnet` package.

```

notFitColumns <- c("id", "chr", "pos", "strand", "sample", "isLatent")

samples <- unique(as.character(integrationData$sample))

sampleMatrix <- do.call(cbind, lapply(samples, function(x)
  integrationData$sample ==
  x))

colnames(sampleMatrix) <- gsub(" ", "_", samples)

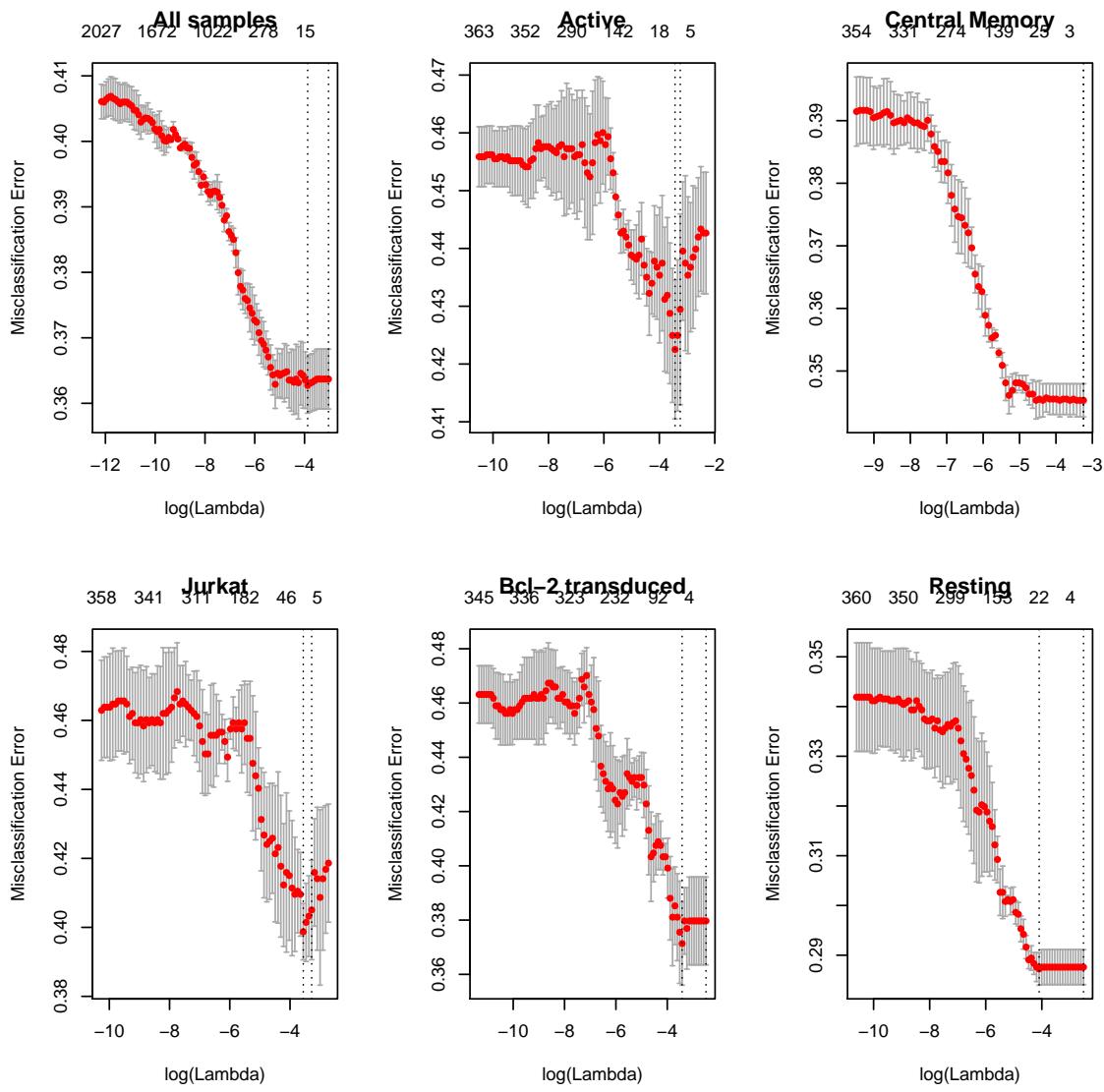
interact <- function(predMatrix, columns, addNames = NULL) {
  out <- do.call(cbind, lapply(1:ncol(columns), function(x)
    predMatrix *
    columns[, x]))
  if (!is.null(addNames)) {
    if (length(addNames) != ncol(columns)) {
      stop(simpleError("Names not same length as columns"))
    }
    colnames(out) <- sprintf("%s_%s", rep(addNames, each =
      ncol(predMatrix)),
      rep(colnames(predMatrix), length(addNames)))
  }
  return(out)
}

fitData <- as.matrix(integrationData[, !colnames(
  integrationData) %in%]

```

```
notFitColumns])  
  
fitData2 <- as.matrix(cbind(interact(fitData, sampleMatrix,  
colnames(sampleMatrix)),  
fitData, sampleMatrix))
```

```
library(glmnet)  
  
penalties <- rep(1, ncol(fitData2))  
  
penalties[ncol(fitData2) - (ncol(sampleMatrix):1) + 1] <- 0  
  
lassoFit <- cv.glmnet(fitData2, integrationData$isLatent,  
family = "binomial",  
type.measure = "class", nfolds = 3, penalty.factor =  
penalties)  
  
seperateFits <- lapply(samples, function(x) cv.glmnet(fitData[  
integrationData$sample ==  
x, ], integrationData$isLatent[integrationData$sample ==  
x], family = "binomial", type.measure = "class", nfolds =  
3))  
  
names(seperateFits) <- samples
```



A.2.3 Correlation

We looked for correlation between the genomic variables and expression status of the proviruses.

```
corMat <- apply(fitData, 2, function(x) sapply(samples,
  function(y) {
    selector <- integrationData$sample == y
```

```

    if (sd(x[selector]) == 0)
      return(0)

    isLatent <- integrationData[selector, "isLatent"]
    cor(as.numeric(isLatent), x[selector], method = "spearman"
    ")
  }))

quantile(cormat, seq(0, 1, 0.1))

##          0%           10%          20%          30%
## -0.185223020 -0.081555830 -0.048938130 -0.030895834
##          40%           50%          60%          70%
## -0.018053321 -0.005613895  0.003580982  0.017822483
##          80%           90%         100%
##  0.036694554  0.062003356  0.170642314

```

If we looked for genomic variables consistently correlated or anti-correlated with proviral expression status with an FDR q-value less than 0.01, no variable was significantly correlated in more than 3 samples.

```

pMat <- apply(fitData, 2, function(x) sapply(samples, function
(y) {
  selector <- integrationData$sample == y
  if (sd(x[selector]) == 0)
    return(NA)
  isLatent <- integrationData[selector, "isLatent"]
  cor.test(as.numeric(isLatent), x[selector], method =
  "spearman",

```

```

exact = FALSE)$p.value
}))

adjustPMat <- pMat

adjustPMat[, ] <- p.adjust(pMat, "fdr")

downPMat <- upPMat <- adjustPMat

downPMat[corMat > 0] <- 1

upPMat[corMat < 0] <- 1

table(apply(upPMat < 0.01 & !is.na(upPMat), 2, sum))

##
##      0     1     2     3
## 298   27   38   10

table(apply(downPMat < 0.01 & !is.na(downPMat), 2, sum))

##
##      0     1     2     3
## 216   36   63   58

```

A.2.4 RNA expression

We fit a logistic regression to a polynomial of log RNA-Seq reads within 5000 bases from Jurkat cells for the Jurkat sample and T cells for the rest.

```

rna <- ifelse(integrationData$sample == "Jurkat",
               integrationData$log_jurkatRNA,

```

```

integrationData$rna_5000)

rna2 <- rna^2

rna3 <- rna^3 # 

rna4 <- rna^4

glmData <- data.frame(isLatent = integrationData$isLatent ,
sample = integrationData$sample ,
rna, rna2, rna3, rna4)

glmMod <- glm(isLatent ~ sample * rna + sample * rna2 + sample
*
rna3 + sample * rna4, data = glmData, family = "binomial")

summary(glmMod)

## 

## Call:
## glm(formula = isLatent ~ sample * rna + sample * rna2 +
## sample *
##     rna3 + sample * rna4, family = "binomial", data =
## glmData)

## 

## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.2899   -0.9864   -0.8676    1.0960    1.6007

## 

## Coefficients:
##                                         Estimate Std. Error z value

```

```

## (Intercept)           1.7623655  0.2138859  8.240
## sampleBcl-2 transduced -2.1625912  0.7061524 -3.062
## sampleCentral Memory      -2.5010063  0.2437685 -10.260
## sampleJurkat            -2.0800202  0.2836871 -7.332
## sampleResting             0.7840481  0.3312247  2.367
## rna                      -0.6567268  0.2344422 -2.801
## rna2                     0.1387703  0.0770589  1.801
## rna3                     -0.0167219  0.0094076 -1.777
## rna4                     0.0007572  0.0003845  1.969
## sampleBcl-2 transduced:rna 0.5750186  0.6366537  0.903
## sampleCentral Memory:rna   0.9067758  0.2750955  3.296
## sampleJurkat:rna          0.5294036  0.3867163  1.369
## sampleResting:rna          0.0366276  0.3436248  0.107
## sampleBcl-2 transduced:rna2 -0.0369353  0.1878816 -0.197
## sampleCentral Memory:rna2   -0.2106715  0.0915492 -2.301
## sampleJurkat:rna2          -0.0766215  0.1641153 -0.467
## sampleResting:rna2          -0.0760450  0.1086998 -0.700
## sampleBcl-2 transduced:rna3 0.0032503  0.0213743  0.152
## sampleCentral Memory:rna3   0.0237064  0.0112661  2.104
## sampleJurkat:rna3          0.0042183  0.0263910  0.160
## sampleResting:rna3          0.0153132  0.0128711  1.190
## sampleBcl-2 transduced:rna4 -0.0002532  0.0008267 -0.306
## sampleCentral Memory:rna4   -0.0009877  0.0004627 -2.135
## sampleJurkat:rna4           0.0001725  0.0014215  0.121
## sampleResting:rna4          -0.0008049  0.0005119 -1.572
## Pr(>|z|)
## (Intercept) < 2e-16 ***

```

```

## sampleBcl-2 transduced          0.00219  **
## sampleCentral Memory           < 2e-16 ***
## sampleJurkat                  2.27e-13 ***
## sampleResting                 0.01793 *
## rna                          0.00509  **
## rna2                         0.07173 .
## rna3                         0.07549 .
## rna4                         0.04891 *
## sampleBcl-2 transduced:rna    0.36643
## sampleCentral Memory:rna      0.00098  ***
## sampleJurkat:rna              0.17101
## sampleResting:rna             0.91511
## sampleBcl-2 transduced:rna2   0.84415
## sampleCentral Memory:rna2     0.02138 *
## sampleJurkat:rna2            0.64059
## sampleResting:rna2           0.48419
## sampleBcl-2 transduced:rna3   0.87913
## sampleCentral Memory:rna3     0.03536 *
## sampleJurkat:rna3            0.87301
## sampleResting:rna3           0.23415
## sampleBcl-2 transduced:rna4   0.75939
## sampleCentral Memory:rna4     0.03280 *
## sampleJurkat:rna4            0.90339
## sampleResting:rna4           0.11585
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

## 
## (Dispersion parameter for binomial family taken to be 1)

## 
##      Null deviance: 17240    on 12435    degrees of freedom
## Residual deviance: 15874    on 12411    degrees of freedom
## AIC: 15924

## 
## Number of Fisher Scoring iterations: 4

```

A.2.5 Strand orientation

We used a Fisher's exact test to check if silent/inducible proviruses were enriched when integrated in the same strand orientation as cellular genes.

```

selector <- integrationData$inGene == 1

strandTable <- with(integrationData[selector, ], table(ifelse(
  isLatent,
  "Silent/Inducible", "Active"), ifelse(inGeneSameStrand ==
  1, "Same", "Diff"), sample))

apply(strandTable, 3, fisher.test)

## $Active

## 
##      Fisher's Exact Test for Count Data

## 
## data: array(newX[, i], d.call, dn.call)
## p-value = 0.06061
## alternative hypothesis: true odds ratio is not equal to 1

```

```
## 95 percent confidence interval:  
## 0.7219466 1.0081995  
  
## sample estimates:  
  
## odds ratio  
## 0.8532127  
  
##  
  
##  
  
## $`Bcl-2 transduced`  
  
##  
  
## Fisher's Exact Test for Count Data  
  
##  
  
## data: array(newX[, i], d.call, dn.call)  
## p-value = 2.177e-05  
  
## alternative hypothesis: true odds ratio is not equal to 1  
  
## 95 percent confidence interval:  
## 1.446896 2.872562  
  
## sample estimates:  
  
## odds ratio  
## 2.036148  
  
##  
  
##  
  
## $`Central Memory`  
  
##  
  
## Fisher's Exact Test for Count Data  
  
##  
  
## data: array(newX[, i], d.call, dn.call)  
## p-value = 0.2907
```

```

## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.9386167 1.2320238
## sample estimates:
## odds ratio
## 1.07529
##
##
## $Jurkat
##
## Fisher's Exact Test for Count Data
##
## data: array(newX[, i], d.call, dn.call)
## p-value = 0.1674
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.9207548 1.5699893
## sample estimates:
## odds ratio
## 1.202007
##
##
## $Resting
##
## Fisher's Exact Test for Count Data
##
## data: array(newX[, i], d.call, dn.call)

```

```

## p-value = 0.5732
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.7825231 1.1405158
## sample estimates:
## odds ratio
## 0.9447415

```

A.2.6 Acetylation

To reduce correlation between acetylation marks, we generated the first ten principal components of the acetylation data and ran a logistic regression against them. We compared the cross validated performance of this regression with a base model only including which dataset the integration site came from. The cross-validation here has been reduced for efficiency but 480-fold cross-validation was used in the paper.

```

acetyl <- integrationData[, !grepl("logDist", colnames(
  integrationData)) &
  grepl("ac", colnames(integrationData))]

acetylPCA <- princomp(acetyl)

cumsum(acetylPCA$sdev[1:10]^2/sum(acetylPCA$sdev^2))

##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
## 0.5947268 0.6786611 0.7267433 0.7610502 0.7833616 0.7964470
##      Comp.7      Comp.8      Comp.9      Comp.10
## 0.8093295 0.8215027 0.8299358 0.8372584

cv.glm <- function(model, K = nrow(thisData), subsets = NULL)
{

```

```

modelCall <- model$call

thisData <- eval(modelCall$data)

n <- nrow(thisData)

if (is.null(subsets))

  subsets <- split(1:n, sample(rep(1:K, length.out = n)))

  )

preds <- lapply(subsets, function(outGroup) {

  subsetData <- thisData[-outGroup, , drop = FALSE]

  predData <- thisData[outGroup, , drop = FALSE]

  thisModel <- modelCall

  thisModel$data <- subsetData

  return(predict(eval(thisModel), predData))

})

pred <- unlist(preds)[order(unlist(subsets))]

subsetId <- rep(1:K, sapply(subsets, length))[order(unlist
  (subsets))]

return(data.frame(pred, subsetId))
}

inData <- data.frame(isLatent = integrationData$isLatent ,
  sample = as.factor(integrationData$sample),
  acetylPCA$score[, 1:10])

modelPreds <- cv.glm(glm(isLatent ~ sample + Comp.1 + Comp.2 +
  Comp.3 + Comp.4 + Comp.5 + Comp.6 + Comp.7 + Comp.8 + Comp
  .9 +
  Comp.10, family = "binomial", data = inData), K = 5)

```

```

basePreds <- cv.glm(glm(isLatent ~ sample, family = "binomial
",
data = inData), subsets = split(1:nrow(inData),
modelPreds$subsetId),
K = 5)

modelCorrect <- sum((modelPreds$pred > 0) ==
integrationData$isLatent)
baseCorrect <- sum((basePreds$pred > 0) ==
integrationData$isLatent)

prop.test(c(baseCorrect, modelCorrect), rep(nrow(
integrationData),
2))

##
##      2-sample test for equality of proportions with
##      continuity correction
##
## data: c(baseCorrect, modelCorrect) out of rep(nrow(
## integrationData), 2)
## X-squared = 0.00017372, df = 1, p-value = 0.9895
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.01187726 0.01219890
## sample estimates:
## prop 1     prop 2
## 0.6362978 0.6361370

```

A.2.7 Gene deserts

We used Fisher's exact test to look for an association between integration outside a gene and proviral expression status.

```
geneTable <- table(integrationData$isLatent ,  
                    integrationData$inGene ,  
                    integrationData$sample)  
  
apply(geneTable , 3 , fisher.test)  
  
## $Active  
##  
##      Fisher's Exact Test for Count Data  
##  
## data: array(newX[, i] , d.call , dn.call)  
## p-value < 2.2e-16  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.3629548 0.5446204  
## sample estimates:  
## odds ratio  
##  0.4452621  
##  
##  
## $`Bcl-2 transduced`  
##  
##      Fisher's Exact Test for Count Data  
##  
## data: array(newX[, i] , d.call , dn.call)
```

```
## p-value = 0.1052
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.9203418 2.3478599
## sample estimates:
## odds ratio
## 1.472224
##
##
## $`Central Memory`
##
## Fisher's Exact Test for Count Data
##
## data: array(newX[, i], d.call, dn.call)
## p-value = 0.7803
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.8525329 1.1253952
## sample estimates:
## odds ratio
## 0.9791165
##
##
## $Jurkat
##
## Fisher's Exact Test for Count Data
##
```

```

## data: array(newX[, i], d.call, dn.call)
## p-value = 0.5443
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.7909269 1.6167285
## sample estimates:
## odds ratio
## 1.127836
##
##
## $Resting
##
## Fisher's Exact Test for Count Data
##
## data: array(newX[, i], d.call, dn.call)
## p-value = 3.071e-08
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.4384828 0.6864112
## sample estimates:
## odds ratio
## 0.5500205

```

We used a two-sample t-test to investigate whether there was a significant difference in distance to the nearest gene between expressed and silent/inducible proviruses integrated outside genes.

```

geneDistData <- integrationData[!integrationData$inGene , c(
  "isLatent",
  "logDist_nearest", "sample")]

by(geneDistData, geneDistData$sample, function(x) t.test(
  logDist_nearest ~
  isLatent, data = x))

## geneDistData$sample: Active

##
##      Welch Two Sample t-test
##
## data: logDist_nearest by isLatent
## t = -2.4539, df = 287.73, p-value = 0.01472
## alternative hypothesis: true difference in means is not
## equal to 0
## 95 percent confidence interval:
## -0.80738340 -0.08867607
## sample estimates:
## mean in group FALSE mean in group TRUE
## 9.608737 10.056767
##
## -----
## geneDistData$sample: Bcl-2 transduced
##
##      Welch Two Sample t-test
##
## data: logDist_nearest by isLatent
## t = 0.40978, df = 86.2, p-value = 0.683

```

```

## alternative hypothesis: true difference in means is not
## equal to 0

## 95 percent confidence interval:
## -0.6309351 0.9586004

## sample estimates:

## mean in group FALSE mean in group TRUE
## 9.036872 8.873039

## -----
## geneDistData$sample: Central Memory

## Welch Two Sample t-test

## data: logDist_nearest by isLatent
## t = -0.07188, df = 861.61, p-value = 0.9427

## alternative hypothesis: true difference in means is not
## equal to 0

## 95 percent confidence interval:
## -0.2371374 0.2203819

## sample estimates:

## mean in group FALSE mean in group TRUE
## 10.19225 10.20063

## -----
## geneDistData$sample: Jurkat

## Welch Two Sample t-test

```

```

## 

## data: logDist_nearest by isLatent
## t = -1.8217, df = 139.56, p-value = 0.07064
## alternative hypothesis: true difference in means is not
## equal to 0
## 95 percent confidence interval:
## -1.26342086 0.05167979
## sample estimates:
## mean in group FALSE mean in group TRUE
## 9.925782 10.531652
##
## -----
## geneDistData$sample: Resting
##
## Welch Two Sample t-test
##
## data: logDist_nearest by isLatent
## t = -5.1275, df = 193.49, p-value = 7.096e-07
## alternative hypothesis: true difference in means is not
## equal to 0
## 95 percent confidence interval:
## -1.2687917 -0.5638568
## sample estimates:
## mean in group FALSE mean in group TRUE
## 9.489931 10.406255

```

To check for a relationship between silent/inducible status and distance to CpG islands, we

used a two sample t-test on the logged distance and saw a significant difference between silent/inducible and expressed proviruses (before accounting for a correlation between being near CpG islands and in genes)

```
t.test(integrationData$logDist_cpg ~ integrationData$isLatent)

##
##      Welch Two Sample t-test
##
## data: integrationData$logDist_cpg by
##       integrationData$isLatent
## t = -2.0233, df = 12381, p-value = 0.04306
## alternative hypothesis: true difference in means is not
## equal to 0
## 95 percent confidence interval:
## -0.105657514 -0.001675563
## sample estimates:
## mean in group FALSE   mean in group TRUE
##                 10.16362             10.21728

sapply(unique(integrationData$sample), function(x) with(
  integrationData[integrationData$sample ==
    x, ], p.adjust(t.test(logDist_cpg ~ isLatent)$p.value,
    method = "bonferroni",
    n = 5)))

##          Active     Central Memory           Jurkat
##          0.512040457     1.000000000     1.000000000
## Bcl-2 transduced           Resting
##          1.000000000     0.005866539
```

Many CpG islands are found near genes. To account for this relationship, we used an ANOVA test including whether the integration site was inside a gene prior to including CpG islands. After including integration inside genes, CpG islands were not significantly associated with silent/inducible status of the proviruses with all samples grouped or individually after Bonferroni correction for multiple comparisons.

```

anova(with(integrationData, glm(isLatent ~ I(logDist_nearest
==

0) + logDist_cpg, family = "binomial")), test = "Chisq")

## Analysis of Deviance Table

##
## Model: binomial, link: logit

##
## Response: isLatent

##
## Terms added sequentially (first to last)

##
##                                     Df Deviance Resid. Df Resid. Dev
## NULL                               12435      17240
## I(logDist_nearest == 0)    1     26.2682    12434      17213
## logDist_cpg                      1      1.1328    12433      17212
##                                     Pr(>Chi)
## NULL
## I(logDist_nearest == 0) 2.971e-07 ***
## logDist_cpg                  0.2872
## ---
## Signif. codes:

```

```

## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sapply(unique(integrationData$sample), function(x) {
  p.adjust(anova(with(integrationData[integrationData$sample
  ==

x, ], glm(isLatent ~ I(logDist_nearest == 0) +
logDist_cpg,
family = "binomial")), test = "Chisq")["logDist_cpg",
"Pr(>Chi)"], method = "bonferroni", n = 5)
})

##          Active      Central Memory           Jurkat
## 1.0000000 1.0000000
## Bcl-2 transduced             Resting
## 1.0000000 0.2007788

```

A.2.8 Alphoid repeats

When analyzing repetitive elements, we treated each read as an independent observation and included reads with multiple alignments to the genome. Additional File 3 is a gzipped csv file containing a row for each read with multiple alignments and one row for each dereplicated integration site with a single alignment with the count variable indicating the number of reads dereplicated to that integration site. There should be 26,190 rows (excluding header) with 14,494 rows of expressed provirus and 11,696 rows of silent/inducible provirus.

```

repeats <- read.csv("AdditionalFile3.csv.gz", check.names =
FALSE,
stringsAsFactors = FALSE)

nrow(repeats)

```

To analyze whether there was an association between proviral expression status and integration within alphoid repeats, we used Fisher's exact test with a Bonferroni correction for five samples. For comparison, we looked at the association between proviral expression and the other repeats in the RepeatMasker database. We did not Bonferroni correct for the multiple repeat types so that the repeats could be compared with the analysis of alphoid repeats (for which we had an a priori hypothesis for an association with latency).

```
dummyX <- rep(c(TRUE, FALSE), 2)

dummyY <- rep(c(TRUE, FALSE), each = 2)

repeatData <- repeats[, !colnames(repeats) %in%
  notRepeatColumns]

repeatData <- repeatData[, apply(repeatData, 2, sum) > 0]

testRepeats <- function(x, repeats) {
  sapply(samples, function(thisSample, repeats) {
    selector <- repeats$sample == thisSample
    repLatent <- rep(repeats$isLatent[selector],
      repeats$count[selector])
    repRepeat <- rep(x[selector], repeats$count[selector])
```

```

fisher.test(table(c(dummyX, repLatent), c(dummyY,
repRepeat)) -
1)p.value
}, repeats)
}

repeatPs <- apply(repeatData, 2, testRepeats, repeats[, notRepeatColumns])

table(apply(repeatPs * 5 < 0.05, 2, sum))

##
##    0     1     2     3
## 611   76   15     1

which(apply(repeatPs * 5 < 0.05, 2, sum) >= 3)

## ALR/Alpha
##          178

p.adjust(repeatPs[, "ALR/Alpha"], "bonferroni")

##          Active   Central Memory        Jurkat
## 5.026890e-02   3.940207e-03   1.027189e-08
## Bcl-2 transduced           Resting
## 1.000000e+00   2.424896e-02

```

A.2.9 Neighbors

We looked at all pairs of viruses on the same chromosome separated by no more than a given distance, e.g. 100 bases, either with all samples pooled or split between within sample

pairs or between sample pairs.

```
allNeighbors <- data.frame(id1 = 0, id2 = 0)[0, ]  
  
ids <- 1:nrow(integrationData)  
  
for (chr in unique(integrationData$chr)) {  
  chrSelector <- integrationData$chr == chr  
  neighborPairs <- data.frame(id1 = rep(ids[chrSelector],  
    sum(chrSelector)),  
    id2 = rep(ids[chrSelector], each = sum(chrSelector)))  
  neighborPairs <- neighborPairs[neighborPairs$id1 <  
    neighborPairs$id2,  
    ]  
  allNeighbors <- rbind(allNeighbors, neighborPairs)  
}  
  
allNeighbors$dist <- abs(integrationData$pos[allNeighbors$id1]  
-  
  integrationData$pos[allNeighbors$id2])  
  
allNeighbors$latent1 <- integrationData$isLatent [  
  allNeighbors$id1]  
  
allNeighbors$latent2 <- integrationData$isLatent [  
  allNeighbors$id2]  
  
allNeighbors$sample1 <- integrationData$sample [  
  allNeighbors$id1]  
  
allNeighbors$sample2 <- integrationData$sample [  
  allNeighbors$id2]
```

The expected number of matching pairs was calculated as $\sum_{j \in \text{samples}} n_{j,d}(\theta_{j,d}\theta_{\neg j,d} + (1 - \theta_{j,d})(1 - \theta_{\neg j,d}))$ for between sample, $\sum_{j \in \text{samples}} n_{j,d}(\theta_{j,d}^2 + (1 - \theta_{j,d})^2)$ for within sample and $n_d(\theta_d^2 + (1 - \theta_d)^2)$ for all pairs, where $n_{j,d}$ is the number of pairs of proviruses separated by no more than d base pairs where the first provirus is from sample j , $\theta_{j,d}$ is the proportion of silent/inducible proviruses in sample j appearing in at least one pair of proviruses separated by less than d base pairs and $\neg j$ means all samples except sample j .

```

dists <- unique(round(10^seq(1, 6, 1)))

pairings <- do.call(rbind, lapply(dists, function(x,
allNeighbors) {
  inSelector <- allNeighbors$dist <= x &
  allNeighbors$sample1 ==
  allNeighbors$sample2
  outSelector <- allNeighbors$dist <= x &
  allNeighbors$sample1 != allNeighbors$sample2
  allSelector <- allNeighbors$dist <= x
  out <- data.frame(dist = x, observedIn = sum(allNeighbors[
    inSelector,
    "latent1"] == allNeighbors[inSelector, "latent2"]),
    observedOut = sum(allNeighbors[outSelector,
    "latent1"] == allNeighbors[outSelector, "latent2"]),
    observedAll = sum(allNeighbors[allSelector, "latent1"])
    ==
    allNeighbors[allSelector, "latent2"]), totalIn =
  
```

```

        sum(inSelector) ,

totalOut = sum(outSelector), totalAll = sum(
    allSelector))

out$expectedIn <- sum(with(allNeighbors[inSelector, ],
    sapply(samples,
        function(x) {

            inLatent <- c(latent1[sample1 == x], latent2[
                sample2 ==
                x])[!duplicated(c(id1[sample1 == x], id2[
                    sample2 ==
                    x]))]

            if (length(inLatent) == 0) return(0)
            return(sum(sample1 == x) * (mean(inLatent)^2 +
                mean(!inLatent)^2))
        })))
}

out$expectedOut <- sum(with(allNeighbors[outSelector, ],
    sapply(samples, function(x) {

        inLatent <- c(latent1[sample1 == x], latent2[
            sample2 ==
            x])[!duplicated(c(id1[sample1 == x], id2[
                sample2 ==
                x]))]

        outLatent <- c(latent1[sample1 != x], latent2[
            sample2 !=
            x])[!duplicated(c(id1[sample1 != x], id2[
                sample2 !=
                x]))]
    })))
}

```

```

    if (length(inLatent) == 0) return(0)

    return(sum(sample1 == x) * (mean(inLatent) * mean(
        outLatent) +
        mean(!inLatent) * mean(!outLatent)))
    }))

out$expectedAll <- sum(with(allNeighbors[allSelector, ],
{
    allLatent <- c(latent1, latent2)[!duplicated(c(id1
        ,
        id2))]

    return(length(latent1) * (mean(allLatent)^2 + mean
        (!allLatent)^2))
})
return(out)
}, allNeighbors))

rownames(pairings) <- pairings$dist

```

To look for more matches than expected by random pairing between neighboring proviruses, we used a one sample Z-test of proportion to compare the observed number of matching pairs with the expected proportion of pairs.

```

combinations <- c(All = "All", `Between sample` = "Out", `
    Within sample` = "In")

lapply(combinations, function(x, pairing) {
    vars <- sprintf(c("observed%s", "expected%s", "total%s"),
        x)
    expectedProb <- pairing[, vars[2]]/pairing[, vars[3]]

```

```

prop.test(pairing[, vars[1]], pairing[, vars[3]], p =
  expectedProb)
}, pairings["100", ])

## $All

##
##      1-sample proportions test with continuity correction
##

## data:  pairing[, vars[1]] out of pairing[, vars[3]], null
## probability expectedProb

## X-squared = 13.002, df = 1, p-value = 0.0003111
## alternative hypothesis: true p is not equal to 0.5000141
## 95 percent confidence interval:
##  0.5586837 0.6962353
## sample estimates:
##   p
## 0.63
##
##
## $`Between sample`

##
##      1-sample proportions test with continuity correction
##

## data:  pairing[, vars[1]] out of pairing[, vars[3]], null
## probability expectedProb

## X-squared = 0.21919, df = 1, p-value = 0.6397
## alternative hypothesis: true p is not equal to 0.4836763
## 95 percent confidence interval:

```

```

##  0.3570532 0.5572662

## sample estimates:

##          p
## 0.4554455

##
## 

## $`Within sample`


##      1-sample proportions test with continuity correction

## 

## data: pairing[, vars[1]] out of pairing[, vars[3]], null
## probability expectedProb
## X-squared = 24.446, df = 1, p-value = 7.644e-07
## alternative hypothesis: true p is not equal to 0.5561437
## 95 percent confidence interval:
##  0.7140170 0.8776751
## sample estimates:
##          p
## 0.8080808

```

A.2.10 Compiling this document

This document was generated using R's Sweave function (<http://en.wikipedia.org/wiki/Sweave>). If you would like to regenerate this document, download Additional Files 2, 3 and 4 from Sherrill-Mix et al.³⁸⁶ and make sure the files are all in the same directory and named AdditionalFile2.csv.gz, AdditionalFile3.csv.gz and AdditionalFile4.Rnw. Then compile by going to that directory and using the commands:

R CMD Sweave AdditionalFile4.Rnw

pdflatex AdditionalFile4.tex

Note that you will need R and L^AT_EX (and the R package glmnet if you would like to rerun the lasso regressions) installed.

BIBLIOGRAPHY

- [1] MS Gottlieb, HM Schanker, PT Fan, A Saxon, JD Weisman and I Pozalski. 1981. Pneumocystis pneumonia—Los Angeles. *MMWR Morb Mortal Wkly Rep*, 30:250–252
- [2] A Friedman-Kien, L Laubenstein, M Marmor, K Hymes, J Green, A Ragaz, J Gottlieb, F Muggia, R Demopoulos and M Weintraub. 1981. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men—New York City and California. *MMWR Morb Mortal Wkly Rep*, 30:305–308
- [3] KB Hymes, T Cheung, JB Greene, NS Prose, A Marcus, H Ballard, DC William and LJ Laubenstein. 1981. Kaposi's sarcoma in homosexual men—a report of eight cases. *Lancet*, 2:598–600. doi: 10.1016/S0140-6736(81)92740-9
- [4] H Masur, MA Michelis, JB Greene, I Onorato, RA Stouwe, RS Holzman, G Wormser, L Brettman, M Lange et al. 1981. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. *N Engl J Med*, 305:1431–1438. doi: 10.1056/NEJM198112103052402
- [5] FP Siegal, C Lopez, GS Hammer, AE Brown, SJ Kornfeld, J Gold, J Hassett, SZ Hirschman, C Cunningham-Rundles and BR Adelsberg. 1981. Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesion. *N Engl J Med*, 305:1439–1444. doi: 10.1056/NEJM198112103052403
- [6] MS Gottlieb, R Schroff, HM Schanker, JD Weisman, PT Fan, RA Wolf and A Saxon. 1981. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med*, 305:1425–1431. doi: 10.1056/NEJM198112103052401
- [7] Y Laor and RA Schwartz. 1979. Epidemiologic aspects of American Kaposi's sarcoma. *J Surg Oncol*, 12:299–303. doi: 10.1002/jso.2930120403
- [8] MB Klein, FA Pereira and I Kantor. 1974. Kaposi Sarcoma complicating systemic lupus erythematosus treated with immunosuppression. *Arch Dermatol*, 110:602–604. doi: 10.1001/archderm.1974.01630100058014
- [9] BD Myers, E Kessler, J Levi, A Pick, JB Rosenfeld and P Tikvah. 1974. Kaposi sarcoma in kidney transplant recipients. *Arch Intern Med*, 133:307–311. doi: 10.1001/archinte.1974.00320140145017
- [10] SB Kapadia and JR Krause. 1977. Kaposi's sarcoma after long-term alkylating agent therapy for multiple myeloma. *South Med J*, 70:1011–1013
- [11] B Safai and RA Good. 1981. Kaposi's sarcoma: a review and recent developments. *CA Cancer J Clin*, 31:2–12. doi: 10.3322/canclin.31.1.2

- [12] Y Chang, E Cesarman, MS Pessin, F Lee, J Culpepper, DM Knowles and PS Moore. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*, 266:1865–1869. doi: 10.1126/science.7997879
- [13] F Sitas, H Carrara, V Beral, R Newton, G Reeves, D Bull, U Jentsch, R Pacella-Norman, D Bourboulia et al. 1999. Antibodies against human herpesvirus 8 in black South African patients with cancer. *N Engl J Med*, 340:1863–1871. doi: 10.1056/NEJM199906173402403
- [14] BA Burke and RA Good. 1973. Pneumocystis carinii infection. *Medicine (Baltimore)*, 52:23–51
- [15] WT Hughes. 1977. Pneumocystis carinii pneumonia. *N Engl J Med*, 297:1381–1383. doi: 10.1056/NEJM197712222972505
- [16] J Gerstoft, A Malchow-Møller, I Bygbjerg, E Dickmeiss, C Enk, P Halberg, S Haahr, M Jacobsen, K Jensen et al. 1982. Severe acquired immunodeficiency in European homosexual men. *Br Med J (Clin Res Ed)*, 285:17–19
- [17] H Masur, MA Michelis, GP Wormser, S Lewin, J Gold, ML Tapper, J Giron, CW Lerner, D Armstrong et al. 1982. Opportunistic infection in previously healthy women. Initial manifestations of a community-acquired cellular immunodeficiency. *Ann Intern Med*, 97:533–539
- [18] A Ammann, M Cowan, D Wara, H Goldman, H Perkins, R Lanzerotti, J Gullett, A Duff, S Dritz and J Chin. 1982. Possible transfusion-associated acquired immune deficiency syndrome (AIDS) — California. *MMWR Morb Mortal Wkly Rep*, 31:652–654
- [19] N Ehrenkranz, J Rubini, R Gunn, C Horsburgh, T Collins, U Hasiba, W Hathaway, W Doig, R Hopkins and J Elliott. 1982. Pneumocystis carinii pneumonia among persons with hemophilia A. *MMWR Morb Mortal Wkly Rep*, 31:365–367
- [20] MC Poon, A Landay, J Alexander, W Birch, M Eyster, H Al-Mondhiry, J Ballard, E Witte, C Hayes et al. 1982. Update on acquired immune deficiency syndrome (AIDS) among patients with hemophilia A. *MMWR Morb Mortal Wkly Rep*, 31:644–6, 652
- [21] JB Greene, GS Sidhu, S Lewin, JF Levine, H Masur, MS Simberkoff, P Nicholas, RC Good, SB Zolla-Pazner et al. 1982. Mycobacterium avium-intracellulare: a cause of disseminated life-threatening infection in homosexuals and drug abusers. *Ann Intern Med*, 97:539–546
- [22] R O'Reilly, D Kirkpatrick, CB Small, R Klein, H Keltz, G Friedland, K Bromberg, S Fikrig, H Mendez et al. 1982. Unexplained immunodeficiency and opportunistic infections in infants—New York, New Jersey, California. *MMWR Morb Mortal Wkly Rep*, 31:665–667
- [23] S Fannin, M Gottlieb, J Weisman, E Rogolsky, T Prendergast, J Chin, A Friedman-

- Kien, L Laubenstein, S Friedman and R Rothenberg. 1982. A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. *MMWR Morb Mortal Wkly Rep*, 31: 305–307
- [24] C Harris, CB Small, G Friedland, R Klein, B Moll, E Emeson, I Spigland, N Steigbigel, R Reiss et al. 1983. Immunodeficiency among female sexual partners of males with acquired immune deficiency syndrome (AIDS) — New York. *MMWR Morb Mortal Wkly Rep*, 31:697–698
 - [25] F Barré-Sinoussi, JC Chermann, F Rey, MT Nugeyre, S Chamaret, J Gruest, C Dauguet, C Axler-Blin, F Vézinet-Brun et al. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220: 868–871
 - [26] RC Gallo, PS Sarin, EP Gelmann, M Robert-Guroff, E Richardson, VS Kalyanaraman, D Mann, GD Sidhu, RE Stahl et al. 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, 220:865–867. doi: 10.1126/science.6601823
 - [27] M Popovic, MG Sarngadharan, E Read and RC Gallo. 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, 224:497–500. doi: 10.1126/science.6200935
 - [28] JA Levy, AD Hoffman, SM Kramer, JA Landis, JM Shimabukuro and LS Oshiro. 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science*, 225:840–842. doi: 10.1126/science.6206563
 - [29] RC Gallo, SZ Salahuddin, M Popovic, GM Shearer, M Kaplan, BF Haynes, TJ Parker, R Redfield, J Oleske and B Safai. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, 224: 500–503. doi: 10.1126/science.6200936
 - [30] MG Sarngadharan, M Popovic, L Bruch, J Schüpbach and RC Gallo. 1984. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science*, 224:506–508. doi: 10.1126/science.6324345
 - [31] B Safai, MG Sarngadharan, JE Groopman, K Arnett, M Popovic, A Sliski, J Schüpbach and RC Gallo. 1984. Seroepidemiological studies of human T-lymphotropic retrovirus type III in acquired immunodeficiency syndrome. *Lancet*, 1:1438–1440. doi: 10.1016/S0140-6736(84)91933-0
 - [32] N Clumeck, F Mascart-Lemone, J de Maubeuge, D Brenez and L Marcelis. 1983. Acquired immune deficiency syndrome in Black Africans. *Lancet*, 1:642. doi: 10.1016/S0140-6736(83)91808-1
 - [33] N Clumeck, J Sonnet, H Taelman, S Cran and P Henrivaux. 1984. Acquired immune

- deficiency syndrome in Belgium and its relation to Central Africa. *Ann N Y Acad Sci*, 437:264–269. doi: 10.1111/j.1749-6632.1984.tb37144.x
- [34] P Van de Perre, D Rouvroy, P Lepage, J Bogaerts, P Kestelyn, J Kayihigi, AC Hekker, JP Butzler and N Clumeck. 1984. Acquired immunodeficiency syndrome in Rwanda. *Lancet*, 2:62–65. doi: 10.1016/S0140-6736(84)90240-X
- [35] P Piot, TC Quinn, H Taelman, FM Feinsod, KB Minlangu, O Wobin, N Mbendi, P Mazebo, K Ndangi and W Stevens. 1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *Lancet*, 2:65–69. doi: 10.1016/S0140-6736(84)90241-1
- [36] JN Nkengasong, W Janssens, L Heyndrickx, K Fransen, PM Ndumbe, J Motte, A Leonaers, M Ngolle, J Ayuk and P Piot. 1994. Genotypic subtypes of HIV-1 in Cameroon. *AIDS*, 8:1405–1412
- [37] J Louwagie, W Janssens, J Mascola, L Heyndrickx, P Hegerich, G van der Groen, FE McCutchan and DS Burke. 1995. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol*, 69: 263–271
- [38] N Vidal, M Peeters, C Mulanga-Kabeya, N Nzilambi, D Robertson, W Ilunga, H Sema, K Tshimanga, B Bongo and E Delaporte. 2000. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol*, 74:10498–10507. doi: 10.1128/JVI.74.22.10498-10507.2000
- [39] A Rambaut, DL Robertson, OG Pybus, M Peeters and EC Holmes. 2001. Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature*, 410:1047–1048. doi: 10.1038/35074179
- [40] C Yang, B Dash, SL Hanna, HS Frances, N Nzilambi, RC Colebunders, M St Louis, TC Quinn, TM Folks and RB Lal. 2001. Predominance of HIV type 1 subtype G among commercial sex workers from Kinshasa, Democratic Republic of Congo. *AIDS Res Hum Retroviruses*, 17:361–365. doi: 10.1089/08892220150503726
- [41] ML Kalish, KE Robbins, D Pieniazek, A Schaefer, N Nzilambi, TC Quinn, ME St Louis, AS Youngpairoj, J Phillips et al. 2004. Recombinant viruses and early global HIV-1 epidemic. *Emerg Infect Dis*, 10:1227–1234. doi: 10.3201/eid1007.030904
- [42] SS Frøland, P Jenum, CF Lindboe, KW Webring, PJ Linnestad and T Böhmer. 1988. HIV-1 infection in Norwegian family before 1970. *Lancet*, 1:1344–1345. doi: 10.1016/S0140-6736(88)92164-2
- [43] RF Garry, MH Witte, AA Gottlieb, M Elvin-Lewis, MS Gottlieb, CL Witte, SS Alexander, WR Cole and W Drake, Jr. 1988. Documentation of an AIDS virus infection in the United States in 1968. *JAMA*, 260:2085–2087. doi: 10.1001/jama.1988.03410140097031

- [44] IC Bygbjerg. 1983. AIDS in a Danish surgeon (Zaire, 1976). *Lancet*, 1:925. doi: 10.1016/S0140-6736(83)91348-X
- [45] J Vandepitte, R Verwilghen and P Zachee. 1983. AIDS and cryptococcosis (Zaire, 1977). *Lancet*, 1:925–926. doi: 10.1016/S0140-6736(83)91349-1
- [46] AJ Nahmias, J Weiss, X Yao, F Lee, R Kodsi, M Schanfield, T Matthews, D Bolognesi, D Durack and A Motulsky. 1986. Evidence for human infection with an HTLV III/LAV-like virus in Central Africa, 1959. *Lancet*, 1:1279–1280. doi: 10.1016/S0140-6736(86)91422-4
- [47] T Zhu, BT Korber, AJ Nahmias, E Hooper, PM Sharp and DD Ho. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature*, 391:594–597. doi: 10.1038/35400
- [48] M Worobey, M Gemmel, DE Teuwen, T Haselkorn, K Kunstman, M Bunce, JJ Muyembe, JMM Kabongo, RM Kalengayi et al. 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*, 455:661–664. doi: 10.1038/nature07390
- [49] B Korber, M Muldoon, J Theiler, F Gao, R Gupta, A Lapedes, BH Hahn, S Wolinsky and T Bhattacharya. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science*, 288:1789–1796. doi: 10.1126/science.288.5472.1789
- [50] M Salemi, K Strimmer, WW Hall, M Duffy, E Delaporte, S Mboup, M Peeters and AM Vandamme. 2001. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *FASEB J*, 15:276–278. doi: 10.1096/fj.00-0449fje
- [51] PM Sharp, E Bailes, RR Chaudhuri, CM Rodenburg, MO Santiago and BH Hahn. 2001. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B Biol Sci*, 356:867–876. doi: 10.1098/rstb.2001.0863
- [52] K Yusim, M Peeters, OG Pybus, T Bhattacharya, E Delaporte, C Mulanga, M Muldoon, J Theiler and B Korber. 2001. Using human immunodeficiency virus type 1 sequences to infer historical features of the acquired immune deficiency syndrome epidemic and human immunodeficiency virus evolution. *Philos Trans R Soc Lond B Biol Sci*, 356: 855–866. doi: 10.1098/rstb.2001.0859
- [53] NR Faria, A Rambaut, MA Suchard, G Baele, T Bedford, MJ Ward, AJ Tatem, JD Sousa, N Arinaminpathy et al. 2014. The early spread and epidemic ignition of HIV-1 in human populations. *Science*, 346:56–61. doi: 10.1126/science.1256739
- [54] M Peeters, C Honoré, T Huet, L Bedjabaga, S Ossari, P Bussi, RW Cooper and E Delaporte. 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS*, 3:625–630

- [55] M Peeters, V Courgnaud and B Abela. 2001. Genetic diversity of lentiviruses in non-human primates. *AIDS Rev*, 3:3–10
- [56] T Huet, R Cheynier, A Meyerhans, G Roelants and S Wain-Hobson. 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature*, 345:356–359. doi: 10.1038/345356a0
- [57] F Gao, E Bailes, DL Robertson, Y Chen, CM Rodenburg, SF Michael, LB Cummins, LO Arthur, M Peeters et al. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, 397:436–441. doi: 10.1038/17130
- [58] BF Keele, F Van Heuverswyn, Y Li, E Bailes, J Takehisa, ML Santiago, F Bibollet-Ruche, Y Chen, LV Wain et al. 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 313:523–526. doi: 10.1126/science.1126531
- [59] F Van Heuverswyn, Y Li, E Bailes, C Neel, B Lafay, BF Keele, KS Shaw, J Takehisa, MH Kraus et al. 2007. Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. *Virology*, 368:155–171. doi: 10.1016/j.virol.2007.06.018
- [60] E Bowen-Jones and S Pendry. 1999. The threat to primates and other mammals from the bushmeat trade in Africa, and how this threat could be diminished. *Oryx*, 33: 233–246. doi: 10.1046/j.1365-3008.1999.00066.x
- [61] BH Hahn, GM Shaw, KM De Cock and PM Sharp. 2000. AIDS as a zoonosis: scientific and public health implications. *Science*, 287:607–614. doi: 10.1126/science.287.5453.607
- [62] M Peeters, V Courgnaud, B Abela, P Auzel, X Pourrut, F Bibollet-Ruche, S Loul, F Liegeois, C Butel et al. 2002. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis*, 8:451–457. doi: 10.3201/eid0805.010522
- [63] ND Wolfe, TA Prosser, JK Carr, U Tamoufe, E Mpoudi-Ngole, JN Torimiro, M LeBreton, FE McCutchan, DL Birx and DS Burke. 2004. Exposure to nonhuman primates in rural Cameroon. *Emerg Infect Dis*, 10:2094–2099. doi: 10.3201/eid1012.040062
- [64] ND Wolfe, W Heneine, JK Carr, AD Garcia, V Shanmugam, U Tamoufe, JN Torimiro, AT Prosser, M Lebreton et al. 2005. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc Natl Acad Sci U S A*, 102: 7994–7999. doi: 10.1073/pnas.0501734102
- [65] ML Kalish, ND Wolfe, CB Ndongmo, J McNicholl, KE Robbins, M Aidoo, PN Fonjungo, G Alemnji, C Zeh et al. 2005. Central African hunters exposed to simian immunodeficiency virus. *Emerg Infect Dis*, 11:1928–1930. doi: 10.3201/eid1112.050394
- [66] PM Sharp and BH Hahn. 2008. AIDS: prehistory of HIV-1. *Nature*, 455:605–606. doi: 10.1038/455605a

- [67] D Vangroenweghe. 2001. The earliest cases of human immunodeficiency virus type 1 group M in Congo-Kinshasa, Rwanda and Burundi and the origin of acquired immune deficiency syndrome. *Philos Trans R Soc Lond B Biol Sci*, 356:923–925. doi: 10.1098/rstb.2001.0876
- [68] A Chitnis, D Rawls and J Moore. 2000. Origin of HIV type 1 in colonial French Equatorial Africa? *AIDS Res Hum Retroviruses*, 16:5–8. doi: 10.1089/088922200309548
- [69] JD de Sousa, V Müller, P Lemey and AM Vandamme. 2010. High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains. *PLoS One*, 5:e9936. doi: 10.1371/journal.pone.0009936
- [70] JD de Sousa, C Alvarez, AM Vandamme and V Müller. 2012. Enhanced heterosexual transmission hypothesis for the origin of pandemic HIV-1. *Viruses*, 4:1950–1983. doi: 10.3390/v4101950
- [71] MTP Gilbert, A Rambaut, G Wlasiuk, TJ Spira, AE Pitchenik and M Worobey. 2007. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A*, 104:18566–18570. doi: 10.1073/pnas.0705329104
- [72] UNAIDS Communications and Global Advocacy. 2014. Fact sheet 2014. URL <http://www.unaids.org/en/resources/campaigns/2014/2014gapreport/factsheet>
- [73] RD Moore and RE Chaisson. 1996. Natural history of opportunistic disease in an HIV-infected urban clinical cohort. *Ann Intern Med*, 124:633–642. doi: 10.7326/0003-4819-124-7-199604010-00003
- [74] R Rothenberg, M Woelfel, R Stoneburner, J Milberg, R Parker and B Truman. 1987. Survival with the acquired immunodeficiency syndrome. Experience with 5833 cases in New York City. *N Engl J Med*, 317:1297–1302. doi: 10.1056/NEJM198711193172101
- [75] S Vella, M Giuliano, P Pezzotti, MG Agresti, C Tomino, M Floridia, D Greco, M Moroni, G Visco and F Milazzo. 1992. Survival of zidovudine-treated patients with AIDS compared with that of contemporary untreated patients. *JAMA*, 267:1232–1236. doi: 10.1001/jama.1992.03480090080031
- [76] KJ Lui, DN Lawrence, WM Morgan, TA Peterman, HW Haverkos and DJ Bregman. 1986. A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc Natl Acad Sci U S A*, 83: 3051–3055
- [77] MM Deschamps, DW Fitzgerald, JW Pape and W Johnson, Jr. 2000. HIV infection in Haiti: natural history and disease progression. *AIDS*, 14:2515–2521
- [78] KM Harrison, R Song and X Zhang. 2010. Life expectancy after HIV diagnosis based on national HIV surveillance data from 25 states, United States. *J Acquir Immune Defic Syndr*, 53:124–130. doi: 10.1097/QAI.0b013e3181b563e7

- [79] Collaborative Group on AIDS Incubation and HIV Survival. 2000. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. *Lancet*, 355:1131–1137. doi: 10.1016/S0140-6736(00)02061-4
- [80] MA Fischl, DD Richman, MH Grieco, MS Gottlieb, PA Volberding, OL Laskin, JM Leedom, JE Groopman, D Mildvan and RT Schooley. 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *N Engl J Med*, 317:185–191. doi: 10.1056/NEJM198707233170401
- [81] MA Fischl, DD Richman, DM Causey, MH Grieco, Y Bryson, D Mildvan, OL Laskin, JE Groopman, PA Volberding and RT Schooley. 1989. Prolonged zidovudine therapy in patients with AIDS and advanced AIDS-related complex. *JAMA*, 262:2405–2410. doi: 10.1001/jama.1989.03430170067030
- [82] PA Volberding, SW Lagakos, MA Koch, C Pettinelli, MW Myers, DK Booth, H Balfour, Jr, RC Reichman, JA Bartlett et al. 1990. Zidovudine in asymptomatic human immunodeficiency virus infection. A controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. *N Engl J Med*, 322:941–949. doi: 10.1056/NEJM199004053221401
- [83] BH Hahn, GM Shaw, ME Taylor, RR Redfield, PD Markham, SZ Salahuddin, F Wong-Staal, RC Gallo, ES Parks and WP Parks. 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science*, 232:1548–1553. doi: 10.1126/science.3012778
- [84] BD Preston, BJ Poiesz and LA Loeb. 1988. Fidelity of HIV-1 reverse transcriptase. *Science*, 242:1168–1171. doi: 10.1126/science.2460924
- [85] JD Roberts, K Bebenek and TA Kunkel. 1988. The accuracy of reverse transcriptase from HIV-1. *Science*, 242:1171–1173. doi: 10.1126/science.2460925
- [86] LM Mansky and HM Temin. 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol*, 69:5087–5094
- [87] LM Mansky. 1996. The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology*, 222:391–400. doi: 10.1006/viro.1996.0436
- [88] ME Abram, AL Ferris, W Shao, WG Alvord and SH Hughes. 2010. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol*, 84: 9864–9878. doi: 10.1128/JVI.00915-10
- [89] V Achuthan, BJ Keith, BA Connolly and JJ DeStefano. 2014. Human immunodeficiency virus reverse transcriptase displays dramatically higher fidelity under physiological magnesium conditions in vitro. *J Virol*, 88:8514–8527. doi: 10.1128/JVI.00752-14

- [90] BA Larder, G Darby and DD Richman. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science*, 243:1731–1734. doi: 10.1126/science.2467383
- [91] BA Larder and SD Kemp. 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science*, 246:1155–1158. doi: 10.1126/science.2479983
- [92] S Land, G Terloar, D McPhee, C Birch, R Doherty, D Cooper and I Gust. 1990. Decreased in vitro susceptibility to zidovudine of HIV isolates obtained from patients with AIDS. *J Infect Dis*, 161:326–329. doi: 10.1093/infdis/161.2.326
- [93] CA Boucher, M Tersmette, JM Lange, P Kellam, RE de Goede, JW Mulder, G Darby, J Goudsmit and BA Larder. 1990. Zidovudine sensitivity of human immunodeficiency viruses from high-risk, symptom-free individuals during therapy. *Lancet*, 336:585–590. doi: 10.1016/0140-6736(90)93391-2
- [94] DD Richman, JM Grimes and SW Lagakos. 1990. Effect of stage of disease and drug dose on zidovudine susceptibilities of isolates of human immunodeficiency virus. *J Acquir Immune Defic Syndr*, 3:743–746
- [95] DD Richman, JC Guatelli, J Grimes, A Tsiatis and T Gingeras. 1991. Detection of mutations associated with zidovudine resistance in human immunodeficiency virus by use of the polymerase chain reaction. *J Infect Dis*, 164:1075–1081. doi: 10.1093/infdis/164.6.1075
- [96] JE Fitzgibbon, RM Howell, CA Haberzettl, SJ Sperber, DJ Gocke and DT Dubin. 1992. Human immunodeficiency virus type 1 pol gene mutations which cause decreased susceptibility to 2',3'-dideoxycytidine. *Antimicrob Agents Chemother*, 36:153–157. doi: 10.1128/AAC.36.1.153
- [97] DD Richman, D Havlir, J Corbeil, D Looney, C Ignacio, SA Spector, J Sullivan, S Cheeseman, K Barringer and D Pauletti. 1994. Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *J Virol*, 68:1660–1666
- [98] R Schuurman, M Nijhuis, R van Leeuwen, P Schipper, D de Jong, P Collis, SA Danner, J Mulder, C Loveday and C Christopherson. 1995. Rapid changes in human immunodeficiency virus type 1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). *J Infect Dis*, 171:1411–1419. doi: 10.1093/infdis/171.6.1411
- [99] JC Schmit, L Ruiz, B Clotet, A Raventos, J Tor, J Leonard, J Desmyter, E De Clercq and AM Vandamme. 1996. Resistance-related mutations in the HIV-1 protease gene of patients treated for 1 year with the protease inhibitor ritonavir (ABT-538). *AIDS*, 10:995–999
- [100] T Creagh-Kirk, P Doi, E Andrews, S Nusinoff-Lehrman, H Tilson, D Hoth and

- DW Barry. 1988. Survival experience among patients with AIDS receiving zidovudine. Follow-up of patients in a compassionate plea program. *JAMA*, 260:3009–3015. doi: 10.1001/jama.1988.03410200065027
- [101] RD Moore, J Keruly, DD Richman, T Creagh-Kirk and RE Chaisson. 1992. Natural history of advanced HIV disease in patients treated with zidovudine. *AIDS*, 6:671–677
- [102] S Vella, B Schwartländer, SP Sow, SP Eholie and RL Murphy. 2012. The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. *AIDS*, 26:1231–1241. doi: 10.1097/QAD.0b013e32835521a3
- [103] JO Kahn, SW Lagakos, DD Richman, A Cross, C Pettinelli, SH Liou, M Brown, PA Volberding, CS Crumpacker and G Beall. 1992. A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. *N Engl J Med*, 327:581–587. doi: 10.1056/NEJM199208273270901
- [104] G Skowron, SA Bozzette, L Lim, CB Pettinelli, HH Schaumburg, J Arezzo, MA Fischl, WG Powderly, DJ Gocke et al. 1993. Alternating and intermittent regimens of zidovudine and dideoxycytidine in patients with AIDS or AIDS-related complex. *Ann Intern Med*, 118:321–330
- [105] DI Abrams, AI Goldman, C Launer, JA Korvick, JD Neaton, LR Crane, M Grodesky, S Wakefield, K Muth and S Kornegay. 1994. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *N Engl J Med*, 330:657–662. doi: 10.1056/NEJM199403103301001
- [106] MD de Jong, M Loewenthal, CA Boucher, I van der Ende, D Hall, P Schipper, A Imrie, HM Weigel, RH Kauffmann and R Koster. 1994. Alternating nevirapine and zidovudine treatment of human immunodeficiency virus type 1-infected persons does not prolong nevirapine activity. *J Infect Dis*, 169:1346–1350. doi: 10.1093/infdis/169.6.1346
- [107] JC Schmit, J Cogniaux, P Hermans, C Van Vaeck, S Sprecher, B Van Remoortel, M Witvrouw, J Balzarini, J Desmyter et al. 1996. Multiple drug resistance to nucleoside analogues and nonnucleoside reverse transcriptase inhibitors in an efficiently replicating human immunodeficiency virus type 1 patient strain. *J Infect Dis*, 174:962–968. doi: 10.1093/infdis/174.5.962
- [108] AC Collier, RW Coombs, MA Fischl, PR Skolnik, D Northfelt, P Boutin, CJ Hooper, LD Kaplan, PA Volberding et al. 1993. Combination therapy with zidovudine and didanosine compared with zidovudine alone in HIV-1 infection. *Ann Intern Med*, 119: 786–793. doi: 10.7326/0003-4819-119-8-199310150-00003
- [109] SM Hammer, DA Katzenstein, MD Hughes, H Gundacker, RT Schooley, RH Haubrich, WK Henry, MM Lederman, JP Phair et al. 1996. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *N Engl J Med*, 335:1081–1090. doi: 10.1056/NEJM199610103351501

- [110] JJ Eron, SL Benoit, J Jemsek, RD MacArthur, J Santana, JB Quinn, DR Kuritzkes, MA Fallon and M Rubin. 1995. Treatment with lamivudine, zidovudine, or both in HIV-positive patients with 200 to 500 CD4+ cells per cubic millimeter. *N Engl J Med*, 333:1662–1669. doi: 10.1056/NEJM199512213332502
- [111] LD Saravolatz, DL Winslow, G Collins, JS Hodges, C Pettinelli, DS Stein, N Markowitz, R Reves, MO Loveless et al. 1996. Zidovudine alone or in combination with didanosine or zalcitabine in HIV-infected patients with the acquired immunodeficiency syndrome or fewer than 200 CD4 cells per cubic millimeter. *N Engl J Med*, 335:1099–1106. doi: 10.1056/NEJM199610103351503
- [112] J Darbyshire, Delta Coordinating Committee et al. 1996. Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. *Lancet*, 348:283–291. doi: 10.1016/S0140-6736(96)05387-1
- [113] RE Dornstife, MH St Clair, AT Huang, TJ Panella, GW Koszalka, CL Burns and DR Averett. 1991. Anti-human immunodeficiency virus synergism by zidovudine (3'-azidothymidine) and didanosine (dideoxyinosine) contrasts with their additive inhibition of normal human marrow progenitor cells. *Antimicrob Agents Chemother*, 35:322–328. doi: 10.1128/AAC.35.2.322
- [114] VA Johnson, DP Merrill, JA Videler, TC Chou, RE Byington, JJ Eron, RT D'Aquila and MS Hirsch. 1991. Two-drug combinations of zidovudine, didanosine, and recombinant interferon-alpha A inhibit replication of zidovudine-resistant human immunodeficiency virus type 1 synergistically in vitro. *J Infect Dis*, 164:646–655. doi: 10.1093/infdis/164.4.646
- [115] SW Cox, K Apéria, J Albert and B Wahren. 1994. Comparison of the sensitivities of primary isolates of HIV type 2 and HIV type 1 to antiviral drugs and drug combinations. *AIDS Res Hum Retroviruses*, 10:1725–1729. doi: 10.1177/095632029300400407
- [116] JY Feng, JK Ly, F Myrick, D Goodman, KL White, ES Svarovskaia, K Borroto-Esoda and MD Miller. 2009. The triple combination of tenofovir, emtricitabine and efavirenz shows synergistic anti-HIV-1 activity in vitro: a mechanism of action study. *Retrovirology*, 6:44. doi: 10.1186/1742-4690-6-44
- [117] BL Jilek, M Zarr, ME Sampah, SA Rabi, CK Bullen, J Lai, L Shen and RF Siliciano. 2012. A quantitative basis for antiretroviral therapy for HIV-1 infection. *Nat Med*, 18: 446–451. doi: 10.1038/nm.2649
- [118] R Kulkarni, R Hluhanich, DM McColl, MD Miller and KL White. 2014. The combined anti-HIV-1 activities of emtricitabine and tenofovir plus the integrase inhibitor elvitegravir or raltegravir show high levels of synergy in vitro. *Antimicrob Agents Chemother*, 58:6145–6150. doi: 10.1128/AAC.03591-14
- [119] YK Chow, MS Hirsch, DP Merrill, LJ Bechtel, JJ Eron, JC Kaplan and RT D'Aquila.

1993. Use of evolutionary limitations of HIV-1 multidrug resistance to optimize therapy. *Nature*, 361:650–654. doi: 10.1038/361650a0
- [120] BA Larder, SD Kemp and PR Harrigan. 1995. Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science*, 269:696–699. doi: 10.1126/science.7542804
- [121] AC Collier, RW Coombs, DA Schoenfeld, RL Bassett, J Timpone, A Baruch, M Jones, K Facey, C Whitacre et al. 1996. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. *N Engl J Med*, 334:1011–1017. doi: 10.1056/NEJM199604183341602
- [122] SM Hammer, KE Squires, MD Hughes, JM Grimes, LM Demeter, JS Currier, J Eron, Jr, JE Feinberg, H Balfour, Jr et al. 1997. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *N Engl J Med*, 337:725–733. doi: 10.1056/NEJM199709113371101
- [123] RM Gulick, JW Mellors, D Havlir, JJ Eron, C Gonzalez, D McMahon, DD Richman, FT Valentine, L Jonas et al. 1997. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med*, 337:734–739. doi: 10.1056/NEJM199709113371102
- [124] JS Montaner, P Reiss, D Cooper, S Vella, M Harris, B Conway, MA Wainberg, D Smith, P Robinson et al. 1998. A randomized, double-blind trial comparing combinations of nevirapine, didanosine, and zidovudine for HIV-infected patients: the INCAS Trial. Italy, The Netherlands, Canada and Australia Study. *JAMA*, 279:930–937. doi: 10.1001/jama.279.12.930
- [125] RD Moore and RE Chaisson. 1999. Natural history of HIV infection in the era of combination antiretroviral therapy. *AIDS*, 13:1933–1942
- [126] Antiretroviral Therapy Cohort Collaboration, M Zwahlen, R Harris, M May, R Hogg, D Costagliola, F de Wolf, J Gill, G Fätkenheuer et al. 2009. Mortality of HIV-infected patients starting potent antiretroviral therapy: comparison with the general population in nine industrialized countries. *Int J Epidemiol*, 38:1624–1633. doi: 10.1093/ije/dyp306
- [127] Antiretroviral Therapy Cohort Collaboration. 2008. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet*, 372:293–299. doi: 10.1016/S0140-6736(08)61113-7
- [128] O Keiser, P Taffé, M Zwahlen, M Battegay, E Bernasconi, R Weber, M Rickenbach and SHIVCS . 2004. All cause mortality in the Swiss HIV Cohort Study from 1990 to 2001 in comparison with the Swiss population. *AIDS*, 18:1835–1843
- [129] AI van Sighem, LAJ Gras, P Reiss, K Brinkman, F de Wolf and ATHENAnocs .

2010. Life expectancy of recently diagnosed asymptomatic HIV-infected patients approaches that of uninfected individuals. *AIDS*, 24:1527–1535. doi: 10.1097/QAD.0b013e32833a3946
- [130] F Nakagawa, M May and A Phillips. 2013. Life expectancy living with HIV: recent estimates and future implications. *Curr Opin Infect Dis*, 26:17–25. doi: 10.1097/QCO.0b013e32835ba6b1
 - [131] LF Johnson, J Mossong, RE Dorrington, M Schomaker, CJ Hoffmann, O Keiser, MP Fox, R Wood, H Prozesky et al. 2013. Life expectancies of South African adults starting antiretroviral treatment: collaborative analysis of cohort studies. *PLoS Med*, 10:e1001418. doi: 10.1371/journal.pmed.1001418
 - [132] AR Cillo, MD Sobolewski, RJ Bosch, E Fyne, M Piatak, Jr, JM Coffin and JW Mellors. 2014. Quantification of HIV-1 latency reversal in resting CD4+ T cells from patients on suppressive antiretroviral therapy. *Proc Natl Acad Sci U S A*, 111:7078–7083. doi: 10.1073/pnas.1402873111
 - [133] TW Chun, D Engel, MM Berrey, T Shea, L Corey and AS Fauci. 1998. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci U S A*, 95:8869–8873
 - [134] JB Whitney, AL Hill, S Sanisetty, P Penalosa-MacMaster, J Liu, M Shetty, L Parenteau, C Cabral, J Shields et al. 2014. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature*, 512:74–77. doi: 10.1038/nature13594
 - [135] X Contreras, M Schwenecker, CS Chen, JM McCune, SG Deeks, J Martin and BM Peterlin. 2009. Suberoylanilide hydroxamic acid reactivates HIV from latently infected cells. *J Biol Chem*, 284:6782–6789. doi: 10.1074/jbc.M807898200
 - [136] S Xing, CK Bullen, NS Shroff, L Shan, HC Yang, JL Manucci, S Bhat, H Zhang, JB Margolick et al. 2011. Disulfiram reactivates latent HIV-1 in a Bcl-2-transduced primary CD4+ T cell model without inducing global T cell activation. *J Virol*, 85: 6060–6064. doi: 10.1128/JVI.02033-10
 - [137] TW Chun, L Carruth, D Finzi, X Shen, JA DiGiuseppe, H Taylor, M Hermankova, K Chadwick, J Margolick et al. 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*, 387:183–188. doi: 10.1038/387183a0
 - [138] D Baltimore. 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226:1209–1211. doi: 10.1038/2261209a0
 - [139] HM Temin and S Mizutani. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 226:1211–1213. doi: 10.1038/2261211a0
 - [140] DP Grandgenett, AC Vora and RD Schiff. 1978. A 32,000-dalton nucleic acid-binding

- protein from avian retravirus cores possesses DNA endonuclease activity. *Virology*, 89: 119–132. doi: 10.1016/0042-6822(78)90046-6
- [141] AT Panganiban and HM Temin. 1984. The retrovirus pol gene encodes a product required for DNA integration: identification of a retrovirus int locus. *Proc Natl Acad Sci U S A*, 81:7885–7889
 - [142] P Schwartzberg, J Colicelli and SP Goff. 1984. Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell*, 37:1043–1052. doi: 10.1016/0092-8674(84)90439-2
 - [143] LA Donehower and HE Varmus. 1984. A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. *Proc Natl Acad Sci U S A*, 81:6461–6465
 - [144] AT Panganiban and HM Temin. 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature*, 306:155–160. doi: 10.1038/306155a0
 - [145] NK Heinzinger, MI Bukrinsky, SA Haggerty, AM Ragland, V Kewalramani, MA Lee, HE Gendelman, L Ratner, M Stevenson and M Emerman. 1994. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A*, 91:7311–7315
 - [146] BK Ganser, S Li, VY Klishko, JT Finch and WI Sundquist. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science*, 283:80–83. doi: 10.1126/science.283.5398.80
 - [147] S Li, CP Hill, WI Sundquist and JT Finch. 2000. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*, 407:409–413. doi: 10.1038/35030177
 - [148] IJL Byeon, X Meng, J Jung, G Zhao, R Yang, J Ahn, J Shi, J Concel, C Aiken et al. 2009. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell*, 139:780–790. doi: 10.1016/j.cell.2009.10.010
 - [149] G Zhao, JR Perilla, EL Yufenyuy, X Meng, B Chen, J Ning, J Ahn, AM Gronenborn, K Schulten et al. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646. doi: 10.1038/nature12162
 - [150] K Lee, Z Ambrose, TD Martin, I Oztop, A Mulky, JG Julias, N Vandegraaff, JG Baumann, R Wang et al. 2010. Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*, 7:221–233. doi: 10.1016/j.chom.2010.02.007
 - [151] M Thali, A Bukovsky, E Kondo, B Rosenwirth, CT Walsh, J Sodroski and

- HG Göttlinger. 1994. Functional association of cyclophilin A with HIV-1 virions. *Nature*, 372:363–365. doi: 10.1038/372363a0
- [152] TR Gamble, FF Vajdos, S Yoo, DK Worthy lake, M Houseweart, WI Sundquist and CP Hill. 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*, 87:1285–1294. doi: 10.1016/S0092-8674(00)81823-1
- [153] T Schaller, KE Ocwieja, J Rasaiyaah, AJ Price, TL Brady, SL Roth, S Hué, AJ Fletcher, K Lee et al. 2011. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog*, 7:e1002439. doi: 10.1371/journal.ppat.1002439
- [154] KE Ocwieja, TL Brady, K Ronen, A Huegel, SL Roth, T Schaller, LC James, GJ Towers, JAT Young et al. 2011. HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog*, 7:e1001313. doi: 10.1371/journal.ppat.1001313
- [155] J Rasaiyaah, CP Tan, AJ Fletcher, AJ Price, C Blondeau, L Hilditch, DA Jacques, DL Selwood, LC James et al. 2013. HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature*, 503:402–405. doi: 10.1038/nature12769
- [156] FD Veronese, R Rahman, TD Copeland, S Oroszlan, RC Gallo and MG Sarngadharan. 1987. Immunological and chemical analysis of P6, the carboxyl-terminal fragment of HIV P15. *AIDS Res Hum Retroviruses*, 3:253–264. doi: 10.1089/aid.1987.3.253
- [157] HG Göttlinger, T Dorfman, JG Sodroski and WA Haseltine. 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A*, 88:3195–3199. doi: 10.1073/pnas.88.8.3195
- [158] B Strack, A Calistri, S Craig, E Popova and HG Göttlinger. 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell*, 114: 689–699. doi: 10.1016/S0092-8674(03)00653-6
- [159] W Paxton, RI Connor and NR Landau. 1993. Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *J Virol*, 67:7229–7237
- [160] FD Bushman, T Fujiwara and R Craigie. 1990. Retroviral DNA integration directed by HIV integration protein in vitro. *Science*, 249:1555–1558. doi: 10.1126/science.2171144
- [161] A Engelman, K Mizuuchi and R Craigie. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, 67:1211–1221. doi: 10.1016/0092-8674(91)90297-C
- [162] GN Maertens, S Hare and P Cherepanov. 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature*, 468:326–329. doi: 10.1038/nature09517

- [163] S Hare, SS Gupta, E Valkov, A Engelman and P Cherepanov. 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, 464:232–236. doi: 10.1038/nature08784
- [164] FD Bushman and R Craigie. 1991. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci U S A*, 88:1339–1343. doi: 10.1073/pnas.88.4.1339
- [165] A Wlodawer, M Miller, M Jaskólski, BK Sathyaranayana, E Baldwin, IT Weber, LM Selk, L Clawson, J Schneider and SB Kent. 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, 245:616–621. doi: 10.1126/science.2548279
- [166] HG Kräusslich, RH Ingraham, MT Skoog, E Wimmer, PV Pallai and CA Carter. 1989. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci U S A*, 86:807–811
- [167] NE Kohl, EA Emini, WA Schleif, LJ Davis, JC Heimbach, RA Dixon, EM Scolnick and IS Sigal. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A*, 85:4686–4690
- [168] AG Dalgleish, PC Beverley, PR Clapham, DH Crawford, MF Greaves and RA Weiss. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 312:763–767. doi: 10.1038/312763a0
- [169] D Klatzmann, E Champagne, S Chamaret, J Gruest, D Guetard, T Hercend, JC Gluckman and L Montagnier. 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*, 312:767–768. doi: 10.1038/312767a0
- [170] JD Lifson, MB Feinberg, GR Reyes, L Rabin, B Banapour, S Chakrabarti, B Moss, F Wong-Staal, KS Steimer and EG Engleman. 1986. Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. *Nature*, 323:725–728. doi: 10.1038/323725a0
- [171] JD Lifson, GR Reyes, MS McGrath, BS Stein and EG Engleman. 1986. AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. *Science*, 232:1123–1127. doi: 10.1126/science.3010463
- [172] PJ Maddon, AG Dalgleish, JS McDougal, PR Clapham, RA Weiss and R Axel. 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 47:333–348. doi: 10.1016/0092-8674(86)90590-8
- [173] Y Feng, CC Broder, PE Kennedy and EA Berger. 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, 272:872–877. doi: 10.1126/science.272.5263.872
- [174] H Choe, M Farzan, Y Sun, N Sullivan, B Rollins, PD Ponath, L Wu, CR Mackay,

- G LaRosa et al. 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 85:1135–1148. doi: 10.1016/S0092-8674(00)81313-6
- [175] J He, Y Chen, M Farzan, H Choe, A Ohagen, S Gartner, J Busciglio, X Yang, W Hofmann et al. 1997. CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature*, 385:645–649. doi: 10.1038/385645a0
- [176] FD Veronese, AL DeVico, TD Copeland, S Oroszlan, RC Gallo and MG Sarngadharan. 1985. Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science*, 229:1402–1405. doi: 10.1126/science.2994223
- [177] S Hallenberger, V Bosch, H Angliker, E Shaw, HD Klenk and W Garten. 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature*, 360:358–361. doi: 10.1038/360358a0
- [178] X Wei, JM Decker, S Wang, H Hui, JC Kappes, X Wu, JF Salazar-Gonzalez, MG Salazar, JM Kilby et al. 2003. Antibody neutralization and escape by HIV-1. *Nature*, 422:307–312. doi: 10.1038/nature01470
- [179] P Zhu, J Liu, J Bess, Jr, E Chertova, JD Lifson, H Grisé, GA Ofek, KA Taylor and KH Roux. 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature*, 441:847–852. doi: 10.1038/nature04817
- [180] EC Holmes, LQ Zhang, P Simmonds, CA Ludlam and AJ Brown. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A*, 89:4835–4839
- [181] R Shankarappa, JB Margolick, SJ Gange, AG Rodrigo, D Upchurch, H Farzadegan, P Gupta, CR Rinaldo, GH Learn et al. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol*, 73:10489–10502
- [182] S Bonhoeffer, EC Holmes and MA Nowak. 1995. Causes of HIV diversity. *Nature*, 376:125. doi: 10.1038/376125a0
- [183] SM Wolinsky, BT Korber, AU Neumann, M Daniels, KJ Kunstman, AJ Whetsell, MR Furtado, Y Cao, DD Ho and JT Safrit. 1996. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science*, 272: 537–542. doi: 10.1126/science.272.5261.537
- [184] HA Ross and AG Rodrigo. 2002. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J Virol*, 76:11715–11720. doi: 10.1128/JVI.76.22.11715-11720.2002
- [185] J Sodroski, C Rosen, F Wong-Staal, SZ Salahuddin, M Popovic, S Arya, RC Gallo and

- WA Haseltine. 1985. Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. *Science*, 227:171–173. doi: 10.1126/science.2981427
- [186] J Sodroski, R Patarca, C Rosen, F Wong-Staal and W Haseltine. 1985. Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. *Science*, 229:74–77. doi: 10.1126/science.2990041
- [187] BR Cullen. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell*, 46:973–982. doi: 10.1016/0092-8674(86)90696-3
- [188] AI Dayton, JG Sodroski, CA Rosen, WC Goh and WA Haseltine. 1986. The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell*, 44:941–947. doi: 10.1016/0092-8674(86)90017-6
- [189] TK Howcroft, K Strebel, MA Martin and DS Singer. 1993. Repression of MHC class I gene promoter activity by two-exon Tat of HIV. *Science*, 260:1320–1322. doi: 10.1126/science.8493575
- [190] Y Bennasser, SY Le, M Benkirane and KT Jeang. 2005. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity*, 22:607–619. doi: 10.1016/j.immuni.2005.03.010
- [191] R Triboulet, B Mari, YL Lin, C Chable-Bessia, Y Bennasser, K Lebrigand, B Cardinaud, T Maurin, P Barbry et al. 2007. Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science*, 315:1579–1582. doi: 10.1126/science.1136319
- [192] S Qian, X Zhong, L Yu, B Ding, P de Haan and K Boris-Lawrie. 2009. HIV-1 Tat RNA silencing suppressor activity is conserved across kingdoms and counteracts translational repression of HIV-1. *Proc Natl Acad Sci U S A*, 106:605–610. doi: 10.1073/pnas.0806822106
- [193] J Lin and BR Cullen. 2007. Analysis of the interaction of primate retroviruses with the human RNA interference machinery. *J Virol*, 81:12218–12226. doi: 10.1128/JVI.01390-07
- [194] BE Meyer and MH Malim. 1994. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev*, 8:1538–1547. doi: 10.1101/gad.8.13.1538
- [195] J Sodroski, WC Goh, C Rosen, A Dayton, E Terwilliger and W Haseltine. 1986. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*, 321:412–417. doi: 10.1038/321412a0
- [196] MB Feinberg, RF Jarrett, A Aldovini, RC Gallo and F Wong-Staal. 1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell*, 46:807–817. doi: 10.1016/0092-8674(86)90062-0
- [197] DM Knight, FA Flomerfelt and J Ghrayeb. 1987. Expression of the art/trs protein

- of HIV and study of its role in viral envelope synthesis. *Science*, 236:837–840. doi: 10.1126/science.3033827
- [198] MH Malim, J Hauber, R Fenrick and BR Cullen. 1988. Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. *Nature*, 335: 181–183. doi: 10.1038/335181a0
- [199] D Gutman and CJ Goldenberg. 1988. Virus-specific splicing inhibitor in extracts from cells infected with HIV-1. *Science*, 241:1492–1495. doi: 10.1126/science.3047873
- [200] MH Malim, J Hauber, SY Le, JV Maizel and BR Cullen. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, 338:254–257. doi: 10.1038/338254a0
- [201] MH Malim, S Bhnlein, J Hauber and BR Cullen. 1989. Functional dissection of the HIV-1 Rev trans-activator—derivation of a trans-dominant repressor of Rev function. *Cell*, 58:205–214. doi: 10.1016/0092-8674(89)90416-9
- [202] G Yu and RL Felsted. 1992. Effect of myristoylation on p27 nef subcellular distribution and suppression of HIV-LTR transcription. *Virology*, 187:46–55. doi: 10.1016/0042-6822(92)90293-X
- [203] JV Garcia and AD Miller. 1991. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature*, 350:508–511. doi: 10.1038/350508a0
- [204] RE Benson, A Sanfridson, JS Ottinger, C Doyle and BR Cullen. 1993. Downregulation of cell-surface CD4 expression by simian immunodeficiency virus Nef prevents viral super infection. *J Exp Med*, 177:1561–1566. doi: 10.1084/jem.177.6.1561
- [205] C Aiken, J Konner, NR Landau, ME Lenburg and D Trono. 1994. Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell*, 76:853–864. doi: 10.1016/0092-8674(94)90360-3
- [206] J Lama, A Mangasarian and D Trono. 1999. Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner. *Curr Biol*, 9:622–631. doi: 10.1016/S0960-9822(99)80284-X
- [207] TM Ross, AE Oran and BR Cullen. 1999. Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein. *Curr Biol*, 9: 613–621. doi: 10.1016/S0960-9822(99)80283-8
- [208] N Michel, I Allespach, S Venzke, OT Fackler and OT Keppler. 2005. The Nef protein of human immunodeficiency virus establishes superinfection immunity by a dual strategy to downregulate cell-surface CCR5 and CD4. *Curr Biol*, 15:714–723. doi: 10.1016/j.cub.2005.02.058
- [209] O Schwartz, V Maréchal, S Le Gall, F Lemonnier and JM Heard. 1996. Endocytosis

- of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat Med*, 2:338–342. doi: 10.1038/nm0396-338
- [210] KL Collins, BK Chen, SA Kalams, BD Walker and D Baltimore. 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature*, 391:397–401. doi: 10.1038/34929
- [211] P Stumptner-Cuvelette, S Morchoisne, M Dugast, S Le Gall, G Raposo, O Schwartz and P Benaroch. 2001. HIV-1 Nef impairs MHC class II antigen presentation and surface expression. *Proc Natl Acad Sci U S A*, 98:12144–12149. doi: 10.1073/pnas.221256498
- [212] AD Blagoveshchenskaya, L Thomas, SF Feliciangeli, CH Hung and G Thomas. 2002. HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway. *Cell*, 111:853–866. doi: 10.1016/S0092-8674(02)01162-5
- [213] XN Xu, B Laffert, GR Screaton, M Kraft, D Wolf, W Kolanus, J Mongkolsapay, AJ McMichael and AS Baur. 1999. Induction of Fas ligand expression by HIV involves the interaction of Nef with the T cell receptor zeta chain. *J Exp Med*, 189:1489–1496. doi: 10.1084/jem.189.9.1489
- [214] JA Schrager and JW Marsh. 1999. HIV-1 Nef increases T cell activation in a stimulus-dependent manner. *Proc Natl Acad Sci U S A*, 96:8167–8172. doi: 10.1073/pnas.96.14.8167
- [215] JK Wang, E Kiyokawa, E Verdin and D Trono. 2000. The Nef protein of HIV-1 associates with rafts and primes T cells for activation. *Proc Natl Acad Sci U S A*, 97: 394–399. doi: 10.1073/pnas.97.1.394
- [216] A Simmons, V Aluvihare and A McMichael. 2001. Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. *Immunity*, 14:763–777. doi: 10.1016/S1074-7613(01)00158-3
- [217] JA Schrager, V Der Minassian and JW Marsh. 2002. HIV Nef increases T cell ERK MAP kinase activity. *J Biol Chem*, 277:6137–6142. doi: 10.1074/jbc.M107322200
- [218] M Schindler, J Münch, O Kutsch, H Li, ML Santiago, F Bibollet-Ruche, MC Müller-Trutwin, FJ Novembre, M Peeters et al. 2006. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell*, 125:1055–1067. doi: 10.1016/j.cell.2006.04.033
- [219] F Kirchhoff, M Schindler, A Specht, N Arhel and J Münch. 2008. Role of Nef in primate lentiviral immunopathogenesis. *Cell Mol Life Sci*, 65:2621–2636. doi: 10.1007/s00018-008-8094-2
- [220] F Kirchhoff. 2009. Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? *Nat Rev Microbiol*, 7:467–476. doi: 10.1038/nrmicro2111

- [221] F Wong-Staal, PK Chanda and J Ghrayeb. 1987. Human immunodeficiency virus: the eighth gene. *AIDS Res Hum Retroviruses*, 3:33–39. doi: 10.1089/aid.1987.3.33
- [222] EA Cohen, EF Terwilliger, Y Jalinoos, J Proulx, JG Sodroski and WA Haseltine. 1990. Identification of HIV-1 vpr product and function. *J Acquir Immune Defic Syndr*, 3: 11–18
- [223] EA Cohen, G Dehni, JG Sodroski and WA Haseltine. 1990. Human immunodeficiency virus vpr product is a virion-associated regulatory protein. *J Virol*, 64:3097–3099
- [224] X Yuan, Z Matsuda, M Matsuda, M Essex and TH Lee. 1990. Human immunodeficiency virus vpr gene encodes a virion-associated protein. *AIDS Res Hum Retroviruses*, 6: 1265–1271. doi: 10.1089/aid.1990.6.1265
- [225] JB Jowett, V Planelles, B Poon, NP Shah, ML Chen and IS Chen. 1995. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. *J Virol*, 69:6304–6313
- [226] F Re, D Braaten, EK Franke and J Luban. 1995. Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34cdc2-cyclin B. *J Virol*, 69:6859–6864
- [227] J He, S Choe, R Walker, P Di Marzio, DO Morgan and NR Landau. 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. *J Virol*, 69:6705–6711
- [228] ME Rogel, LI Wu and M Emerman. 1995. The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. *J Virol*, 69:882–888
- [229] CM de Noronha, MP Sherman, HW Lin, MV Cavrois, RD Moir, RD Goldman and WC Greene. 2001. Dynamic disruptions in nuclear envelope architecture and integrity induced by HIV-1 Vpr. *Science*, 294:1105–1108. doi: 10.1126/science.1063957
- [230] WC Goh, ME Rogel, CM Kinsey, SF Michael, PN Fultz, MA Nowak, BH Hahn and M Emerman. 1998. HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. *Nat Med*, 4:65–71. doi: 10.1038/nm0198-065
- [231] RA Subbramanian, A Kessous-Elbaz, R Lodge, J Forget, XJ Yao, D Bergeron and EA Cohen. 1998. Human immunodeficiency virus type 1 Vpr is a positive regulator of viral transcription and infectivity in primary human macrophages. *J Exp Med*, 187: 1103–1111. doi: 10.1084/jem.187.7.1103
- [232] SA Stewart, B Poon, JB Jowett and IS Chen. 1997. Human immunodeficiency virus type 1 Vpr induces apoptosis following cell cycle arrest. *J Virol*, 71:5579–5592
- [233] LD Shostak, J Ludlow, J Fisk, S Pursell, BJ Rimel, D Nguyen, JD Rosenblatt and

- V Planelles. 1999. Roles of p53 and caspases in the induction of cell cycle arrest and apoptosis by HIV-1 vpr. *Exp Cell Res*, 251:156–165. doi: 10.1006/excr.1999.4568
- [234] AM Sheehy, NC Gaddis, JD Choi and MH Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 418: 646–650. doi: 10.1038/nature00939
 - [235] R Mariani, D Chen, B Schröfelbauer, F Navarro, R König, B Bollman, C Münk, H Nymark-McMahon and NR Landau. 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, 114:21–31. doi: 10.1016/S0092-8674(03)00515-4
 - [236] AM Sheehy, NC Gaddis and MH Malim. 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med*, 9:1404–1407. doi: 10.1038/nm945
 - [237] M Marin, KM Rose, SL Kozak and D Kabat. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med*, 9:1398–1403. doi: 10.1038/nm946
 - [238] X Yu, Y Yu, B Liu, K Luo, W Kong, P Mao and XF Yu. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science*, 302:1056–1060. doi: 10.1126/science.1089591
 - [239] RS Harris, KN Bishop, AM Sheehy, HM Craig, SK Petersen-Mahrt, IN Watt, MS Neuberger and MH Malim. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell*, 113:803–809. doi: 10.1016/S0092-8674(03)00423-9
 - [240] B Mangeat, P Turelli, G Caron, M Friedli, L Perrin and D Trono. 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, 424:99–103. doi: 10.1038/nature01709
 - [241] H Zhang, B Yang, RJ Pomerantz, C Zhang, SC Arunachalam and L Gao. 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*, 424:94–98. doi: 10.1038/nature01707
 - [242] D Lecossier, F Bouchonnet, F Clavel and AJ Hance. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*, 300:1112. doi: 10.1126/science.1083338
 - [243] EA Cohen, EF Terwilliger, JG Sodroski and WA Haseltine. 1988. Identification of a protein encoded by the vpu gene of HIV-1. *Nature*, 334:532–534. doi: 10.1038/334532a0
 - [244] K Strelbel, T Klimkait and MA Martin. 1988. A novel gene of HIV-1, vpu, and its 16-kilodalton product. *Science*, 241:1221–1223. doi: 10.1126/science.3261888
 - [245] RL Willey, F Maldarelli, MA Martin and K Strelbel. 1992. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J Virol*, 66:7193–7200

- [246] S Bour, U Schubert and K Strelbel. 1995. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. *J Virol*, 69:1510–1520
- [247] WL Marshall, DC Diamond, MM Kowalski and RW Finberg. 1992. High level of surface CD4 prevents stable human immunodeficiency virus infection of T-cell transfectants. *J Virol*, 66:5492–5499
- [248] MJ Cortés, F Wong-Staal and J Lama. 2002. Cell surface CD4 interferes with the infectivity of HIV-1 particles released from T cells. *J Biol Chem*, 277:1770–1779. doi: 10.1074/jbc.M109807200
- [249] B Crise, L Buonocore and JK Rose. 1990. CD4 is retained in the endoplasmic reticulum by the human immunodeficiency virus type 1 glycoprotein precursor. *J Virol*, 64: 5585–5593
- [250] S Bour, F Boulerice and MA Wainberg. 1991. Inhibition of gp160 and CD4 maturation in U937 cells after both defective and productive infections by human immunodeficiency virus type 1. *J Virol*, 65:6387–6396
- [251] SJD Neil, T Zang and PD Bieniasz. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*, 451:425–430. doi: 10.1038/nature06553
- [252] K Strelbel, T Klimkait, F Maldarelli and MA Martin. 1989. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J Virol*, 63:3784–3791
- [253] S Koenig, HE Gendelman, JM Orenstein, MC Dal Canto, GH Pezeshkpour, M Yungbluth, F Janotta, A Aksamit, MA Martin and AS Fauci. 1986. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science*, 233:1089–1093. doi: 10.1126/science.3016903
- [254] DD Ho, AU Neumann, AS Perelson, W Chen, JM Leonard and M Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 373:123–126. doi: 10.1038/373123a0
- [255] X Wei, SK Ghosh, ME Taylor, VA Johnson, EA Emini, P Deutsch, JD Lifson, S Bonhoeffer, MA Nowak and BH Hahn. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373:117–122. doi: 10.1038/373117a0
- [256] AS Perelson, P Essunger, Y Cao, M Vesalanen, A Hurley, K Saksela, M Markowitz and DD Ho. 1997. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, 387:188–191. doi: 10.1038/387188a0
- [257] O Lambotte, Y Taoufik, MG de Goér, C Wallon, C Goujard and JF Delfraissy. 2000. Detection of infectious HIV in circulating monocytes from patients on prolonged highly active antiretroviral therapy. *J Acquir Immune Defic Syndr*, 23:114–119

- [258] T Zhu, D Muthui, S Holte, D Nickle, F Feng, S Brodie, Y Hwangbo, JI Mullins and L Corey. 2002. Evidence for human immunodeficiency virus type 1 replication in vivo in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *J Virol*, 76:707–716. doi: 10.1128/JVI.76.2.707-716.2002
- [259] D Finzi, M Hermankova, T Pierson, LM Carruth, C Buck, RE Chaisson, TC Quinn, K Chadwick, J Margolick et al. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*, 278:1295–1300. doi: 10.1126/science.278.5341.1295
- [260] JK Wong, M Hezareh, HF Günthard, DV Havlir, CC Ignacio, CA Spina and DD Richman. 1997. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science*, 278:1291–1295. doi: 10.1126/science.278.5341.1291
- [261] AR Bellamy, SC Gillies and JD Harvey. 1974. Molecular weight of two oncornavirus genomes: derivation from particle molecular weights and RNA content. *J Virol*, 14: 1388–1393
- [262] HJ Kung, JM Bailey, N Davidson, MO Nicolson and RM McAllister. 1975. Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol*, 16:397–411
- [263] HJ Kung, S Hu, W Bender, JM Bailey, N Davidson, MO Nicolson and RM McAllister. 1976. RD-114, baboon, and woolly monkey viral RNA's compared in size and structure. *Cell*, 7:609–620. doi: 10.1016/0092-8674(76)90211-7
- [264] AT Panganiban and D Fiore. 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science*, 241:1064–1069. doi: 10.1126/science.2457948
- [265] WS Hu and HM Temin. 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A*, 87:1556–1560. doi: 10.1073/pnas.87.4.1556
- [266] WS Hu and HM Temin. 1990. Retroviral recombination and reverse transcription. *Science*, 250:1227–1233. doi: 10.1126/science.1700865
- [267] JA Hoxie, BS Haggarty, JL Rackowski, N Pillsbury and JA Levy. 1985. Persistent non-cytopathic infection of normal human T lymphocytes with AIDS-associated retrovirus. *Science*, 229:1400–1402. doi: 10.1126/science.2994222
- [268] T Folks, DM Powell, MM Lightfoote, S Benn, MA Martin and AS Fauci. 1986. Induction of HTLV-III/LAV from a nonvirus-producing T-cell line: implications for latency. *Science*, 231:600–602. doi: 10.1126/science.3003906
- [269] ARW Schröder, P Shinn, H Chen, C Berry, JR Ecker and F Bushman. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, 110: 521–529. doi: 10.1016/S0092-8674(02)00864-4

- [270] C Wang, Y Mitsuya, B Gharizadeh, M Ronaghi and SR W. 2007. Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. *Genome Research*, 17:1195–1201. doi: 10.1101/gr.6468307
- [271] SM Berget, C Moore and PA Sharp. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA*, 74:3171–3175
- [272] LT Chow, RE Gelinas, TR Broker and RJ Roberts. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12:1–8. doi: 10.1016/0092-8674(77)90180-5
- [273] N Lohse, ABE Hansen, G Pedersen, G Kronborg, J Gerstoft, HT Sørensen, M Vaeth and N Obel. 2007. Survival of persons with and without HIV infection in Denmark, 1995–2005. *Ann Intern Med*, 146:87–95
- [274] DD Richman, DM Margolis, M Delaney, WC Greene, D Hazuda and RJ Pomerantz. 2009. The challenge of finding a cure for HIV infection. *Science*, 323:1304–1307. doi: 10.1126/science.1165706
- [275] AB Hutchinson, PG Farnham, HD Dean, DU Ekwueme, C del Rio, L Kamimoto and SE Kellerman. 2006. The economic burden of HIV in the United States in the era of highly active antiretroviral therapy: evidence of continuing racial and ethnic differences. *J Acquir Immune Defic Syndr*, 43:451–457. doi: 10.1097/01.qai.0000243090.32866.4e
- [276] BR Schackman, KA Gebo, RP Walensky, E Losina, T Muccio, PE Sax, MC Weinstein, GR Seage, RD Moore and KA Freedberg. 2006. The lifetime cost of current human immunodeficiency virus care in the United States. *Med Care*, 44:990–997. doi: 10.1097/01.mlr.0000228021.89490.2a
- [277] T Fukuhara, T Hosoya, S Shimizu, K Sumi, T Oshiro, Y Yoshinaka, M Suzuki, N Yamamoto, LA Herzenberg et al. 2006. Utilization of host SR protein kinases and RNA-splicing machinery during viral replication. *Proc Natl Acad Sci USA*, 103: 11329–11333. doi: 10.1073/pnas.0604616103
- [278] D Mandal, Z Feng and CM Stoltzfus. 2010. Excessive RNA splicing and inhibition of HIV-1 replication induced by modified U1 small nuclear RNAs. *J Virol*, 84:12790–12800. doi: 10.1128/JVI.01257-10
- [279] YH Zheng, HF Yu and BM Peterlin. 2003. Human p32 protein relieves a post-transcriptional block to HIV replication in murine cells. *Nat Cell Biol*, 5:611–618. doi: 10.1038/ncb1000
- [280] S Kammler, M Otte, I Hauber, J Kjems, J Hauber and H Schaal. 2006. The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*, 3:89. doi: 10.1186/1742-4690-3-89
- [281] D Dowling, S Nasr-Esfahani, CH Tan, K O'Brien, JL Howard, DA Jans, DF j Purcell,

- CM Stoltzfus and S Sonza. 2008. HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. *Retrovirology*, 5:18. doi: 10.1186/1742-4690-5-18
- [282] C Gélinas and HM Temin. 1986. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosis viruses. *Proc Natl Acad Sci USA*, 83:9211–9215
- [283] SA Herman and JM Coffin. 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science*, 236:845–848. doi: 10.1126/science.3033828
- [284] NH Shin, D Hartigan-O'Connor, JK Pfeiffer and A Telesnitsky. 2000. Replication of lengthened Moloney murine leukemia virus genomes is impaired at multiple stages. *J Virol*, 74:2694–2702
- [285] CM Stoltzfus. 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv Virus Res*, 74:1–40. doi: 10.1016/S0065-3527(09)74001-1
- [286] SY Kim, R Byrn, J Groopman and D Baltimore. 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol*, 63:3708–3713
- [287] RJ Pomerantz, D Trono, MB Feinberg and D Baltimore. 1990. Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. *Cell*, 61:1271–1276. doi: 10.1016/0092-8674(90)90691-7
- [288] AL Brass, DM Dykxhoorn, Y Benita, N Yan, A Engelman, RJ Xavier, J Lieberman and SJ Elledge. 2008. Identification of host proteins required for HIV infection through a functional genomic screen. *Science*, 319:921–926. doi: 10.1126/science.1152725
- [289] R König, Y Zhou, D Elleeder, TL Diamond, GMC Bonamy, JT Irelan, CY Chiang, BP Tu, PDD Jesus et al. 2008. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, 135:49–60. doi: 10.1016/j.cell.2008.07.032
- [290] FD Bushman, N Malani, J Fernandes, I D'Orso, G Cagney, TL Diamond, H Zhou, DJ Hazuda, AS Espeseth et al. 2009. Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog*, 5:e1000437. doi: 10.1371/journal.ppat.1000437
- [291] S Jäger, P Cimermancic, N Gulbahce, JR Johnson, KE McGovern, SC Clarke, M Shales, G Mercenne, L Pache et al. 2012. Global landscape of HIV-human protein complexes. *Nature*, 481:365–370. doi: 10.1038/nature10719
- [292] A Bansal, J Carlson, J Yan, OT Akinsiku, M Schaefer, S Sabbaj, A Bet, DN Levy, S Heath et al. 2010. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J Exp Med*, 207:51–59. doi: 10.1084/jem.20092060

- [293] N Bakkour, YL Lin, S Maire, L Ayadi, F Mahuteau-Betzer, CH Nguyen, C Mettling, P Portales, D Grierson et al. 2007. Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. *PLoS Pathog*, 3: 1530–1539. doi: 10.1371/journal.ppat.0030159
- [294] MB Asparuhova, G Marti, S Liu, F Serhan, D Trono and D Schmperli. 2007. Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. *J Gene Med*, 9:323–334. doi: 10.1002/jgm.1027
- [295] TO Tange, TH Jensen and J Kjems. 1996. In vitro interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32. *J Biol Chem*, 271:10066–10072. doi: 10.1074/jbc.271.17.10066
- [296] R Berro, K Kehn, C de la Fuente, A Pumfery, R Adair, J Wade, AM Colberg-Poley, J Hiscott and F Kashanchi. 2006. Acetylated Tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. *J Virol*, 80:3189–3204. doi: 10.1128/JVI.80.7.3189-3204.2006
- [297] J Bohne, A Schambach and D Zychlinski. 2007. New way of regulating alternative splicing in retroviruses: the promoter makes a difference. *J Virol*, 81:3652–3656. doi: 10.1128/JVI.02105-06
- [298] JA Jablonski, AL Amelio, M Giacca and M Caputi. 2010. The transcriptional transactivator Tat selectively regulates viral splicing. *Nucleic Acids Res*, 38:1249–1260. doi: 10.1093/nar/gkp1105
- [299] M Kuramitsu, C Hashizume, N Yamamoto, A Azuma, M Kamata, N Yamamoto, Y Tanaka and Y Aida. 2005. A novel role for Vpr of human immunodeficiency virus type 1 as a regulator of the splicing of cellular pre-mRNA. *Microbes Infect*, 7:1150–1160. doi: 10.1016/j.micinf.2005.03.022
- [300] C Hashizume, M Kuramitsu, X Zhang, T Kurosawa, M Kamata and Y Aida. 2007. Human immunodeficiency virus type 1 Vpr interacts with spliceosomal protein SAP145 to mediate cellular pre-mRNA splicing inhibition. *Microbes Infect*, 9:490–497. doi: 10.1016/j.micinf.2007.01.013
- [301] NM Kopelman, D Lancet and I Yanai. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet*, 37:588–589. doi: 10.1038/ng1575
- [302] Y Xing and C Lee. 2005. Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proc Natl Acad Sci USA*, 102:13526–13531. doi: 10.1073/pnas.0501213102
- [303] Z Su, J Wang, J Yu, X Huang and X Gu. 2006. Evolution of alternative splicing after gene duplication. *Genome Res*, 16:182–189. doi: 10.1101/gr.4197006

- [304] FL Watson, R Pttmann-Holgado, F Thomas, DL Lamar, M Hughes, M Kondo, VI Rebel and D Schmucker. 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science*, 309:1874–1878. doi: 10.1126/science.1116887
- [305] VW Pollard and MH Malim. 1998. The HIV-1 Rev protein. *Annu Rev Microbiol*, 52: 491–532. doi: 10.1146/annurev.micro.52.1.491
- [306] DF Purcell and MA Martin. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol*, 67:6365–6378
- [307] ME Klotman, S Kim, A Buchbinder, A DeRossi, D Baltimore and F Wong-Staal. 1991. Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes. *Proc Natl Acad Sci USA*, 88: 5011–5015
- [308] DM Benko, S Schwartz, GN Pavlakis and BK Felber. 1990. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. *J Virol*, 64:2505–2518
- [309] K Fujita, J Silver and K Peden. 1992. Changes in both gp120 and gp41 can account for increased growth potential and expanded host range of human immunodeficiency virus type 1. *J Virol*, 66:4445–4451
- [310] RM McAllister, MB Gardner, AE Greene, C Bradt, WW Nichols and BH Landing. 1971. Cultivation in vitro of cells derived from a human osteosarcoma. *Cancer*, 27: 397–402
- [311] T Kwan, D Benovoy, C Dias, S Gurd, D Serre, H Zuzan, TA Clark, A Schweitzer, MK Staples et al. 2007. Heritability of alternative splicing in the human genome. *Genome Res*, 17:1210–1218. doi: 10.1101/gr.6281007
- [312] J Hull, S Campino, K Rowlands, MS Chan, RR Copley, MS Taylor, K Rockett, G Elvidge, B Keating et al. 2007. Identification of common genetic variation that modulates alternative splicing. *PLoS Genet*, 3:e99. doi: 10.1371/journal.pgen.0030099
- [313] ET Wang, R Sandberg, S Luo, I Khrebtukova, L Zhang, C Mayr, SF Kingsmore, GP Schroth and CB Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456:470–476. doi: 10.1038/nature07509
- [314] Y Barash, JA Calarco, W Gao, Q Pan, X Wang, O Shai, BJ Blencowe and BJ Frey. 2010. Deciphering the splicing code. *Nature*, 465:53–59. doi: 10.1038/nature09000
- [315] MT Vahey, ME Nau, LL Jagodzinski, J Valley-Ogunro, M Taubman, NL Michael and MG Lewis. 2002. Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV type 1 RF. *AIDS Res Hum Retroviruses*, 18:179–192. doi: 10.1089/08892220252781239

- [316] AB van 't Wout, GK Lehrman, SA Mikheeva, GC O'Keeffe, MG Katze, RE Bumgarner, GK Geiss and JI Mullins. 2003. Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)T-cell lines. *J Virol*, 77:1392–1402. doi: 10.1128/JVI.77.2.1392-1402.2003
- [317] R Mitchell, CY Chiang, C Berry and F Bushman. 2003. Global analysis of cellular transcription following infection with an HIV-based vector. *Mol Ther*, 8:674–687. doi: 10.1016/S1525-0016(03)00215-6
- [318] M Rotger, KK Dang, J Fellay, EL Heinzen, S Feng, P Descombes, KV Shianna, D Ge, HF Gnathard et al. 2010. Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog*, 6:e1000781. doi: 10.1371/journal.ppat.1000781
- [319] ST Chang, P Sova, X Peng, J Weiss, GL Law, RE Palermo and MG Katze. 2011. Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line. *MBio*, 2. doi: 10.1128/mBio.00134-11
- [320] R Tewhey, JB Warner, M Nakano, B Libby, M Medkova, PH David, SK Kotsopoulos, ML Samuels, JB Hutchison et al. 2009. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol*, 27:1025–1031. doi: 10.1038/nbt.1583
- [321] JC Marioni, CE Mason, SM Mane, M Stephens and Y Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*, 18:1509–1517. doi: 10.1101/gr.079558.108
- [322] R Morin, M Bainbridge, A Fejes, M Hirst, M Krzywinski, T Pugh, H McDonald, R Varhol, S Jones and M Marra. 2008. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*, 45:81–94. doi: 10.2144/000112900
- [323] J Eid, A Fehr, J Gray, K Luong, J Lyle, G Otto, P Peluso, D Rank, P Baybayan et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323:133–138. doi: 10.1126/science.1162986
- [324] M Schaefer, M Brown, W Kilembe, S Allen, Y Guo, E Hunter and E Paxinos. Single-molecule complete HIV-1 genome sequencing from 2 linked transmission pairs. In *Conference on Retroviruses and Opportunistic Infections*, 2012
- [325] E Buratti and FE Baralle. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol*, 24:10505–10514. doi: 10.1128/MCB.24.24.10505-10514.2004
- [326] JA Jablonski, E Buratti, C Stuani and M Caputi. 2008. The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J Virol*, 82:8038–8050. doi: 10.1128/JVI.00721-08

- [327] PJ Shepard and KJ Hertel. 2008. Conserved RNA secondary structures promote alternative splicing. *RNA*, 14:1463–1469. doi: 10.1261/rna.1069408
- [328] M Alló, V Buggiano, JP Fededa, E Petrillo, I Schor, M de la Mata, E Agirre, M Plass, E Eyras et al. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol*, 16:717–724. doi: 10.1038/nsmb.1620
- [329] H Tilgner, C Nikolaou, S Althammer, M Sammeth, M Beato, J Valcrcel and R Guig. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, 16:996–1001. doi: 10.1038/nsmb.1658
- [330] S Schwartz, E Meshorer and G Ast. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*, 16:990–995. doi: 10.1038/nsmb.1659
- [331] TL Crabb, BJ Lam and KJ Hertel. 2010. Retention of spliceosomal components along ligated exons ensures efficient removal of multiple introns. *RNA*, 16:1786–1796. doi: 10.1261/rna.2186510
- [332] K Takahara, U Schwarze, Y Imamura, GG Hoffman, H Toriello, LT Smith, PH Byers and DS Greenspan. 2002. Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptor-site mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I. *Am J Hum Genet*, 71:451–465. doi: 10.1086/342099
- [333] M de la Mata, C Lafaille and AR Kornblihtt. 2010. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA*, 16:904–912. doi: 10.1261/rna.1993510
- [334] AM Zahler, KM Neugebauer, WS Lane and MB Roth. 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science*, 260:219–222. doi: 10.1126/science.8385799
- [335] CW Smith and J Valcárcel. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, 25:381–388. doi: 10.1016/S0968-0004(00)01604-2
- [336] J Ule, G Stefani, A Mele, M Ruggiu, X Wang, B Taneri, T Gaasterland, BJ Blencowe and RB Darnell. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444:580–586. doi: 10.1038/nature05304
- [337] HY Xiong, Y Barash and BJ Frey. 2011. Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context. *Bioinformatics*, 27:2554–2562. doi: 10.1093/bioinformatics/btr444
- [338] JT Witten and J Ule. 2011. Understanding splicing regulation through RNA splicing maps. *Trends Genet*, 27:89–97. doi: 10.1016/j.tig.2010.12.001

- [339] MM O'Reilly, MT McNally and KL Beemon. 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology*, 213:373–385. doi: 10.1006/viro.1995.0010
- [340] BA Amendt, D Hesslein, LJ Chang and CM Stoltzfus. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. *Mol Cell Biol*, 14:3960–3970. doi: 10.1128/MCB.14.6.3960
- [341] JD Levingood, C Rollins, CHJ Mishler, CA Johnson, G Miner, P Rajan, BM Znosko and BS Tolbert. 2012. Solution structure of the HIV-1 exon splicing silencer 3. *J Mol Biol*, 415:680–698. doi: 10.1016/j.jmb.2011.11.034
- [342] M Caputi, M Freund, S Kammler, C Asang and H Schaal. 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J Virol*, 78:6517–6526. doi: 10.1128/JVI.78.12.6517-6526.2004
- [343] C Asang, I Hauber and H Schaal. 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic Acids Res*, 36:1450–1463. doi: 10.1093/nar/gkm1147
- [344] TO Tange, CK Damgaard, S Guth, J Valcrcel and J Kjems. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J*, 20: 5748–5758. doi: 10.1093/emboj/20.20.5748
- [345] A Tranell, EM Feny and S Schwartz. 2010. Serine- and arginine-rich proteins 55 and 75 (SRp55 and SRp75) induce production of HIV-1 vpr mRNA by inhibiting the 5'-splice site of exon 3. *J Biol Chem*, 285:31537–31547. doi: 10.1074/jbc.M109.077453
- [346] CM Stoltzfus and JM Madsen. 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Curr HIV Res*, 4: 43–55. doi: 10.2174/157016206775197655
- [347] P Legrain and M Rosbash. 1989. Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell*, 57:573–583. doi: 10.1016/0092-8674(89)90127-X
- [348] U Fischer, S Meyer, M Teufel, C Heckel, R Lhrmann and G Rautmann. 1994. Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *EMBO J*, 13: 4105–4112
- [349] KA Jones and BM Peterlin. 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem*, 63:717–743. doi: 10.1146/annurev.bi.63.070194.003441
- [350] TW McCloskey, M Ott, E Tribble, SA Khan, S Teichberg, MO Paul, S Pahwa, E Verdin

- and N Chirmule. 1997. Dual role of HIV Tat in regulation of apoptosis in T cells. *J Immunol*, 158:1014–1019
- [351] GR Campbell, E Pasquier, J Watkins, V Bourgarel-Rey, V Peyrot, D Esquieu, P Barbier, J de Mareuil, D Braguer et al. 2004. The glutamine-rich region of the HIV-1 Tat protein is involved in T-cell apoptosis. *J Biol Chem*, 279:48197–48204. doi: 10.1074/jbc.M406195200
 - [352] HB Miller, TJ Robinson, R Gordn, AJ Hartemink and MA Garcia-Blanco. 2011. Identification of Tat-SF1 cellular targets by exon array analysis reveals dual roles in transcription and splicing. *RNA*, 17:665–674. doi: 10.1261/rna.2462011
 - [353] RA Fouchier, BE Meyer, JH Simon, U Fischer, AV Albright, F Gonzlez-Scarano and MH Malim. 1998. Interaction of the human immunodeficiency virus type 1 Vpr protein with the nuclear pore complex. *J Virol*, 72:6004–6013
 - [354] AK Gubitz, W Feng and G Dreyfuss. 2004. The SMN complex. *Exp Cell Res*, 296: 51–56. doi: 10.1016/j.yexcr.2004.03.022
 - [355] C Trapnell, BA Williams, G Pertea, A Mortazavi, G Kwan, MJ van Baren, SL Salzberg, BJ Wold and L Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 28:511–515. doi: 10.1038/nbt.1621
 - [356] MF Rogers, J Thomas, AS Reddy and A Ben-Hur. 2012. SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biol*, 13:R4. doi: 10.1186/gb-2012-13-1-r4
 - [357] A Ryo, Y Suzuki, K Ichiyama, T Wakatsuki, N Kondoh, A Hada, M Yamamoto and N Yamamoto. 1999. Serial analysis of gene expression in HIV-1-infected T cell lines. *FEBS Lett*, 462:182–186. doi: 10.1016/S0014-5793(99)01526-4
 - [358] G Lefebvre, S Desfarges, F Uyttebroeck, M Muoz, N Beerenwinkel, J Rougemont, A Telenti and A Ciuffi. 2011. Analysis of HIV-1 expression level and sense of transcription by high-throughput sequencing of the infected cell. *J Virol*, 85:6205–6211. doi: 10.1128/JVI.00252-11
 - [359] RS Yallow and SA Berson. 1960. Immunoassay of endogenous plasma insulin in man. *J Clin Invest*, 39:1157–1175. doi: 10.1172/JCI104130
 - [360] E Engvall and P Perlmann. 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry*, 8:871–874. doi: 10.1016/0019-2791(71)90454-X
 - [361] BK Van Weemen and AH Schuurs. 1971. Immunoassay using antigen-enzyme conjugates. *FEBS Lett*, 15:232–236. doi: 10.1016/0014-5793(71)80319-8

- [362] JW Ward, AJ Grindon, PM Feorino, C Schable, M Parvin and JR Allen. 1986. Laboratory and epidemiologic evaluation of an enzyme immunoassay for antibodies to HTLV-III. *JAMA*, 256:357–361. doi: 10.1001/jama.1986.03380030059028
- [363] H Towbin, T Staehelin and J Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*, 76:4350–4354
- [364] Centers for Disease Control. 1985. Provisional Public Health Service inter-agency recommendations for screening donated blood and plasma for antibody to the virus causing acquired immunodeficiency syndrome. *MMWR Morb Mortal Wkly Rep*, 34:1–5
- [365] DS Burke and RR Redfield. 1986. False-positive Western blot tests for antibodies to HTLV-III. *JAMA*, 256:347. doi: 10.1001/jama.1986.03380030049013
- [366] DS Burke, JF Brundage, RR Redfield, JJ Damato, CA Schable, P Putman, R Visintine and HI Kim. 1988. Measurement of the false positive rate in a screening program for human immunodeficiency virus infections. *N Engl J Med*, 319:961–964. doi: 10.1056/NEJM198810133191501
- [367] RJ Chappel, KM Wilson and EM Dax. 2009. Immunoassays for the diagnosis of HIV: meeting future needs by enhancing the quality of testing. *Future Microbiol*, 4:963–982. doi: 10.2217/fmb.09.77
- [368] B Weber, EH Fall, A Berger and HW Doerr. 1998. Reduction of diagnostic window by new fourth-generation human immunodeficiency virus screening assays. *J Clin Microbiol*, 36:2235–2239
- [369] B Weber, L Görtler, R Thorstensson, U Michl, A Mühlbacher, P Bürgisser, R Villaescusa, A Eiras, C Gabriel et al. 2002. Multicenter evaluation of a new automated fourth-generation human immunodeficiency virus screening assay with a sensitive antigen detection module and high specificity. *J Clin Microbiol*, 40:1938–1946. doi: 10.1128/JCM.40.6.1938-1946.2002
- [370] WJ Kassler, C Haley, WK Jones, AR Gerber, EJ Kennedy and JR George. 1995. Performance of a rapid, on-site human immunodeficiency virus antibody assay in a public health setting. *J Clin Microbiol*, 33:2899–2902
- [371] Centers for Disease Control and Prevention. 1998. Update: HIV counseling and testing using rapid tests—United States, 1995. *MMWR Morb Mortal Wkly Rep*, 47:211–215
- [372] Centers for Disease Control and Prevention. 2002. Approval of a new rapid test for HIV antibody. *MMWR Morb Mortal Wkly Rep*, 51:1051–1052
- [373] D Gallo, JR George, JH Fitchen, AS Goldstein and MS Hindahl. 1997. Evaluation of a system using oral mucosal transudate for HIV-1 antibody screening and confirmatory

- testing. OraSure HIV Clinical Trials Group. *JAMA*, 277:254–258. doi: 10.1001/jama.1997.03540270080030
- [374] KP Delaney, BM Branson, A Uniyal, PR Kerndt, PA Keenan, K Jafa, AD Gardner, DJ Jamieson and M Bulterys. 2006. Performance of an oral fluid rapid HIV-1/2 test: experience from four CDC studies. *AIDS*, 20:1655–1660. doi: 10.1097/01.aids.0000238412.75324.82
- [375] C Semá Baltazar, C Raposo, IV Jani, D Shodell, D Correia, C Gonçalves da Silva, M Kalou, H Patel and B Parekh. 2014. Evaluation of performance and acceptability of two rapid oral fluid tests for HIV detection in Mozambique. *J Clin Microbiol*, 52: 3544–3548. doi: 10.1128/JCM.01098-14
- [376] TC Granade, BS Parekh, SK Phillips and JS McDougal. 2004. Performance of the OraQuick and Hema-Strip rapid HIV antibody detection assays by non-laboratorians. *J Clin Virol*, 30:229–232. doi: 10.1016/j.jcv.2003.12.006
- [377] N Pant Pai, J Sharma, S Shivkumar, S Pillay, C Vadnais, L Joseph, K Dheda and RW Peeling. 2013. Supervised and unsupervised self-testing for HIV in high- and low-risk populations: a systematic review. *PLoS Med*, 10:e1001414. doi: 10.1371/journal.pmed.1001414
- [378] C Hart, G Schochetman, T Spira, A Lifson, J Moore, J Galphin, J Sninsky and CY Ou. 1988. Direct detection of HIV RNA expression in seropositive subjects. *Lancet*, 2: 596–599. doi: 10.1016/S0140-6736(88)90639-3
- [379] CY Ou, S Kwok, SW Mitchell, DH Mack, JJ Sninsky, JW Krebs, P Feorino, D Warfield and G Schochetman. 1988. DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells. *Science*, 239:295–297. doi: 10.1126/science.3336784
- [380] EF Long. 2011. HIV screening via fourth-generation immunoassay or nucleic acid amplification test in the United States: a cost-effectiveness analysis. *PLoS One*, 6: e27625. doi: 10.1371/journal.pone.0027625
- [381] SA Fiscus, B Cheng, SM Crowe, L Demeter, C Jennings, V Miller, R Respass, W Stevens and FfCHIVRAVLAWG . 2006. HIV-1 viral load assays for resource-limited settings. *PLoS Med*, 3:e417. doi: 10.1371/journal.pmed.0030417
- [382] S Wang, F Xu and U Demirci. 2010. Advances in developing HIV-1 viral load assays for resource-limited settings. *Biotechnol Adv*, 28:770–781. doi: 10.1016/j.biotechadv.2010.06.004
- [383] P Mee, KL Fielding, S Charalambous, GJ Churchyard and AD Grant. 2008. Evaluation of the WHO criteria for antiretroviral treatment failure among adults in South Africa. *AIDS*, 22:1971–1977. doi: 10.1097/QAD.0b013e32830e4cd8
- [384] JJG van Oosterhout, L Brown, R Weigel, JJ Kumwenda, D Mzinganjira, N Saukila,

- B Mhango, T Hartung, S Phiri and MC Hosseinpour. 2009. Diagnosis of antiretroviral therapy failure in Malawi: poor performance of clinical and immunological WHO criteria. *Trop Med Int Health*, 14:856–861. doi: 10.1111/j.1365-3156.2009.02309.x
- [385] MC Hosseinpour, JJG van Oosterhout, R Weigel, S Phiri, D Kamwendo, N Parkin, SA Fiscus, JAE Nelson, JJ Eron and J Kumwenda. 2009. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. *AIDS*, 23:1127–1134. doi: 10.1097/QAD.0b013e32832ac34e
- [386] S Sherrill-Mix, MK Lewinski, M Famiglietti, A Bosque, N Malani, KE Ocwieja, CC Berry, D Looney, L Shan et al. 2013. HIV latency and integration site placement in five cell-based models. *Retrovirology*, 10:90. doi: 10.1186/1742-4690-10-90
- [387] TW Chun, D Finzi, J Margolick, K Chadwick, D Schwartz and RF Siliciano. 1995. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nat Med*, 1:1284–1290
- [388] RT Davey, N Bhat, C Yoder, TW Chun, JA Metcalf, R Dewar, V Natarajan, RA Lemppicki, JW Adelsberger et al. 1999. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc Natl Acad Sci U S A*, 96:15109–15114
- [389] D Finzi, J Blankson, JD Siliciano, JB Margolick, K Chadwick, T Pierson, K Smith, J Lisziewicz, F Lori et al. 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med*, 5:512–517. doi: 10.1038/8394
- [390] JD Siliciano, J Kajdas, D Finzi, TC Quinn, K Chadwick, JB Margolick, C Kovacs, SJ Gange and RF Siliciano. 2003. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med*, 9:727–728. doi: 10.1038/nm880
- [391] LS Weinberger, RD Dar and ML Simpson. 2008. Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet*, 40:466–470. doi: 10.1038/ng.116
- [392] A Singh, B Razooky, CD Cox, ML Simpson and LS Weinberger. 2010. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys J*, 98:L32–L34. doi: 10.1016/j.bpj.2010.03.001
- [393] BS Razooky and LS Weinberger. 2011. Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. *Methods*, 53:68–77. doi: 10.1016/j.ymeth.2010.12.006
- [394] HJ Muller. 1930. Types of visible variations induced by X-rays in *Drosophila*. *J Genet*, 22:299–334

- [395] M Gaszner and G Felsenfeld. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*, 7:703–713. doi: 10.1038/nrg1925
- [396] A Jordan, P Defechereux and E Verdin. 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J*, 20:1726–1738. doi: 10.1093/emboj/20.7.1726
- [397] A Jordan, D Bisgrove and E Verdin. 2003. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J*, 22:1868–1877. doi: 10.1093/emboj/cdg188
- [398] R Pearson, YK Kim, J Hokello, K Lassen, J Friedman, M Tyagi and J Karn. 2008. Epigenetic silencing of human immunodeficiency virus (HIV) transcription by formation of restrictive chromatin structures at the viral long terminal repeat drives the progressive entry of HIV into latency. *J Virol*, 82:12291–12303. doi: 10.1128/JVI.01383-08
- [399] F Romerio, MN Gabriel and DM Margolis. 1997. Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. *J Virol*, 71:9375–9382
- [400] JJ Coull, F Romerio, JM Sun, JL Volker, KM Galvin, JR Davie, Y Shi, U Hansen and DM Margolis. 2000. The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *J Virol*, 74:6790–6799. doi: 10.1128/JVI.74.15.6790-6799.2000
- [401] G He and DM Margolis. 2002. Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. *Mol Cell Biol*, 22:2965–2973. doi: 10.1128/MCB.22.9.2965-2973.2002
- [402] MK Lewinski, D Bisgrove, P Shinn, H Chen, C Hoffmann, S Hannenhalli, E Verdin, CC Berry, JR Ecker and FD Bushman. 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol*, 79:6610–6619. doi: 10.1128/JVI.79.11.6610-6619.2005
- [403] L Shan, HC Yang, SA Rabi, HC Bravo, NS Shroff, RA Irizarry, H Zhang, JB Margolick, JD Siliciano and RF Siliciano. 2011. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J Virol*, 85:5384–5393. doi: 10.1128/JVI.02536-10
- [404] MJ Pace, EH Graf, LM Agosto, AM Mexas, F Male, T Brady, FD Bushman and U O'Doherty. 2012. Directly infected resting CD4+ T cells can produce HIV Gag without spreading infection in a model of HIV latency. *PLoS Pathog*, 8:e1002818. doi: 10.1371/journal.ppat.1002818
- [405] T Lenasi, X Contreras and BM Peterlin. 2008. Transcriptional interference antagonizes

- proviral gene expression to promote HIV latency. *Cell Host Microbe*, 4:123–133. doi: 10.1016/j.chom.2008.05.016
- [406] Y Han, YB Lin, W An, J Xu, HC Yang, K O’Connell, D Dordai, JD Boeke, JD Siliciano and RF Siliciano. 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe*, 4:134–146. doi: 10.1016/j.chom.2008.06.008
- [407] L Shan, K Deng, NS Shroff, CM Durand, SA Rabi, HC Yang, H Zhang, JB Margolick, JN Blankson and RF Siliciano. 2012. Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity*, 36:491–501. doi: 10.1016/j.jimmuni.2012.01.014
- [408] D Boehm, V Calvanese, RD Dar, S Xing, S Schroeder, L Martins, K Aull, PC Li, V Planelles et al. 2013. BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism. *Cell Cycle*, 12:452–462. doi: 10.4161/cc.23309
- [409] A Savarino, A Mai, S Norelli, SE Daker, S Valente, D Rotili, L Altucci, AT Palamarra and E Garaci. 2009. “Shock and kill” effects of class I-selective histone deacetylase inhibitors in combination with the glutathione synthesis inhibitor buthionine sulfoximine in cell line models for HIV-1 quiescence. *Retrovirology*, 6:52. doi: 10.1186/1742-4690-6-52
- [410] NM Archin, AL Liberty, AD Kashuba, SK Choudhary, JD Kuruc, AM Crooks, DC Parker, EM Anderson, MF Kearney et al. 2012. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, 487:482–485. doi: 10.1038/nature11286
- [411] A Bosque and V Planelles. 2009. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood*, 113:58–65. doi: 10.1182/blood-2008-07-168393
- [412] A Bosque and V Planelles. 2011. Studies of HIV-1 latency in an ex vivo model that uses primary central memory T cells. *Methods*, 53:54–61. doi: 10.1016/j.ymeth.2010.10.002
- [413] X Wu, Y Li, B Crise and SM Burgess. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science*, 300:1749–1751. doi: 10.1126/science.1083413
- [414] RS Mitchell, BF Beitzel, ARW Schroder, P Shinn, H Chen, CC Berry, JR Ecker and FD Bushman. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*, 2:e234. doi: 10.1371/journal.pbio.0020234
- [415] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012
- [416] C Berry, S Hannenhalli, J Leipzig and FD Bushman. 2006. Selection of target sites

- for mobile DNA integration in the human genome. *PLoS Comput Biol*, 2:e157. doi: 10.1371/journal.pcbi.0020157
- [417] GP Wang, A Ciuffi, J Leipzig, CC Berry and FD Bushman. 2007. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res*, 17:1186–1194. doi: 10.1101/gr.6286907
 - [418] H Mochizuki, JP Schwartz, K Tanaka, RO Brady and J Reiser. 1998. High-titer human immunodeficiency virus type 1-based vector systems for gene delivery into nondividing cells. *J Virol*, 72:8873–8883
 - [419] Y Han, K Lassen, D Monie, AR Sedaghat, S Shimoji, X Liu, TC Pierson, JB Margolick, RF Siliciano and JD Siliciano. 2004. Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J Virol*, 78:6122–6133. doi: 10.1128/JVI.78.12.6122-6133.2004
 - [420] G Plesa, J Dai, C Baytop, JL Riley, CH June and U O'Doherty. 2007. Addition of deoxynucleosides enhances human immunodeficiency virus type 1 integration and 2LTR formation in resting CD4+ T cells. *J Virol*, 81:13938–13942. doi: 10.1128/JVI.01745-07
 - [421] N Malani. hiReadsProcessor R package. URL <http://github.com/malnirav/hiReadsProcessor>
 - [422] WJ Kent. 2002. BLAT—the BLAST-like alignment tool. *Genome Res*, 12:656–664. doi: 10.1101/gr.229202
 - [423] CC Berry, K Ocwieja, N Malani and FD Bushman. 2014. Comparing DNA integration site clusters with Scan Statistics. *Bioinformatics*, 30:1493–1500. doi: 10.1093/bioinformatics/btu035
 - [424] J Ernst and M Kellis. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*, 28:817–825. doi: 10.1038/nbt.1662
 - [425] AS Hinrichs, D Karolchik, R Baertsch, GP Barber, G Bejerano, H Clawson, M Diekhans, TS Furey, RA Harte et al. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*, 34:D590–D598. doi: 10.1093/nar/gkj144
 - [426] KR Rosenbloom, CA Sloan, VS Malladi, TR Dreszer, K Learned, VM Kirkup, MC Wong, M Maddren, R Fang et al. 2013. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*, 41:D56–D63. doi: 10.1093/nar/gks1172
 - [427] J Han, SG Park, JB Bae, J Choi, JM Lyu, SH Park, HS Kim, YJ Kim, S Kim and TY Kim. 2012. The characteristics of genome-wide DNA methylation in naïve CD4+ T cells of patients with psoriasis or atopic dermatitis. *Biochem Biophys Res Commun*, 422:157–163. doi: 10.1016/j.bbrc.2012.04.128

- [428] LR Meyer, AS Zweig, AS Hinrichs, D Karolchik, RM Kuhn, M Wong, CA Sloan, KR Rosenbloom, G Roe et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res*, 41:D64–D69. doi: 10.1093/nar/gks1048
- [429] Z Wang, C Zang, JA Rosenfeld, DE Schones, A Barski, S Cuddapah, K Cui, TY Roh, W Peng et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*, 40:897–903. doi: 10.1038/ng.154
- [430] A Barski, S Cuddapah, K Cui, TY Roh, DE Schones, Z Wang, G Wei, I Chepelev and K Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129:823–837. doi: 10.1016/j.cell.2007.05.009
- [431] Z Wang, C Zang, K Cui, DE Schones, A Barski, W Peng and K Zhao. 2009. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, 138:1019–1031. doi: 10.1016/j.cell.2009.06.049
- [432] DE Schones, K Cui, S Cuddapah, TY Roh, A Barski, Z Wang, G Wei and K Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132: 887–898. doi: 10.1016/j.cell.2008.02.022
- [433] F Hsu, WJ Kent, H Clawson, RM Kuhn, M Diekhans and D Haussler. 2006. The UCSC Known Genes. *Bioinformatics*, 22:1036–1046. doi: 10.1093/bioinformatics/btl048
- [434] J Friedman, T Hastie and R Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*, 33:1–22
- [435] IH Greger, F Demarchi, M Giacca and NJ Proudfoot. 1998. Transcriptional interference perturbs the binding of Sp1 to the HIV-1 promoter. *Nucleic Acids Res*, 26:1294–1301
- [436] A De Marco, C Biancotto, A Knezevich, P Maiuri, C Vardabasso and A Marcello. 2008. Intragenic transcriptional cis-activation of the human immunodeficiency virus 1 does not result in allele-specific inhibition of the endogenous gene. *Retrovirology*, 5:98. doi: 10.1186/1742-4690-5-98
- [437] JS Waye and HF Willard. 1987. Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. *Nucleic Acids Res*, 15:7549–7569. doi: 10.1093/nar/15.18.7549
- [438] J Jurka, VV Kapitonov, A Pavlicek, P Klonowski, O Kohany and J Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*, 110:462–467. doi: 10.1159/000084979
- [439] E Verdin, P Paras and C Van Lint. 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J*, 12: 3249–3259
- [440] C Van Lint, S Emiliani, M Ott and E Verdin. 1996. Transcriptional activation and

chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J*, 15:1112–1120

- [441] KG Lassen, KX Ramyar, JR Bailey, Y Zhou and RF Siliciano. 2006. Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. *PLoS Pathog*, 2:e68. doi: 10.1371/journal.ppat.0020068
- [442] M Dieudonné, P Maiuri, C Biancotto, A Knezevich, A Kula, M Lusic and A Marcello. 2009. Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery. *EMBO J*, 28:2231–2243. doi: 10.1038/emboj.2009.141
- [443] RF Siliciano and WC Greene. 2011. HIV Latency. *Cold Spring Harb Perspect Med*, 1: a007096. doi: 10.1101/cshperspect.a007096
- [444] M Lusic, B Marini, H Ali, B Lucic, R Luzzati and M Giacca. 2013. Proximity to PML nuclear bodies regulates HIV-1 latency in CD4+ T cells. *Cell Host Microbe*, 13: 665–677. doi: 10.1016/j.chom.2013.05.006
- [445] KE Ocieja, S Sherrill-Mix, R Mukherjee, R Custers-Allen, P David, M Brown, S Wang, DR Link, J Olson et al. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res*, 40:10345–10355. doi: 10.1093/nar/gks753
- [446] Q Pan, O Shai, LJ Lee, BJ Frey and BJ Blencowe. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40:1413–1415. doi: 10.1038/ng.259
- [447] F Pagani, M Raponi and FE Baralle. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci U S A*, 102: 6368–6372. doi: 10.1073/pnas.0502288102
- [448] K Wang, R Wernersson and S Brunak. 2011. The strength of intron donor splice sites in human genes displays a bell-shaped pattern. *Bioinformatics*, 27:3079–3084. doi: 10.1093/bioinformatics/btr532
- [449] C Carrera, M Pinilla, L Pérez-Alvarez and MM Thomson. 2010. Identification of unusual and novel HIV type 1 spliced transcripts generated in vivo. *AIDS Res Hum Retroviruses*, 26:815–820. doi: 10.1089/aid.2010.0011
- [450] M Lützelberger, LS Reinert, AT Das, B Berkhout and J Kjems. 2006. A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. *J Biol Chem*, 281: 18644–18651. doi: 10.1074/jbc.M513698200
- [451] J Salfeld, HG Gttlinger, RA Sia, RE Park, JG Sodroski and WA Haseltine. 1990. A tripartite HIV-1 tat-env-rev fusion protein. *EMBO J*, 9:965–970
- [452] S Schwartz, BK Felber, DM Benko, EM Fenyö and GN Pavlakis. 1990. Cloning and

- functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol*, 64:2519–2529
- [453] J Smith, A Azad and N Deacon. 1992. Identification of two novel human immunodeficiency virus type 1 splice acceptor sites in infected T cell lines. *J Gen Virol*, 73 (Pt 7):1825–1828
- [454] JA Jablonski and M Caputi. 2009. Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *J Virol*, 83:981–992. doi: 10.1128/JVI.01801-08
- [455] A Tranell, S Tingsborg, EM Feny and S Schwartz. 2011. Inhibition of splicing by serine-arginine rich protein 55 (SRp55) causes the appearance of partially spliced HIV-1 mRNAs in the cytoplasm. *Virus Res*, 157:82–91. doi: 10.1016/j.virusres.2011.02.010
- [456] H Zhou, M Xu, Q Huang, AT Gates, XD Zhang, JC Castle, E Stec, M Ferrer, B Strulovici et al. 2008. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe*, 4:495–504. doi: 10.1016/j.chom.2008.10.004
- [457] Y Zhu, G Chen, F Lv, X Wang, X Ji, Y Xu, J Sun, L Wu, YT Zheng and G Gao. 2011. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc Natl Acad Sci U S A*, 108:15834–15839. doi: 10.1073/pnas.1101676108
- [458] MJ Saltarelli, E Hadziyannis, CE Hart, JV Harrison, BK Felber, TJ Spira and GN Pavlakis. 1996. Analysis of human immunodeficiency virus type 1 mRNA splicing patterns during disease progression in peripheral blood mononuclear cells from infected individuals. *AIDS Res Hum Retroviruses*, 12:1443–1456. doi: 10.1089/aid.1996.12.1443
- [459] E Delgado, C Carrera, P Nebreda, A Fernndez-Garca, M Pinilla, V Garca, L Prez-Ivarez and MM Thomson. 2012. Identification of new splice sites used for generation of rev transcripts in human immunodeficiency virus type 1 subtype C primary isolates. *PLoS One*, 7:e30574. doi: 10.1371/journal.pone.0030574
- [460] P Grabowski. 2011. Alternative splicing takes shape during neuronal development. *Curr Opin Genet Dev*, 21:388–394. doi: 10.1016/j.gde.2011.03.005
- [461] M Llorian and CWJ Smith. 2011. Decoding muscle alternative splicing. *Curr Opin Genet Dev*, 21:380–387. doi: 10.1016/j.gde.2011.03.006
- [462] JY Ip, A Tong, Q Pan, JD Topp, BJ Blencowe and KW Lynch. 2007. Global analysis of alternative splicing during T-cell activation. *RNA*, 13:563–572. doi: 10.1261/rna.457207
- [463] JD Topp, J Jackson, AA Melton and KW Lynch. 2008. A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA*, 14:2038–2049. doi: 10.1261/rna.1212008

- [464] S Sonza, HP Mutimer, K O'Brien, P Ellery, JL Howard, JH Axelrod, NJ Deacon, SM Crowe and DFJ Purcell. 2002. Selectively reduced tat mRNA heralds the decline in productive human immunodeficiency virus type 1 infection in monocyte-derived macrophages. *J Virol*, 76:12611–12621
- [465] R Collman, JW Balliet, SA Gregory, H Friedman, DL Kolson, N Nathanson and A Srinivasan. 1992. An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1. *J Virol*, 66: 7517–7521
- [466] H Deng, R Liu, W Ellmeier, S Choe, D Unutmaz, M Burkhart, P Di Marzio, S Marmon, RE Sutton et al. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381:661–666. doi: 10.1038/381661a0
- [467] NR Landau and DR Littman. 1992. Packaging system for rapid production of murine leukemia virus vectors with variable tropism. *J Virol*, 66:5110–5113
- [468] N Srinivasakumar, N Chazal, C Helga-Maria, S Prasad, ML Hammarskjöld and D Rekosh. 1997. The effect of viral regulatory protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. *J Virol*, 71:5841–5848
- [469] DC Shugars, MS Smith, DH Glueck, PV Nantermet, F Seillier-Moiseiwitsch and R Swanstrom. 1993. Analysis of human immunodeficiency virus type 1 nef gene sequences present in vivo. *J Virol*, 67:4639–4650
- [470] KJ Travers, CS Chin, DR Rank, JS Eid and SW Turner. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res*, 38:e159. doi: 10.1093/nar/gkq543
- [471] TA Thanaraj and F Clark. 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res*, 29:2581–2593. doi: 10.1093/nar/29.12.2581
- [472] M Aebi, H Hornig, RA Padgett, J Reiser and C Weissmann. 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell*, 47:555–565
- [473] M Burset, IA Seledtsov and VV Solovyev. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res*, 28:4364–4375. doi: 10.1093/nar/28.21.4364
- [474] M Burset, IA Seledtsov and VV Solovyev. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res*, 29:255–259. doi: 10.1093/nar/29.1.255
- [475] N Sheth, X Roca, ML Hastings, T Roeder, AR Krainer and R Sachidanandam. 2006.

- Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res*, 34: 3955–3967. doi: 10.1093/nar/gkl556
- [476] JC Guatelli, TR Gingeras and DD Richman. 1990. Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. *J Virol*, 64:4093–4098
- [477] C Kuiken, B Foley, T Leitner, C Apetrei, B Hahn, I Mizrahi, J Mullins, A Rambaut, S Wolinsky and B Korber. 2010. HIV Sequence Compendium 2010. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, New Mexico. URL <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2010compendium.html>
- [478] C Burge, T Tuschl and P Sharp. 1999. Splicing of precursors to mRNAs by the spliceosomes. *Cold Spring Harbor Monograph Archive*, 37. doi: 10.1101/087969589.37.525
- [479] TEM Abbink and B Berkhout. 2008. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol*, 82:3090–3098. doi: 10.1128/JVI.01479-07
- [480] K Verhoef, PS Bilodeau, JL van Wamel, J Kjems, CM Stoltzfus and B Berkhout. 2001. Repair of a Rev-minus human immunodeficiency virus type 1 mutant by activation of a cryptic splice site. *J Virol*, 75:3495–3500. doi: 10.1128/JVI.75.7.3495-3500.2001
- [481] AM Zahler, CK Damgaard, J Kjems and M Caputi. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J Biol Chem*, 279:10077–10084. doi: 10.1074/jbc.M312743200
- [482] S Sherrill-Mix, K Ocieja and F Bushman. Under Review. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of distinctive genomic repeats. *Retrovirology*
- [483] S Wain-Hobson, P Sonigo, O Danos, S Cole and M Alizon. 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell*, 40:9–17. doi: 10.1016/0092-8674(85)90303-4
- [484] SK Arya, C Guo, SF Josephs and F Wong-Staal. 1985. Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). *Science*, 229:69–73
- [485] RA Marciniak and PA Sharp. 1991. HIV-1 Tat protein promotes formation of more-processive elongation complexes. *EMBO J*, 10:4189–4196
- [486] P Wei, ME Garber, SM Fang, WH Fischer and KA Jones. 1998. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, 92:451–462. doi: 10.1016/S0092-8674(00)80939-3

- [487] S Kanazawa, T Okamoto and BM Peterlin. 2000. Tat competes with CIITA for the binding to P-TEFb and blocks the expression of MHC class II genes in HIV infection. *Immunity*, 12:61–70. doi: 10.1016/S1074-7613(00)80159-4
- [488] M Barboric, JHN Yik, N Czudnochowski, Z Yang, R Chen, X Contreras, M Geyer, B Matija Peterlin and Q Zhou. 2007. Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res*, 35:2003–2012. doi: 10.1093/nar/gkm063
- [489] SK O'Brien, H Cao, R Nathans, A Ali and TM Rana. 2010. P-TEFb kinase complex phosphorylates histone H1 to regulate expression of cellular and HIV-1 genes. *J Biol Chem*, 285:29713–29720. doi: 10.1074/jbc.M110.125997
- [490] L Muniz, S Egloff, B Ughy, BE Jády and T Kiss. 2010. Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. *PLoS Pathog*, 6:e1001152. doi: 10.1371/journal.ppat.1001152
- [491] J Corbeil, D Sheeter, D Genini, S Rought, L Leoni, P Du, M Ferguson, DR Masys, JB Welsh et al. 2001. Temporal gene regulation during HIV-1 infection of human CD4+ T cells. *Genome Res*, 11:1198–1204. doi: 10.1101/gr.180201
- [492] CH Woelk, F Ottone, CR Plotkin, P Du, CD Royer, SE Rought, J Lozach, R Sasik, RS Kornbluth et al. 2004. Interferon gene expression following HIV type 1 infection of monocyte-derived macrophages. *AIDS Res Hum Retroviruses*, 20:1210–1222. doi: 10.1089/0889222042545009
- [493] MD Hyrcza, C Kovacs, M Loutfy, R Halpenny, L Heisler, S Yang, O Wilkins, M Ostrowski and SD Der. 2007. Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. *J Virol*, 81:3477–3486. doi: 10.1128/JVI.01552-06
- [494] JQ Wu, DE Dwyer, WB Dyer, YH Yang, B Wang and NK Saksena. 2008. Transcriptional profiles in CD8+ T cells from HIV+ progressors on HAART are characterized by coordinated up-regulation of oxidative phosphorylation enzymes and interferon responses. *Virology*, 380:124–135. doi: 10.1016/j.virol.2008.06.039
- [495] AJ Smith, Q Li, SW Wietgrefe, TW Schacker, CS Reilly and AT Haase. 2010. Host genes associated with HIV-1 replication in lymphatic tissue. *J Immunol*, 185:5417–5424. doi: 10.4049/jimmunol.1002197
- [496] M Imbeault, K Giguère, M Ouellet and MJ Tremblay. 2012. Exon level transcriptomic profiling of HIV-1-infected CD4(+) T cells reveals virus-induced genes and host environment favorable for viral replication. *PLoS Pathog*, 8:e1002861. doi: 10.1371/journal.ppat.1002861
- [497] P Mohammadi, S Desfarges, I Bartha, B Joos, N Zangerer, M Muoz, HF Gnethard,

- N Beerenswinkel, A Telenti and A Ciuffi. 2013. 24 hours in the life of HIV-1 in a T cell line. *PLoS Pathog*, 9:e1003161. doi: 10.1371/journal.ppat.1003161
- [498] X Peng, P Sova, RR Green, MJ Thomas, MJ Korth, S Proll, J Xu, Y Cheng, K Yi et al. 2014. Deep sequencing of HIV-infected cells: insights into nascent transcription and host-directed therapy. *J Virol*, 88:8768–8782. doi: 10.1128/JVI.00768-14
- [499] C de la Fuente, F Santiago, L Deng, C Eadie, I Zilberman, K Kehn, A Maddukuri, S Baylor, K Wu et al. 2002. Gene expression profile of HIV-1 Tat expressing cells: a close interplay between proliferative and differentiation signals. *BMC Biochem*, 3:14. doi: 10.1186/1471-2091-3-14
- [500] B Langmead, C Trapnell, M Pop and SL Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10:R25. doi: 10.1186/gb-2009-10-3-r25
- [501] GR Grant, MH Farkas, AD Pizarro, NF Lahens, J Schug, BP Brunk, CJ Stoeckert, JB Hogenesch and EA Pierce. 2011. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics*, 27:2518–2528. doi: 10.1093/bioinformatics/btr427
- [502] Q Li, AJ Smith, TW Schacker, JV Carlis, L Duan, CS Reilly and AT Haase. 2009. Microarray analysis of lymphatic tissue reveals stage-specific, gene expression signatures in HIV-1 infection. *J Immunol*, 183:1975–1982. doi: 10.4049/jimmunol.0803222
- [503] A Subramanian, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102:15545–15550. doi: 10.1073/pnas.0506580102
- [504] WJ Kent, CW Sugnet, TS Furey, KM Roskin, TH Pringle, AM Zahler and D Haussler. 2002. The human genome browser at UCSC. *Genome Res*, 12:996–1006. doi: 10.1101/gr.229102
- [505] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin and GPDPS . 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25:2078–2079. doi: 10.1093/bioinformatics/btp352
- [506] RP Subramanian, JH Wildschutte, C Russo and JM Coffin. 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, 8:90. doi: 10.1186/1742-4690-8-90
- [507] G La Mantia, D Maglione, G Pengue, A Di Cristofano, A Simeone, L Lanfrancone and L Lania. 1991. Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embryonal carcinoma cells. *Nucleic Acids Res*, 19:1513–1520

- [508] G La Mantia, B Majello, A Di Cristofano, M Strazzullo, G Minchietti and L Lania. 1992. Identification of regulatory elements within the minimal promoter region of the human endogenous ERV9 proviruses: accurate transcription initiation is controlled by an Inr-like element. *Nucleic Acids Res*, 20:4129–4136. doi: 10.1093/nar/20.16.4129
- [509] KE Plant, SJ Routledge and NJ Proudfoot. 2001. Intergenic transcription in the human beta-globin gene cluster. *Mol Cell Biol*, 21:6507–6514. doi: 10.1128/MCB.21.19.6507-6514.2001
- [510] J Ling, W Pi, R Bollag, S Zeng, M Keskintepe, H Saliman, S Krantz, B Whitney and D Tuan. 2002. The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. *J Virol*, 76:2410–2423. doi: 10.1128/jvi.76.5.2410-2423.2002
- [511] X Yu, X Zhu, W Pi, J Ling, L Ko, Y Takeda and D Tuan. 2005. The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. *J Biol Chem*, 280:35184–35194. doi: 10.1074/jbc.M508138200
- [512] RC Edgar. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113. doi: 10.1186/1471-2105-5-113
- [513] M Rotger, J Dalmau, A Rauch, P McLaren, SE Bosinger, R Martinez, NG Sandler, A Roque, J Liebner et al. 2011. Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *J Clin Invest*, 121:2391–2400. doi: 10.1172/JCI45235
- [514] K Breuer, AK Foroushani, MR Laird, C Chen, A Sribnaia, R Lo, GL Winsor, REW Hancock, FSL Brinkman and DJ Lynn. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*, 41:D1228–D1233. doi: 10.1093/nar/gks1147
- [515] I Rusinova, S Forster, S Yu, A Kannan, M Massee, H Cumming, R Chapman and PJ Hertzog. 2013. Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res*, 41:D1040–D1046. doi: 10.1093/nar/gks1215
- [516] ST Chang, MJ Thomas, P Sova, RR Green, RE Palermo and MG Katze. 2013. Next-generation sequencing of small RNAs from HIV-infected cells identifies phased microRNA expression patterns and candidate novel microRNAs differentially expressed upon infection. *MBio*, 4:e00549–e00512. doi: 10.1128/mBio.00549-12
- [517] Z Kalender Atak, K De Keersmaecker, V Gianfelici, E Geerdens, R Vandepoel, D Pauwels, M Porcu, I Lahortiga, V Brys et al. 2012. High accuracy mutation detection in leukemia on a selected panel of cancer genes. *PLoS One*, 7:e38463. doi: 10.1371/journal.pone.0038463
- [518] ES Patel and LJ Chang. 2012. Synergistic effects of interleukin-7 and pre-T cell

- receptor signaling in human T cell development. *J Biol Chem*, 287:33826–33835. doi: 10.1074/jbc.M112.380113
- [519] M Imbeault, M Ouellet and MJ Tremblay. 2009. Microarray study reveals that HIV-1 induces rapid type-I interferon-dependent p53 mRNA up-regulation in human primary CD4+ T cells. *Retrovirology*, 6:5. doi: 10.1186/1742-4690-6-5
- [520] S Iwase, Y Furukawa, J Kikuchi, M Nagai, Y Terui, M Nakamura and H Yamada. 1997. Modulation of E2F activity is linked to interferon-induced growth suppression of hematopoietic cells. *J Biol Chem*, 272:12406–12414. doi: 10.1074/jbc.272.19.12406
- [521] RW Johnstone, JA Kerry and JA Trapani. 1998. The human interferon-inducible protein, IFI 16, is a repressor of transcription. *J Biol Chem*, 273:17172–17177. doi: 10.1074/jbc.273.27.17172
- [522] BR Williams. 1999. PKR; a sentinel kinase for cellular stress. *Oncogene*, 18:6112–6120. doi: 10.1038/sj.onc.1203127
- [523] CV Ramana, N Grammatikakis, M Chernov, H Nguyen, KC Goh, BR Williams and GR Stark. 2000. Regulation of c-myc expression by IFN-gamma through Stat1-dependent and -independent pathways. *EMBO J*, 19:263–272. doi: 10.1093/emboj/19.2.263
- [524] SL Liang, D Quirk and A Zhou. 2006. RNase L: its biological roles and regulation. *IUBMB Life*, 58:508–514. doi: 10.1080/15216540600838232
- [525] F Maldarelli, C Xiang, G Chamoun and SL Zeichner. 1998. The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. *Virus Res*, 53:39–51
- [526] A Monette, L Ajamian, M López-Lastra and AJ Mouland. 2009. Human immunodeficiency virus type 1 (HIV-1) induces the cytoplasmic retention of heterogeneous nuclear ribonucleoprotein A1 by disrupting nuclear import: implications for HIV-1 gene expression. *J Biol Chem*, 284:31350–31362. doi: 10.1074/jbc.M109.048736
- [527] R Contreras-Galindo, P López, R Vélez and Y Yamamura. 2007. HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) in vitro. *AIDS Res Hum Retroviruses*, 23:116–122. doi: 10.1089/aid.2006.0117
- [528] R Contreras-Galindo, MH Kaplan, S He, AC Contreras-Galindo, MJ Gonzalez-Hernandez, F Kappes, D Dube, SM Chan, D Robinson et al. 2013. HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res*, 23:1505–1513. doi: 10.1101/gr.144303.112
- [529] N Bhardwaj, F Maldarelli, J Mellors and JM Coffin. 2014. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses

- in vivo but not to increased virion production. *J Virol*, 88:11108–11120. doi: 10.1128/JVI.01623-14
- [530] RB Jones, H Song, Y Xu, KE Garrison, AA Buzdin, N Anwar, DV Hunter, S Mujib, V Mihajlovic et al. 2013. LINE-1 retrotransposable element DNA accumulates in HIV-1-infected cells. *J Virol*, 87:13307–13320. doi: 10.1128/JVI.02257-13
- [531] P Medstrand and DL Mager. 1998. Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol*, 72:9782–9787
- [532] C Macfarlane and P Simmonds. 2004. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J Mol Evol*, 59:642–656. doi: 10.1007/s00239-004-2656-1
- [533] K Büscher, U Trefzer, M Hofmann, W Sterry, R Kurth and J Denner. 2005. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res*, 65:4172–4180. doi: 10.1158/0008-5472.CAN-04-2983
- [534] G Howard, R Eiges, F Gaudet, R Jaenisch and A Eden. 2008. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*, 27:404–408. doi: 10.1038/sj.onc.1210631
- [535] RC Iskow, MT McCabe, RE Mills, S Torene, WS Pittard, AF Neuwald, EG Van Meir, PM Vertino and SE Devine. 2010. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, 141:1253–1261. doi: 10.1016/j.cell.2010.05.020
- [536] E Lee, R Iskow, L Yang, O Gokcumen, P Haseley, LJ Luquette, 3rd, JG Lohr, CC Harris, L Ding et al. 2012. Landscape of somatic retrotransposition in human cancers. *Science*, 337:967–971. doi: 10.1126/science.1222077
- [537] SW Criscione, Y Zhang, W Thompson, JM Sedivy and N Neretti. 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics*, 15:583. doi: 10.1186/1471-2164-15-583
- [538] AW Whisnant, HP Bogerd, O Flores, P Ho, JG Powers, N Sharova, M Stevenson, CH Chen and BR Cullen. 2013. In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. *MBio*, 4:e000193. doi: 10.1128/mBio.00193-13
- [539] NF Lahens, IH Kavakli, R Zhang, K Hayer, MB Black, H Dueck, A Pizarro, J Kim, R Irizarry et al. 2014. IVT-seq reveals extreme bias in RNA sequencing. *Genome Biol*, 15:R86. doi: 10.1186/gb-2014-15-6-r86
- [540] RD Hockett, JM Kilby, CA Derdeyn, MS Saag, M Sillers, K Squires, S Chiz, MA Nowak, GM Shaw and RP Bucy. 1999. Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA. *J Exp Med*, 189: 1545–1554. doi: 10.1084/jem.189.10.1545

- [541] RJ De Boer, RM Ribeiro and AS Perelson. 2010. Current estimates for HIV-1 production imply rapid viral clearance in lymphoid tissues. *PLoS Comput Biol*, 6: e1000906. doi: 10.1371/journal.pcbi.1000906
- [542] T Ikeda, J Shibata, K Yoshimura, A Koito and S Matsushita. 2007. Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J Infect Dis*, 195:716–725. doi: 10.1086/510915
- [543] TA Wagner, S McLaughlin, K Garg, CYK Cheung, BB Larsen, S Styrchak, HC Huang, PT Edlefsen, JI Mullins and LM Frenkel. 2014. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*, 345:570–573. doi: 10.1126/science.1256304
- [544] F Maldarelli, X Wu, L Su, FR Simonetti, W Shao, S Hill, J Spindler, AL Ferris, JW Mellors et al. 2014. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*, 345:179–183. doi: 10.1126/science.1254194
- [545] LB Cohn, IT Silva, TY Oliveira, RA Rosales, EH Parrish, GH Learn, BH Hahn, JL Czartoski, MJ McElrath et al. 2015. HIV-1 integration landscape during latent and active infection. *Cell*, 160:420–432. doi: 10.1016/j.cell.2015.01.020
- [546] T Brady, YN Lee, K Ronen, N Malani, CC Berry, PD Bieniasz and FD Bushman. 2009. Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev*, 23:633–642. doi: 10.1101/gad.1762309
- [547] B Marini, A Kertesz-Farkas, H Ali, B Lucic, K Lisek, L Manganaro, S Pongor, R Luzzati, A Recchia et al. 2015. Nuclear architecture dictates HIV-1 integration site selection. *Nature*. doi: 10.1038/nature14226
- [548] M Cavazzana-Calvo, E Payen, O Negre, G Wang, K Hehir, F Fusil, J Down, M Denaro, T Brady et al. 2010. Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia. *Nature*, 467:318–322. doi: 10.1038/nature09328
- [549] S Hacein-Bey-Abina, A Garrigue, GP Wang, J Soulier, A Lim, E Morillon, E Clappier, L Caccavelli, E Delabesse et al. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest*, 118:3132–3142. doi: 10.1172/JCI35700
- [550] A Moiani, Y Paleari, D Sartori, R Mezzadra, A Miccio, C Cattoglio, F Cocchiarella, MR Lidonnici, G Ferrari and F Mavilio. 2012. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest*, 122:1653–1666. doi: 10.1172/JCI61852
- [551] D Cesana, J Sgualdino, L Rudiloso, S Merella, L Naldini and E Montini. 2012. Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *J Clin Invest*, 122:1667–1676. doi: 10.1172/JCI62189

- [552] S Pääbo, DM Irwin and AC Wilson. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem*, 265:4718–4721
- [553] SJ Odelberg, RB Weiss, A Hata and R White. 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Res*, 23:2049–2057. doi: 10.1093/nar/23.11.2049
- [554] XC Zeng and SX Wang. 2002. Evidence that BmTXK beta-BmKCT cDNA from Chinese scorpion *Buthus martensii* Karsch is an artifact generated in the reverse transcription process. *FEBS Lett*, 520:183–4; author reply 185
- [555] B Tasic, CE Nabholz, KK Baldwin, Y Kim, EH Rueckert, SA Ribich, P Cramer, Q Wu, R Axel and T Maniatis. 2002. Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol Cell*, 10:21–33
- [556] M Geiszt, K Lekstrom and TL Leto. 2004. Analysis of mRNA transcripts from the NAD(P)H oxidase 1 (Nox1) gene. Evidence against production of the NADPH oxidase homolog-1 short (NOH-1S) transcript variant. *J Biol Chem*, 279:51661–51668. doi: 10.1074/jbc.M409325200
- [557] J Cocquet, A Chong, G Zhang and RA Veitia. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics*, 88:127–131. doi: 10.1016/j.ygeno.2005.12.013
- [558] CJ McManus, JD Coolon, MO Duff, J Eipper-Mains, BR Graveley and PJ Wittkopp. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res*, 20: 816–825. doi: 10.1101/gr.102491.109
- [559] B Cogné, R Snyder, P Lindenbaum, JB Dupont, R Redon, P Moullier and A Leger. 2014. NGS library preparation may generate artifactual integration sites of AAV vectors. *Nat Med*, 20:577–578. doi: 10.1038/nm.3578
- [560] E Gilboa, SW Mitra, S Goff and D Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell*, 18:93–100. doi: 10.1016/0092-8674(79)90357-X
- [561] GX Luo and J Taylor. 1990. Template switching by reverse transcriptase during DNA synthesis. *J Virol*, 64:4321–4328
- [562] J Houseley and D Tollervey. 2010. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS One*, 5:e12271. doi: 10.1371/journal.pone.0012271
- [563] A Meyerhans, JP Vartanian and S Wain-Hobson. 1990. DNA recombination during PCR. *Nucleic Acids Res*, 18:1687–1691
- [564] DJG Lahr and LA Katz. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques*, 47:857–866. doi: 10.2144/000113219

- [565] W Al-Ahmadi, L Al-Haj, FA Al-Mohanna, RH Silverman and KSA Khabar. 2009. RNase L downmodulation of the RNA-binding protein, HuR, and cellular growth. *Oncogene*, 28:1782–1791. doi: 10.1038/onc.2009.16
- [566] RB Jones, KE Garrison, S Mujib, V Mihajlovic, N Aidarus, DV Hunter, E Martin, VM John, W Zhan et al. 2012. HERV-K-specific T cells eliminate diverse HIV-1/2 and SIV primary isolates. *J Clin Invest*, 122:4473–4489. doi: 10.1172/JCI64560
- [567] K Boller, O Janssen, H Schuldes, RR Tönjes and R Kurth. 1997. Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K. *J Virol*, 71:4581–4588
- [568] KE Garrison, RB Jones, DA Meiklejohn, N Anwar, LC Ndhlovu, JM Chapman, AL Erickson, A Agrawal, G Spotts et al. 2007. T cell responses to human endogenous retroviruses in HIV-1 infection. *PLoS Pathog*, 3:e165. doi: 10.1371/journal.ppat.0030165
- [569] R Tandon, D SenGupta, LC Ndhlovu, RGS Vieira, RB Jones, VA York, VA Vieira, ER Sharp, AA Wiznia et al. 2011. Identification of human endogenous retrovirus-specific T cell responses in vertically HIV-1-infected subjects. *J Virol*, 85:11526–11531. doi: 10.1128/JVI.05418-11
- [570] D SenGupta, R Tandon, RGS Vieira, LC Ndhlovu, R Lown-Hecht, CE Ormsby, L Loh, RB Jones, KE Garrison et al. 2011. Strong human endogenous retrovirus-specific T cell responses are associated with control of HIV-1 in chronic infection. *J Virol*, 85: 6977–6985. doi: 10.1128/JVI.00179-11
- [571] W Pi, Z Yang, J Wang, L Ruan, X Yu, J Ling, S Krantz, C Isaacs, SJ Conway et al. 2004. The LTR enhancer of ERV-9 human endogenous retrovirus is active in oocytes and progenitor cells in transgenic zebrafish and humans. *Proc Natl Acad Sci U S A*, 101:805–810. doi: 10.1073/pnas.0307698100
- [572] XHF Zhang and LA Chasin. 2006. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci USA*, 103:13427–13432. doi: 10.1073/pnas.0603042103
- [573] FA Santoni, J Guerra and J Luban. 2012. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology*, 9:111. doi: 10.1186/1742-4690-9-111
- [574] NV Fuchs, S Loewer, GQ Daley, Z Izsvák, J Löwer and R Löwer. 2013. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. *Retrovirology*, 10:115. doi: 10.1186/1742-4690-10-115
- [575] A Fort, K Hashimoto, D Yamada, M Salimullah, CA Keya, A Saxena, A Bonetti, I Voineagu, N Bertin et al. 2014. Deep transcriptome profiling of mammalian stem

- cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet*, 46:558–566. doi: 10.1038/ng.2965
- [576] J Wang, G Xie, M Singh, AT Ghanbarian, T Raskó, A Szvetnik, H Cai, D Besser, A Prigione et al. 2014. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, 516:405–409. doi: 10.1038/nature13804
- [577] B Joos, M Fischer, H Kuster, SK Pillai, JK Wong, J Böni, B Hirscherl, R Weber, A Trkola et al. 2008. HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci U S A*, 105:16725–16730. doi: 10.1073/pnas.0804192105
- [578] TP Brennan, JO Woods, AR Sedaghat, JD Siliciano, RF Siliciano and CO Wilke. 2009. Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4+ T cells reveals a novel source of residual viremia in patients on antiretroviral therapy. *J Virol*, 83:8470–8481. doi: 10.1128/JVI.02568-08
- [579] TA Wagner, JL McKernan, NH Tobin, KA Tapia, JI Mullins and LM Frenkel. 2013. An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J Virol*, 87:1770–1778. doi: 10.1128/JVI.01985-12
- [580] MF Kearney, J Spindler, W Shao, S Yu, EM Anderson, A O’Shea, C Rehm, C Poethke, N Kovacs et al. 2014. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog*, 10:e1004010. doi: 10.1371/journal.ppat.1004010
- [581] KE Ocwieja, S Sherrill-Mix, C Liu, J Song, H Bau and FD Bushman. 2015. A reverse transcription loop-mediated isothermal amplification assay optimized to detect multiple HIV subtypes. *PLoS One*, 10:e0117852. doi: 10.1371/journal.pone.0117852
- [582] CJL Murray, KF Ortblad, C Guinovart, SS Lim, TM Wolock, DA Roberts, EA Dansereau, N Graetz, RM Barber et al. 2014. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 384:1005–1070. doi: 10.1016/S0140-6736(14)60844-8
- [583] KA Sollis, PW Smit, S Fiscus, N Ford, M Vitoria, S Essajee, D Barnett, B Cheng, SM Crowe et al. 2014. Systematic review of the performance of HIV viral load technologies on plasma samples. *PLoS One*, 9:e85869. doi: 10.1371/journal.pone.0085869
- [584] C Liu, M Mauk, R Gross, FD Bushman, PH Edelstein, RG Collman and HH Bau. 2013. Membrane-based, sedimentation-assisted plasma separator for point-of-care applications. *Anal Chem*, 85:10463–10470. doi: 10.1021/ac402459h
- [585] KA Curtis, DL Rudolph, I Nejad, J Singleton, A Beddoe, B Weigl, P LaBarre and

- SM Owen. 2012. Isothermal amplification using a chemical heating device for point-of-care detection of HIV-1. *PLoS One*, 7:e31432. doi: 10.1371/journal.pone.0031432
- [586] T Notomi, H Okayama, H Masubuchi, T Yonekawa, K Watanabe, N Amino and T Hase. 2000. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*, 28:E63
- [587] KA Curtis, DL Rudolph and SM Owen. 2008. Rapid detection of HIV-1 by reverse-transcription, loop-mediated isothermal amplification (RT-LAMP). *J Virol Methods*, 151:264–270. doi: 10.1016/j.jviromet.2008.04.011
- [588] KA Curtis, DL Rudolph and SM Owen. 2009. Sequence-specific detection method for reverse transcription, loop-mediated isothermal amplification of HIV-1. *J Med Virol*, 81:966–972. doi: 10.1002/jmv.21490
- [589] Y Zeng, X Zhang, K Nie, X Ding, BZ Ring, L Xu, L Dai, X Li, W Ren et al. 2014. Rapid quantitative detection of Human immunodeficiency virus type 1 by a reverse transcription-loop-mediated isothermal amplification assay. *Gene*, 541:123–128. doi: 10.1016/j.gene.2014.03.015
- [590] N Hosaka, N Ndembí, A Ishizaki, S Kageyama, K Numazaki and H Ichimura. 2009. Rapid detection of human immunodeficiency virus type 1 group M by a reverse transcription-loop-mediated isothermal amplification assay. *J Virol Methods*, 157: 195–199. doi: 10.1016/j.jviromet.2009.01.004
- [591] KA Curtis, PL Niedzwiedz, AS Youngpairoj, DL Rudolph and SM Owen. 2014. Real-time detection of HIV-2 by reverse transcription-loop-mediated isothermal amplification. *J Clin Microbiol*, 52:2674–2676. doi: 10.1128/JCM.00935-14
- [592] C Kuiken, H Yoon, W Abfalsterer, B Gaschen, C Lo and B Korber. 2013. Viral genome analysis and knowledge management. *Methods Mol Biol*, 939:253–261. doi: 10.1007/978-1-62703-107-3_16
- [593] M Manak, S Sina, B Anekella, I Hewlett, E Sanders-Buell, V Ragupathy, J Kim, M Vermeulen, SL Stramer et al. 2012. Pilot studies for development of an HIV subtype panel for surveillance of global diversity. *AIDS Res Hum Retroviruses*, 28:594–606. doi: 10.1089/AID.2011.0271
- [594] J Louwagie, FE McCutchan, M Peeters, TP Brennan, E Sanders-Buell, GA Eddy, G van der Groen, K Fransen, GM Gershay-Damet and R Deley. 1993. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS*, 7:769–780
- [595] L Buonaguro, ML Tornesello and FM Buonaguro. 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol*, 81:10209–10219. doi: 10.1128/JVI.00872-07

- [596] NF Parrish, F Gao, H Li, EE Giorgi, HJ Barbian, EH Parrish, L Zajic, SS Iyer, JM Decker et al. 2013. Phenotypic properties of transmitted founder HIV-1. *Proc Natl Acad Sci U S A*, 110:6626–6633. doi: 10.1073/pnas.1304288110
- [597] AG Abimiku, TL Stern, A Zwandor, PD Markham, C Calef, S Kyari, WC Saxinger, RC Gallo, M Robert-Guroff and MS Reitz. 1994. Subgroup G HIV type 1 isolates from Nigeria. *AIDS Res Hum Retroviruses*, 10:1581–1583
- [598] Zhang, Chung and Oldenburg. 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen*, 4:67–73. doi: 10.1177/108705719900400206
- [599] C Liu, E Geva, M Mauk, X Qiu, WR Abrams, D Malamud, K Curtis, SM Owen and HH Bau. 2011. An isothermal amplification reactor with an integrated isolation membrane for point-of-care detection of infectious diseases. *Analyst*, 136:2069–2076. doi: 10.1039/c1an00007a
- [600] CA Spina, J Anderson, NM Archin, A Bosque, J Chan, M Famiglietti, WC Greene, A Kashuba, SR Lewin et al. 2013. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. *PLoS Pathog*, 9:e1003834. doi: 10.1371/journal.ppat.1003834
- [601] G Lehrman, IB Hogue, S Palmer, C Jennings, CA Spina, A Wiegand, AL Landay, RW Coombs, DD Richman et al. 2005. Depletion of latent HIV-1 infection in vivo: a proof-of-concept study. *Lancet*, 366:549–555. doi: 10.1016/S0140-6736(05)67098-5
- [602] NM Archin, M Cheema, D Parker, A Wiegand, RJ Bosch, JM Coffin, J Eron, M Cohen and DM Margolis. 2010. Antiretroviral intensification and valproic acid lack sustained effect on residual HIV-1 viremia or resting CD4+ cell infection. *PLoS One*, 5:e9390. doi: 10.1371/journal.pone.0009390
- [603] AM Spivak, A Andrade, E Eisele, R Hoh, P Bacchetti, NN Bumpus, F Emad, R Buckheit, 3rd, EF McCance-Katz et al. 2014. A pilot study assessing the safety and latency-reversing activity of disulfiram in HIV-1-infected adults on antiretroviral therapy. *Clin Infect Dis*, 58:883–890. doi: 10.1093/cid/cit813
- [604] AL Hill, DIS Rosenbloom, F Fu, MA Nowak and RF Siliciano. 2014. Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. *Proc Natl Acad Sci U S A*, 111:13475–13480. doi: 10.1073/pnas.1406663111
- [605] ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489:57–74. doi: 10.1038/nature11247
- [606] T Barrett, SE Wilhite, P Ledoux, C Evangelista, IF Kim, M Tomashevsky, KA Marshall, KH Phillippy, PM Sherman et al. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*, 41:D991–D995. doi: 10.1093/nar/gks1193

- [607] D Karolchik, GP Barber, J Casper, H Clawson, MS Cline, M Diekhans, TR Dreszer, PA Fujita, L Guruvadoo et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res*, 42:D764–D770. doi: 10.1093/nar/gkt1168
- [608] M Goldman, B Craft, T Swatloski, M Cline, O Morozova, M Diekhans, D Haussler and J Zhu. 2015. The UCSC Cancer Genomics Browser: update 2015. *Nucleic Acids Res*, 43:D812–D817. doi: 10.1093/nar/gku1073
- [609] F Cunningham, MR Amode, D Barrell, K Beal, K Billis, S Brent, D Carvalho-Silva, P Clapham, G Coates et al. 2015. Ensembl 2015. *Nucleic Acids Res*, 43:D662–D669. doi: 10.1093/nar/gku1010
- [610] ML Metzker. 2010. Sequencing technologies - the next generation. *Nat Rev Genet*, 11: 31–46. doi: 10.1038/nrg2626
- [611] ER Mardis. 2011. A decade's perspective on DNA sequencing technology. *Nature*, 470: 198–203. doi: 10.1038/nature09796
- [612] K Wetterstrand. 2015. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). URL www.genome.gov/sequencingcosts
- [613] DP Depledge, AL Palser, SJ Watson, IYC Lai, ER Gray, P Grant, RK Kanda, E Leproust, P Kellam and J Breuer. 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One*, 6:e27805. doi: 10.1371/journal.pone.0027805
- [614] TR Mercer, MB Clark, J Crawford, ME Brunck, DJ Gerhardt, RJ Taft, LK Nielsen, ME Dinger and JS Mattick. 2014. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat Protoc*, 9:989–1009. doi: 10.1038/nprot.2014.058
- [615] JJ Mosher, B Bowman, EL Bernberg, O Shevchenko, J Kan, J Korlach and LA Kaplan. 2014. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *J Microbiol Methods*, 104:59–60. doi: 10.1016/j.mimet.2014.06.012
- [616] AS Mikheyev and MMY Tin. 2014. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour*, 14:1097–1102. doi: 10.1111/1755-0998.12324
- [617] M Jain, IT Fiddes, KH Miga, HE Olsen, B Paten and M Akeson. 2015. Improved data analysis for the MinION nanopore sequencer. *Nat Methods*, 12:351–356. doi: 10.1038/nmeth.3290
- [618] A Kilianski, JL Haas, EJ Corriveau, AT Liem, KL Willis, DR Kadavy, CN Rosenzweig and SS Minot. 2015. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience*, 4:12. doi: 10.1186/s13742-015-0051-z

- [619] S Jünemann, FJ Sedlazeck, K Prior, A Albersmeier, U John, J Kalinowski, A Mellmann, A Goesmann, A von Haeseler et al. 2013. Updating benchtop sequencing performance comparison. *Nat Biotechnol*, 31:294–296. doi: 10.1038/nbt.2522
- [620] Illumina, Inc. 2015. System specification sheet: MiSeq system. URL <http://www.illumina.com/products/miseq-reagent-kit-v3.html>
- [621] D Rossell, C Stephan-Otto Attolini, M Kroiss and A Stöcker. 2014. Quantifying alternative splicing from paired-end RNA-sequencing data. *Ann Appl Stat*, 8:309–330. doi: 10.1214/13-AOAS687
- [622] N Bray, H Pimentel, P Melsted and L Pachter. 2015. Near-optimal RNA-Seq quantification. *arXiv preprint*, page 1505.02710
- [623] NL Michael, MT Vahey, L d'Arcy, PK Ehrenberg, JD Mosca, J Rappaport and RR Redfield. 1994. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J Virol*, 68:979–987
- [624] S Landry, M Halin, S Lefort, B Audet, C Vaquero, JM Mesnard and B Barbeau. 2007. Detection, characterization and regulation of antisense transcripts in HIV-1. *Retrovirology*, 4:71. doi: 10.1186/1742-4690-4-71
- [625] NCT Schopman, M Willemsen, YP Liu, T Bradley, A van Kampen, F Baas, B Berkhouit and J Haasnoot. 2012. Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs. *Nucleic Acids Res*, 40:414–427. doi: 10.1093/nar/gkr719
- [626] M Kobayashi-Ishihara, M Yamagishi, T Hara, Y Matsuda, R Takahashi, A Miyake, K Nakano, T Yamochi, T Ishida and T Watanabe. 2012. HIV-1-encoded antisense RNA suppresses viral replication for a prolonged period. *Retrovirology*, 9:38. doi: 10.1186/1742-4690-9-38
- [627] S Saayman, A Ackley, AMW Turner, M Famiglietti, A Bosque, M Clemson, V Planelles and KV Morris. 2014. An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol Ther*, 22:1164–1175. doi: 10.1038/mt.2014.29
- [628] CT Berger, A Llano, JM Carlson, ZL Brumme, MA Brockman, S Cedeño, PR Harrigan, DE Kaufmann, D Heckerman et al. 2015. Immune screening identifies novel T cell targets encoded by antisense reading frames of HIV-1. *J Virol*, 89:4015–4019. doi: 10.1128/JVI.03435-14
- [629] LB Ludwig, JL Ambrus, KA Krawczyk, S Sharma, S Brooks, CB Hsiao and SA Schwartz. 2006. Human Immunodeficiency Virus-Type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. *Retrovirology*, 3:80. doi: 10.1186/1742-4690-3-80
- [630] C Torresilla, É Larocque, S Landry, M Halin, Y Coulombe, JY Masson, JM Mesnard

- and B Barbeau. 2013. Detection of the HIV-1 minus-strand-encoded antisense protein and its association with autophagy. *J Virol*, 87:5089–5105. doi: 10.1128/JVI.00225-13
- [631] JZ Levin, M Yassour, X Adiconis, C Nusbaum, DA Thompson, N Friedman, A Gnirke and A Regev. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods*, 7:709–715. doi: 10.1038/nmeth.1491
- [632] J Podnar, H Deiderick, G Huerta and S Hunicke-Smith. 2014. Next-generation sequencing RNA-Seq library construction. *Curr Protoc Mol Biol*, 106:4.21.1–4.21.19. doi: 10.1002/0471142727.mb0421s106
- [633] S Cardinaud, A Moris, M Février, PS Rohrlich, L Weiss, P Langlade-Demoyen, FA Lemonnier, O Schwartz and A Habel. 2004. Identification of cryptic MHC I-restricted epitopes encoded by HIV-1 alternative reading frames. *J Exp Med*, 199: 1053–1063. doi: 10.1084/jem.20031869
- [634] CT Berger, JM Carlson, CJ Brumme, KL Hartman, ZL Brumme, LM Henry, PC Rosato, A Piechocka-Trocha, MA Brockman et al. 2010. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *J Exp Med*, 207:61–75. doi: 10.1084/jem.20091808
- [635] NT Ingolia, S Ghaemmaghami, JRS Newman and JS Weissman. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324:218–223. doi: 10.1126/science.1168978
- [636] NT Ingolia, LF Lareau and JS Weissman. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147: 789–802. doi: 10.1016/j.cell.2011.10.002
- [637] NT Ingolia. 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet*, 15:205–213. doi: 10.1038/nrg3645
- [638] NJ Maness, AD Walsh, SM Piaskowski, J Furlott, HL Kolar, AT Bean, NA Wilson and DI Watkins. 2010. CD8+ T cell recognition of cryptic epitopes is a ubiquitous feature of AIDS virus infection. *J Virol*, 84:11569–11574. doi: 10.1128/JVI.01419-10
- [639] A Bet, EA Maze, A Bansal, S Sterrett, A Gross, S Graff-Dubois, A Samri, A Guihot, C Katlama et al. 2015. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology*, 12:15. doi: 10.1186/s12977-015-0135-y
- [640] S Sonza, HP Mutimer, R Oelrichs, D Jardine, K Harvey, A Dunne, DF Purcell, C Birch and SM Crowe. 2001. Monocytes harbour replication-competent, non-latent HIV-1 in patients on highly active antiretroviral therapy. *AIDS*, 15:17–22
- [641] M Hermankova, JD Siliciano, Y Zhou, D Monie, K Chadwick, JB Margolick, TC Quinn and RF Siliciano. 2003. Analysis of human immunodeficiency virus type 1 gene

- expression in latently infected resting CD4+ T lymphocytes in vivo. *J Virol*, 77: 7383–7392. doi: 10.1128/JVI.77.13.7383-7392.2003
- [642] N Soriano-Sarabia, RE Bateson, NP Dahl, AM Crooks, JD Kuruc, DM Margolis and NM Archin. 2014. Quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. *J Virol*, 88:14070–14077. doi: 10.1128/JVI.01900-14
- [643] BF Keele, EE Giorgi, JF Salazar-Gonzalez, JM Decker, KT Pham, MG Salazar, C Sun, T Grayson, S Wang et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA*, 105:7552–7557. doi: 10.1073/pnas.0802203105
- [644] JF Salazar-Gonzalez, MG Salazar, BF Keele, GH Learn, EE Giorgi, H Li, JM Decker, S Wang, J Baalwa et al. 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med*, 206:1273–1289. doi: 10.1084/jem.20090378
- [645] MP Wentz, BE Moore, MW Cloyd, SM Berget and LA Donehower. 1997. A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. *J Virol*, 71:8542–8551
- [646] JM Madsen and CM Stoltzfus. 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. *J Virol*, 79:10478–10486. doi: 10.1128/JVI.79.16.10478-10486.2005
- [647] S Paca-Uccaralertkun, CK Damgaard, P Auewarakul, A Thitithanyanont, P Suphaphiphat, M Essex, J Kjems and TH Lee. 2006. The effect of a single nucleotide substitution in the splicing silencer in the tat/rev intron on HIV type 1 envelope expression. *AIDS Research & Human Retroviruses*, 22:76–82. doi: 10.1089/aid.2006.22.76
- [648] D Mandal, Z Feng and CM Stoltzfus. 2008. Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. *J Virol*, 82:1600–1604. doi: 10.1128/JVI.02295-07
- [649] R Wong, A Balachandran, AY Mao, W Dobson, S Gray-Owen and A Cochrane. 2011. Differential effect of CLK SR Kinases on HIV-1 gene expression: potential novel targets for therapy. *Retrovirology*, 8:47. doi: 10.1186/1742-4690-8-47
- [650] RW Wong, A Balachandran, MA Ostrowski and A Cochrane. 2013. Digoxin suppresses HIV-1 replication by altering viral RNA processing. *PLoS Pathog*, 9:e1003241. doi: 10.1371/journal.ppat.1003241
- [651] TH Finkel, G Tudor-Williams, NK Banda, MF Cotton, T Curiel, C Monks, TW Baba, RM Ruprecht and A Kupfer. 1995. Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes. *Nat Med*, 1:129–134. doi: 10.1038/nm0295-129

- [652] G Doitsh, NLK Galloway, X Geng, Z Yang, KM Monroe, O Zepeda, PW Hunt, H Hatano, S Sowinski et al. 2014. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature*, 505:509–514. doi: 10.1038/nature12940
- [653] B Bahbouhi and L al Harthi. 2004. Enriching for HIV-infected cells using anti-gp41 antibodies indirectly conjugated to magnetic microbeads. *Biotechniques*, 36:139–147
- [654] S Hrvatin, F Deng, CW O'Donnell, DK Gifford and DA Melton. 2014. MARIS: method for analyzing RNA following intracellular sorting. *PLoS One*, 9:e89459. doi: 10.1371/journal.pone.0089459
- [655] KM Monroe, Z Yang, JR Johnson, X Geng, G Doitsh, NJ Krogan and WC Greene. 2014. IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV. *Science*, 343:428–432. doi: 10.1126/science.1243640
- [656] M Wilkinson. 1988. A rapid and convenient method for isolation of nuclear, cytoplasmic and total cellular RNA. *Nucleic Acids Res*, 16:10934. doi: 10.1093/nar/16.22.1093
- [657] HW Trask, R Cowper-Sal-lari, MA Sartor, J Gui, CV Heath, J Renuka, AJ Higgins, P Andrews, M Korc et al. 2009. Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. *RNA*, 15: 1917–1928. doi: 10.1261/rna.1677409
- [658] BW Solnestam, H Stranneheim, J Hällman, M Käller, E Lundberg, J Lundeberg and P Akan. 2012. Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics*, 13:574. doi: 10.1186/1471-2164-13-574
- [659] M Lagos-Quintana, R Rauhut, W Lendeckel and T Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858. doi: 10.1126/science.1064921
- [660] V Ambros. 2004. The functions of animal microRNAs. *Nature*, 431:350–355. doi: 10.1038/nature02871
- [661] P Landgraf, M Rusu, R Sheridan, A Sewer, N Iovino, A Aravin, S Pfeffer, A Rice, AO Kamphorst et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129:1401–1414. doi: 10.1016/j.cell.2007.04.040
- [662] Z Klase, P Kale, R Winograd, MV Gupta, M Heydarian, R Berro, T McCaffrey and F Kashanchi. 2007. HIV-1 TAR element is processed by Dicer to yield a viral micro-RNA involved in chromatin remodeling of the viral LTR. *BMC Mol Biol*, 8:63. doi: 10.1186/1471-2199-8-63
- [663] DL Ouellet, I Plante, P Landry, C Barat, ME Janelle, L Flamand, MJ Tremblay and P Provost. 2008. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Res*, 36:2353–2365. doi: 10.1093/nar/gkn076

- [664] Z Klase, R Winograd, J Davis, L Carpio, R Hildreth, M Heydarian, S Fu, T McCaffrey, E Meiri et al. 2009. HIV-1 TAR miRNA protects against apoptosis by altering cellular gene expression. *Retrovirology*, 6:18. doi: 10.1186/1742-4690-6-18
- [665] S Omoto, M Ito, Y Tsutsumi, Y Ichikawa, H Okuyama, EA Brisibe, NK Saksena and YR Fujii. 2004. HIV-1 nef suppression by virally encoded microRNA. *Retrovirology*, 1: 44. doi: 10.1186/1742-4690-1-44
- [666] C Chable-Bessia, O Meziane, D Latreille, R Triboulet, A Zamborlini, A Wagschal, JM Jacquet, J Reynes, Y Levy et al. 2009. Suppression of HIV-1 replication by microRNA effectors. *Retrovirology*, 6:26. doi: 10.1186/1742-4690-6-26
- [667] S Pfeffer, A Sewer, M Lagos-Quintana, R Sheridan, C Sander, FA Grässer, LF van Dyk, CK Ho, S Shuman et al. 2005. Identification of microRNAs of the herpesvirus family. *Nat Methods*, 2:269–276. doi: 10.1038/nmeth746
- [668] TL Sung and AP Rice. 2009. miR-198 inhibits HIV-1 gene expression and replication in monocytes and its mechanism of action appears to involve repression of cyclin T1. *PLoS Pathog*, 5:e1000263. doi: 10.1371/journal.ppat.1000263
- [669] G Swaminathan, F Rossi, LJ Sierra, A Gupta, S Navas-Martín and J Martín-García. 2012. A role for microRNA-155 modulation in the anti-HIV-1 effects of Toll-like receptor 3 stimulation in macrophages. *PLoS Pathog*, 8:e1002937. doi: 10.1371/journal.ppat.1002937
- [670] HS Zhang, TC Wu, WW Sang and Z Ruan. 2012. MiR-217 is involved in Tat-induced HIV-1 long terminal repeat (LTR) transactivation by down-regulation of SIRT1. *Biochim Biophys Acta*, 1823:1017–1023. doi: 10.1016/j.bbampcr.2012.02.014
- [671] HS Zhang, XY Chen, TC Wu, WW Sang and Z Ruan. 2012. MiR-34a is involved in Tat-induced HIV-1 long terminal repeat (LTR) transactivation through the SIRT1/NF κ B pathway. *FEBS Lett*, 586:4203–4207. doi: 10.1016/j.febslet.2012.10.023
- [672] K Chiang, H Liu and AP Rice. 2013. miR-132 enhances HIV-1 replication. *Virology*, 438:1–4. doi: 10.1016/j.virol.2012.12.016
- [673] E Orecchini, M Doria, A Michienzi, E Giuliani, L Vassena, SA Ciafrè, MG Farace and S Galardi. 2014. The HIV-1 Tat protein modulates CD4 expression in human T cells through the induction of miR-222. *RNA Biol*, 11:334–338. doi: 10.4161/rna.28372
- [674] L Farberov, E Herzig, S Modai, O Isakov, A Hizi and N Shomron. 2015. MicroRNA-mediated regulation of p21 and TASK1 cellular restriction factors enhances HIV-1 infection. *J Cell Sci*, 128:1607–1616. doi: 10.1242/jcs.167817
- [675] J Huang, F Wang, E Argyris, K Chen, Z Liang, H Tian, W Huang, K Squires, G Verlinghieri and H Zhang. 2007. Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. *Nat Med*, 13:1241–1247. doi: 10.1038/nm1639

- [676] K Chiang, TL Sung and AP Rice. 2012. Regulation of cyclin T1 and HIV-1 Replication by microRNAs in resting CD4+ T lymphocytes. *J Virol*, 86:3244–3252. doi: 10.1128/JVI.05065-11
- [677] K Chiang and AP Rice. 2012. MicroRNA-mediated restriction of HIV-1 in resting CD4+ T cells and monocytes. *Viruses*, 4:1390–1409. doi: 10.3390/v4091390
- [678] PJ Kanki, DJ Hamel, JL Sankalé, C Hsieh, I Thior, F Barin, SA Woodcock, A Guèye-Ndiaye, E Zhang et al. 1999. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis*, 179:68–73. doi: 10.1086/314557
- [679] P Kaleebu, N French, C Mahe, D Yirrell, C Watera, F Lyagoba, J Nakiyingi, A Rutebeemberwa, D Morgan et al. 2002. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J Infect Dis*, 185:1244–1250. doi: 10.1086/340130
- [680] JM Baeten, B Chohan, L Lavreys, V Chohan, RS McClelland, L Certain, K Mandaliya, W Jaoko and J Overbaugh. 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect Dis*, 195:1177–1180. doi: 10.1086/512682
- [681] N Kiwanuka, O Laeyendecker, M Robb, G Kigozi, M Arroyo, F McCutchan, LA Eller, M Eller, F Makumbi et al. 2008. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J Infect Dis*, 197:707–713. doi: 10.1086/527416
- [682] B Renjifo, P Gilbert, B Chaplin, G Msamanga, D Mwakagile, W Fawzi, M Essex, TV and HIVS Group. 2004. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS*, 18:1629–1636
- [683] GC John-Stewart, RW Nduati, CM Rousseau, DA Mbori-Ngacha, BA Richardson, S Rainwater, DD Pantaleeff and J Overbaugh. 2005. Subtype C Is associated with increased vaginal shedding of HIV-1. *J Infect Dis*, 192:492–496. doi: 10.1086/431514
- [684] W Huang, SH Eshleman, J Toma, S Fransen, E Stawiski, EE Paxinos, JM Whitcomb, AM Young, D Donnell et al. 2007. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol*, 81:7885–7893. doi: 10.1128/JVI.00218-07
- [685] J Snoeck, R Kantor, RW Shafer, K Van Laethem, K Deforche, AP Carvalho, B Wynhoven, MA Soares, P Cane et al. 2006. Discordances between interpretation algorithms for genotypic resistance to protease and reverse transcriptase inhibitors of human immunodeficiency virus are subtype dependent. *Antimicrob Agents Chemother*, 50: 694–701. doi: 10.1128/AAC.50.2.694-701.2006
- [686] PJ Easterbrook, M Smith, J Mullen, S O'Shea, I Chrystie, A de Ruiter, ID Tatt, AM Geretti and M Zuckerman. 2010. Impact of HIV-1 viral subtype on disease

progression and response to antiretroviral therapy. *J Int AIDS Soc*, 13:4. doi: 10.1186/1758-2652-13-4

- [687] AU Scherrer, B Ledergerber, V von Wyl, J Böni, S Yerly, T Klimkait, P Bürgisser, A Rauch, B Hirscher et al. 2011. Improved virological outcome in White patients infected with HIV-1 non-B subtypes compared to subtype B. *Clin Infect Dis*, 53: 1143–1152. doi: 10.1093/cid/cir669
- [688] C Liu, MM Sadik, MG Mauk, PH Edelstein, FD Bushman, R Gross and HH Bau. 2014. Nuclemeter: a reaction-diffusion based method for quantifying nucleic acids undergoing enzymatic amplification. *Sci Rep*, 4:7335. doi: 10.1038/srep07335
- [689] MG Mauk, C Liu, M Sadik and HH Bau. 2015. Microfluidic devices for nucleic acid (NA) isolation, isothermal NA amplification, and real-time detection. *Methods Mol Biol*, 1256:15–40. doi: 10.1007/978-1-4939-2172-0_2
- [690] A Piantadosi, B Chohan, V Chohan, RS McClelland and J Overbaugh. 2007. Chronic HIV-1 infection frequently fails to protect against superinfection. *PLoS Pathog*, 3:e177. doi: 10.1371/journal.ppat.0030177
- [691] RLR Powell, MM Urbanski, S Burda, T Kinge and PN Nyambi. 2009. High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. *J Acquir Immune Defic Syndr*, 50:84–92. doi: 10.1097/QAI.0b013e31818d5a40
- [692] K Ronen, CO McCoy, FA Matsen, DF Boyd, S Emery, K Odem-Davis, W Jaoko, K Mandaliya, RS McClelland et al. 2013. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLoS Pathog*, 9:e1003593. doi: 10.1371/journal.ppat.1003593
- [693] AD Redd, D Ssemwanga, J Vandepitte, SK Wendel, N Ndembí, J Bukenya, S Nakubulwa, H Grosskurth, CM Parry et al. 2014. Rates of HIV-1 superinfection and primary HIV-1 infection are similar in female sex workers in Uganda. *AIDS*, 28:2147–2152. doi: 10.1097/QAD.0000000000000365
- [694] AD Redd, CE Mullis, D Serwadda, X Kong, C Martens, SM Ricklefs, AAR Tobian, C Xiao, MK Grabowski et al. 2012. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J Infect Dis*, 206:267–274. doi: 10.1093/infdis/jis325
- [695] S Jost, MC Bernard, L Kaiser, S Yerly, B Hirscher, A Samri, B Autran, LE Goh and L Perrin. 2002. A patient with HIV-1 superinfection. *N Engl J Med*, 347:731–736. doi: 10.1056/NEJMoa020263
- [696] G Fang, B Weiser, C Kuiken, SM Philpott, S Rowland-Jones, F Plummer, J Kimani, B Shi, R Kaul et al. 2004. Recombination following superinfection by HIV-1. *AIDS*, 18:153–159

- [697] G Blick, RM Kagan, E Coakley, C Petropoulos, L Maroldo, P Greiger-Zanolungo, S Gretz and T Garton. 2007. The probable source of both the primary multidrug-resistant (MDR) HIV-1 strain found in a patient with rapid progression to AIDS and a second recombinant MDR strain found in a chronically HIV-1-infected patient. *J Infect Dis*, 195:1250–1259. doi: 10.1086/512240
- [698] GS Gottlieb, DC Nickle, MA Jensen, KG Wong, RA Kaslow, JC Shepherd, JB Margolick and JI Mullins. 2007. HIV type 1 superinfection with a dual-tropic virus and rapid progression to AIDS: a case report. *Clin Infect Dis*, 45:501–509. doi: 10.1086/520024
- [699] H Streeck, B Li, AFY Poon, A Schneidewind, AD Gladden, KA Power, D Daskalakis, S Bazner, R Zuniga et al. 2008. Immune-driven recombination and loss of control after HIV superinfection. *J Exp Med*, 205:1789–1796. doi: 10.1084/jem.20080281
- [700] O Clerc, S Colombo, S Yerly, A Telenti and M Cavassini. 2010. HIV-1 elite controllers: beware of super-infections. *J Clin Virol*, 47:376–378. doi: 10.1016/j.jcv.2010.01.013
- [701] DM Smith, JK Wong, GK Hightower, CC Ignacio, KK Koelsch, CJ Petropoulos, DD Richman and SJ Little. 2005. HIV drug resistance acquired through superinfection. *AIDS*, 19:1251–1256
- [702] M Pernas, C Casado, R Fuentes, MJ Pérez-Elías and C López-Galíndez. 2006. A dual superinfection and recombination within HIV-1 subtype B 12 years after primoinfection. *J Acquir Immune Defic Syndr*, 42:12–18. doi: 10.1097/01.qai.0000214810.65292.73
- [703] DL Robertson, PM Sharp, FE McCutchan and BH Hahn. 1995. Recombination in HIV-1. *Nature*, 374:124–126. doi: 10.1038/374124b0
- [704] MH Malim and M Emerman. 2001. HIV-1 sequence variation: drift, shift, and attenuation. *Cell*, 104:469–472. doi: 10.1016/S0092-8674(01)00234-3
- [705] SK Gire, A Goba, KG Andersen, RSG Sealson, DJ Park, L Kanneh, S Jalloh, M Momoh, M Fullah et al. 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, 345:1369–1372. doi: 10.1126/science.1259657
- [706] WHO Ebola Response Team. 2014. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med*, 371:1481–1495. doi: 10.1056/NEJMoa1411100
- [707] World Health Organization. 2015. Ebola situation report: 13 May 2014. URL <http://apps.who.int/ebola/en/current-situation/ebola-situation-report-13-may-2015>
- [708] G Chowell and H Nishiura. 2014. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Med*, 12:196. doi: 10.1186/s12916-014-0196-0

- [709] AS Fauci. 2014. Ebola—underscoring the global disparities in health care resources. *N Engl J Med*, 371:1084–1086. doi: 10.1056/NEJMp1409494
- [710] World Health Organization. 2015. Interim guidance on the use of rapid Ebola antigen detection tests. URL <http://www.who.int/csr/resources/publications/ebola/ebola-antigen-detection/en/>
- [711] Y Kurosaki, A Takada, H Ebihara, A Grolla, N Kamo, H Feldmann, Y Kawaoka and J Yasuda. 2007. Rapid and simple detection of Ebola virus by reverse transcription-loop-mediated isothermal amplification. *J Virol Methods*, 141:78–83. doi: 10.1016/j.jviromet.2006.11.031
- [712] T Hoenen, D Safronetz, A Groseth, KR Wollenberg, OA Koita, B Diarra, IS Fall, FC Haidara, F Diallo et al. 2015. Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science*, 348:117–119. doi: 10.1126/science.aaa5646