1	Title: The HERV-K HML-2 integration within RASGRF2 is associated with
2	intravenous drug abuse and modulates transcription in a cell-line model
3	
4	
5	
6	+Correspondence should be addressed to:
7	
8	aris.katzourakis@zoo.ox.ac.uk
9	gmagi@med.uoa.gr
10	
11	
12	This file includes:
13	Materials and Methods
14	Supplementary Figures
15	Other Supplementary Materials for this manuscript includes the following:
16	Dataset S1
17	

18	Materials and Methods
19	
20	Wet Lab
21	
22	Design of target enrichment and high throughput sequencing
23	We developed a custom library-preparation method for the targeted
24	enrichment of HK2 LTR and their flanking sequences. To do this, we designed
25	custom 120bp-long biotinylated probes from IDT to perform a capture of the
26	first and the last ~360bp of the HK2-LTR sequence (5 probes overlapping by
27	60bp at each LTR side). The libraries were prepared with a target insert size
28	at library Quality Control (QC) of 1200bp. From trial runs, we knew that a
29	library QC showing an insert of approximately 1200bp would produce
30	sequencing data with a considerably shorter insert. The prepared libraries
31	were validated and multiplexed before 2X300bp paired-end sequencing on an
32	Illumina MiSeq instrument.
33	
34	Cells and reagents
35	General cell culture reagents were purchased from Sigma-Aldrich
36	(Poole, U.K.) The human embryonic kidney HEK 293 cell line was purchased
37	for this specific project from the European Collection of Authenticated Cell
38	Cultures (ECCAC; Porton Down, U.K.) and the teratocarcinoma cell line
39	NCCIT was obtained from the American Type Culture Collection (ATCC, LGC
40	Standards, Teddington, U.K.). Both of the cell lines came with quality
41	certificates for their identity and being free from mycoplasma contamination.

The cells were incubated in a humidified atmosphere at 37°C with 5% CO<sub>2</sub>
and cultured as described below.

44 HEK 293 cells were cultured in Dulbecco's Modified Eagle's Medium 45 (DMEM) with high glucose (4500mg/dL), L-glutamine and sodium 46 bicarbonate. It was supplemented with 10% heat-inactivated foetal bovine 47 serum, 1% penicillin-streptomycin and 2mM L-glutamine. The cells were 48 allowed to grow to 100% confluence in a T75 flask and were typically 49 subcultured twice per week using trypsin-EDTA to dislodge them from the 50 substratum. The cell suspension was diluted 1:20 into a new flask. Two days 51 after seeding a flask, the media was replaced and the spent media filtered 52 through a 0.2µM syringe filter. Aliquots of the filtered media were stored at -53 80°C until needed for use as conditioned media in genome engineering. 54 The NCCIT cells were cultured in RPMI growth medium supplemented

55 with 10% v/v foetal bovine serum, 1% v/v penicillin-streptomycin and 2mM L-56 glutamine. Two days after seeding the flask, the media was replaced with 57 fresh growth media due to the elevated respiration rate. They were allowed to 58 grow to 100% confluence in a T75 flask and were typically subcultured every 59 four days using trypsin-EDTA to dislodge them from the substratum. On each passage, approximately 1x10<sup>6</sup> cells were reserved for later gDNA extraction. 60 61 SCR7 was obtained from Stratech Scientific Limited (Newmarket, 62 U.K.). Puromycin was purchased from Cayman Chemicals (Cambridge 63 Bioscience, Cambridge, U.K.)

64

65 Genome Engineering

Page 4 of 19

### 67 -CRISPR/Cas9 plasmids

68 The Cas9 and sgRNA plasmids were designed by Andrew Bassett and Joey 69 Riepsaame of Genome Engineering Oxford (GEO) to introduce the HERV 70 sequence into the RASGRF2 gene. We designed the HDR plasmid and this 71 was synthesised using GeneArt (Thermo Fisher Scientific). Subsequently, for 72 the HAP1 transfection experiments we sub-cloned the HDR construct into a 73 new backbone to remove the CMV promoter and increase targeting 74 efficiencies. This construct was converted into a minicircle (Supplementary 75 Figure 3).

76

### 77 -Transfection of cell-lines with CRISPR/Cas9 plasmids

78 HEK 293 cells were seeded into a 6-well plate at a density of 1x10<sup>6</sup> 79 cells/well. The following day, the cells were co-transfected with 1µg each of 80 the sgRNA and HDR plasmids using calcium phosphate. The transfection 81 mixture was left on the cells overnight and then replaced with fresh growth 82 media. The cells were allowed to recover for 24h, then the selection agent 83 (puromycin) was added. Selection was conducted for 7d, with the media 84 (containing selection agent) replaced every 1-2d. The cells were then 85 examined by PCR for the presence of the insertion. 86 Subsequently, cells that were positive for the integration were subjected to 87 limiting dilution in order to create clonal cultures. Briefly, the cells were 88 detached using trypsin, counted and a dilution made to produce a 89 concentration of 100 cells/mL in 80% normal growth media and 20% 90 conditioned media. These cells were then used to seed a 96 well plate

91 (100 $\mu$ L/well). The next day, the plate was examined and wells containing a

Page 5 of 19

92 single cell were marked. The plate was then returned to the incubator and 93 monitored for 7-10 days. When the positive wells reached a high density, they 94 were subcultured and used to seed cluster plates with larger wells. This was 95 repeated until the cells were growing in 6wp. Finally, once the density of the 96 6-well plate was high enough, some cells were removed for gDNA extraction 97 (Qiagen DNeasy Blood, Cells and Tissue kit) and PCR analysis. 98 A refinement of the above method involved using the NHEJ inhibitor 99 SCR7. Briefly, the cells were seeded as above for transfection. The following 100 day, the cells were treated with SCR7 2h before transfection. Importantly, the 101 SCR7-containing media was not removed prior to the transfection. The 102 transfection procedure was then conducted as described above. 103 104 - CRISPR/Cas9 editing of eHAP1 cells 105 Since HEK293s select the WT alleles over those with the RASGRF2-int, we 106 tried to edit a haploid cell line (eHAP1), using the same HDR synthesised 107 plasmid as for the HEK293s. For the integration screening, we used two 108 different PCRs. Each consisted of a primer binding within the 109 newly introduced DNA, the other binding outside (recognizing the WT 110 sequence). We succeeded in screening individually picked and expanded 111 clones but unfortunately, none of these survived upon expansion to have 112 integrated the HDR construct (Supplementary Figure 3). In our 2nd attempt, 113 we first sub-cloned the HDR construct into a new backbone to remove the 114 promoter and increase targeting efficiencies. This construct was converted 115 into a minicircle, devoid of any bacterial backbone (improves HDR 116 efficiencies) (more info at: www.systembio.com/products/gene-expression-

117	systems/minicircle-technology/) eHap1 cells were targeted once more, using
118	200.000 cells - 5 ug Cas9 protein - 1 ug sgRNA - 1 ug minicircle DNA. We
119	Nucleofected the cells following the Horizon Discovery protocol (1575V - 10
120	ms - 3 pulses). After nucleofection, cells were allowed to recover for 48h and
121	then seeded at 5000 cells per 10 cm petri dish to grow into
122	individual colonies. The remainder of cells (bulk) was used for
123	triple genotyping PCR analysis. Unfortunately, the screening of the bulk DNA
124	showed no sign of integrated DNA
125	
126	-Controls of normal RASGRF2 transcription
127	We used as normal controls for the RASGRF2 the unedited HEK293
128	as well as edited HEK293s that "lost" the RASGRF2-int 48-days post-
129	transfection (see also Figure 1). The first one is a "wild type" control, while the
130	second one is a control that shows that transcription of RASGRF2 is restored
131	when the integration is lost.
132	
133	PCR screening for RASGRF2-int patient samples and qPCR on cell-lines
134	To estimate the prevalence of the RASGRF2-polymorphic integration in
135	PWIDs and non-PWIDs, we designed two pairs of primers (one common
136	reverse), which specifically amplify either the allele with the integration alone
137	(primers LTR_splc_F - dn_intg_R) or both the wild type and the integrated

alleles, giving distinct PCR products (primers up\_intg\_F - dn\_intg\_R) (DatasetS1). This method allows the screening and the genotyping of multiple samples

140 simultaneously (Figure 1, a-d).

141 We further investigated the impact of the HK2-LTR integration on the 142 expression of the host RASGRF2 gene. We used CRISPR/Cas9 technology to edit the HEK293 cell line and to manually insert the LTR sequence in the 143 pre-integration intronic site between exons 17 and 18 (see above). Before 144 145 proceeding with the CRISPR/Cas9 engineering experiments we genotyped 146 our 293s cell line (ECACC 85120602) and we found it negative for the 147 RASGRF2 LTR integration. Using publicly available data from www.hek293genome.org/v1/data.php we retrieved 148 the rs26907 SNP 149 genotypes for most of the available 293s derivatives and we found them 150 heterozygous. In addition, we did not find rs26907 in published Neanderthal 151 genomes whereas RASGRF2-int has been identified, while the frequency of 152 rs26907 in East Asian populations is similar to other populations, where 153 RASGRF2-int has not been identified (see main text). These observations do 154 not indicate linkage between rs26907 and RASGRF2-int..

155 The LTR and the flanking host sequences were previously confirmed 156 with Sanger step-sequencing in patient samples positive for the integration. 157 We confirmed the integration and we genotyped the cell line in multiple postclonal-selection time-points using allele-specific PCR designs. The edited 158 159 allele was present in the heterozygous state in the selected clones, as the 160 wild-type allele-specific PCR was also positive after the transfection. The 161 forward primer (into conf F) of the primer-set used for the confirmation of the integration, was annealing upstream to the left homologous arm used in the 162 163 HDR plasmid, allowing us to distinguish the edited alleles from the HDR 164 plasmid (Figure 1).

The expression levels of RASGRF2 exons were evaluated using multiple SYBR-Green qPCRs (primers RAS2-3, 16-17, 17-18, 18-19). We found decreased expression levels of exons around the integration, which were reversed to normal after the selection of the wildtype alleles, 48 days post-transfection (T75 flask).

170

171 Assessment of HK2 LTR promoter in plasmids

172 The HERVK-113 LTR *in vitro* synthesized dsDNA fragment (Invitrogen) 173 was cloned into the HincII restriction site of pUC19 resulting HK-113-174 LTR/pUC19 recombinant plasmid. The HK-113-LTR XhoI excised fragment 175 from HK-113-LTR/pUC19 plasmid was cloned into the XhoI restriction site of 176 pGEM-luc vector (Promega) both in forward and reverse orientation resulting 177 plasmids HK-113-LTR-Luc/F, HK-113-LTR-Luc/R respectively.

The human cell lines Huh7 (hepatocarcinoma) and HeLa (cervical carcinoma) were each maintained in Dulbecco's modified Eagle medium supplemented with 2 mM glutamine, 10% (v/v) heat-inactivated fetal calf serum and 100 U/ml penicillin/streptomycin.

For DNA transfections, Huh7 and HeLa cells were transiently transfected using JetPEI reagent (PolyPlus) according to the manufacturer's protocol. Following transfection, cells were harvested, seeded into 48-well plates, collected at the appropriate time points and subjected to luciferase assays with a commercially available kit (Promega). Luciferase activity was normalized to total cell protein in order to yield relative luciferase activity (RLA).

Huh7 and HeLa cells were transfected with pGEM-luc vector (control) and HK-113-LTR Luc/F, HK-113-LTR-Luc/R plasmids. 48h post-transfection, cells were lysed and luciferase activity was measured; luciferase readouts were normalised to total cell protein.

193

194 **Expression patterns of read-through transcripts** 

Page 9 of 19

We found significantly increased expression levels over the LTR-ends compared to the LTR-starts (Wilcoxon, p<0.05) (Figure 2). We also found that elements with increased expression levels over the LTR-start would present decreased expression (or would be inactive) over the LTR-end and viceversa. (Figure 2)

201 We tested the expression levels in combination with the type but also 202 the estimated age of each element as described in Subramanian et al (1). We 203 found that full-length elements are significantly more active compared to the 204 solo-LTR integrations, over their LTR-end edges (Wilcoxon, p<0.05) (Figure 205 2). We asked for at least 5 reads per fragment to evaluate a HK2-edge as 206 active and we estimated their expression (active vs. not-active) and the 207 magnitude (<500 vs. >500 reads) per edge, in combination with their age 208 (younger than 5my, 5-20my and older than 20my). We found that the 209 youngest elements (younger than 5my) are more likely to be active versus the 210 very old ones (older than 20my) over the LTR-ends. The same trend was 211 observed for the magnitude of the expression, with the youngest elements to 212 be more expressed versus the older ones, both over the LTR-start and the 213 LTR-end regions (z-test, one-tailed, p<0.05) (Figure 2).



## 217 Supplementary Figures

218

219 **Supplementary Figure 1**: Nucleotide alignment and transcription regulatory 220 motifs annotation of: (1) Sanger step sequencing consensus of the RASGRF2 221 - integrated solo LTR. (2) Synthetic fragment inserted into the HDR plasmid, 222 which was used for the CRISPR/Cas9-mediated editing of the HEK293 cells 223 (identical to 1). (3) Synthetic fragment used for the Luciferase expression 224 vector construction. (4, 5) HERV-K 113 LTRs, regions: 1-968bp and 8501-225 9472bp, respectively. The K113 LTR sequences only differs by <1% 226 compared to the RASGRF2 solo LTR integration, while none of the observed 227 sequence differences is located within promoter/regulatory sites, e.g. the four 228 GC-boxes which are targets of the transcription factors Sp1 and Sp3 (2), the 229 transcription initiation site (Inr), the TATA boxes -like and the Beta globin 230 distal promoter -like sequence or the binding site of the transcription factor 231 OTF-2 (3) or the 5'-enhancer complex binding site (nt 62-83) and the YY1 transcription factor (nt 64-80) (4). The U3, R and U5 LTR regions are 232 233 indicated in green, red and blue arrows respectively.

Consensus	Ť	G	ТС	5 G	i G	G	А	А	10 A	A	G	C.	A	A G	i A	G	A	20 G	A 1	- c	A	G	A	тι	ГG	30 T	т	A	с	т	G	т	G	т	с	40 T (	БΤ	G
1. Patient_sanger_consus		•			•	•	•	•	·	·	•	•			•	•	·	•	• •		•	·	•	•		•	·	•	•	•	·	•	•	·	•		•	•
2. HDR_plasmid_insert	·	•		•	•	•	•	·	•	·	•	•	•		·	•	•	•		•	•	·	•	•		•	•	•	•	•	•	•	•	•				•
3. luc_vector_insert	•	•		•	•	•	•	·	·	•	•	•	•		•	·	•	•		•	•	•	•	•		·	•	·	•	•		•	•	•	•		•	•
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:			:	:	:	:	:	:	:	:			:	:	:	:	J3 re	Jion	:	:	:	•		:	:	:	:	:	:	:	:	:	• •	· ·		
Consensus	т	A	GΑ	A A	A A	50 G	А	A	G	т	A	G.	A	5 A	60 T	А	G	G,	AC	БА	с	т	С	70 C /	чτ	т	т	т	G	т	т	A	80 T	G	т.	ΑY	′т	А
1. Patient_sanger_consus	•	•		•	•	·	•	·	•	•	•	•	•		•	•		•				62-8	3_5' (	enha	ncer	com	olex l	oindi	ng s	ite			•	•			•	•
2. HDR_plasmid_insert		•			•	·	·	•	·	•	•	•	•		·	·	•	•				•	Y	Υ1 b	indir	ıg sit	e .	•	•	·				•	•	· c	•	•
3. luc_vector_insert	•	•			•	·	·	•	·	•	•	•			•	·	•	•			•	•	•	•		•	•	·	·	•	•	•	•	•	•	· ٦	• •	•
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:		:	:			:	:	:	:				:	•	:	:	: I3 re	nion	:	:	:	• •	· ·	:	:	:	:	:	:	:	:	:	• •	· (		:
Consensus 1. Patient_sanger_consus	A	G.	۹ • •	0 A A	A ·	А	т ·	т ·	C	т ·	т ·	с <sup>1</sup>	.00 T (	5 C	: C	т ·	т ·	G.	A (	i A	110 T	т ·	с <sup>.</sup>	т ( • •	5 Т	т ·	A	A	т ·	120 C	T ·	A	T I	R A	A ·	с с 	: т	T ·
2. HDR_plasmid_insert		•			•	•	•		•	•	•	•	•		·	•		•			·		•	•		•	•	·	•	•			· .	A				·
3. luc_vector_insert		•			•			•	·	•	•	•	•		•	•		•			•		•	•			·	·	•	•			•	G			•	•
4. K113_LTR_1-968 5. K113_LTR_8501-9472	:	:	· ·	:	:	:		:	:	:	:				:		:	:	 	:		:	:	•	 	:	:	:	:	:	:	:	•	G G		 	:	:
Consensus 1. Patient_sanger_consus	130 A	C ·	с с · ·	: C	: с	А	A	C ·	C ·	140 C	Y T	G	т (	5 C	: т	C ·	т ·	с <sup>1</sup>	.50 T C	gion 5 A •	A	A	с,	А Т • • •	ГG	т ·	160 G	C ·	т ·	G ·	T ·	G	T I	C	M.	A C	70 T	C ·
2. HDR_plasmid_insert		•								•	т													•						•					с			
3. luc_vector_insert											с													•											A			
4. K113_LTR_1-968 5. K113 LTR 8501-9472	•	:	· ·	:	:	÷	:	:	:	:	c c	:	:		:	:	:	:		:	:	:	:		 	÷	:	:	:	:	:	·	:	:	A A	· ·	:	:
	>																		12	nion																		3
	<u> </u>																		<b>J</b> 5 Te	JION																		
Consensus 1. Patient_sanger_consus	A	G.	А G 	5 Т	- т	G	180 A	A	т ·	G	G.	A	т -	ГА 	A	190 G	G	G	C (	G G	т ·	G	с,	A (	po G G	A	т ·	G	т ·	G	C ·	т ·	T -	210 T	G	тт 	Ā	A ·
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert	A	G.	A G 	; т	- т	G ·	180 A	А	т •	G ·	G.	A	T -	ГА 	. A	190 G	G ·	G ·	C (	G G	т •	G ·	с,	2  A (	ро G G	A	т	G	т •	G ·	C •	т ' •	т <sup>2</sup>	210 T	G	тт 	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert	A	G . •	A G · · ·	5 T • •	т , ,	G • •	180 A	A	т • •	G	G.	A	т - • •	ГА  	• A • •	190 G	G ·	G	C (	5 G	т	G	с,	2) A (	po 5 G 	<b>A</b>	т	G ·	т • •	G	C • •	т ' • •	т <sup>2</sup> • •	210 T	G	тт •••	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472	A	G . • •	A G	5 T • •	т • • •	G • • •	180 A	<b>A</b> • •	T • •	G	G . • •	A '		<b>F A</b>	• A • • •	190 G	G • •	G	C (	G G · · · · · · · · · · · · · · · · · ·	т • • •	G	с,	2  • •	po G G	A	т • •	G	т • •	G • •	C • •	т .	T .	210 T ·	G '	тт ••••	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus	A	G . •	A G	5 T	- т	G	180 A	A .	т	G	G . • •	Α • •	T -	Γ Α · · ·	· A · ·	190 G	G • •	G	C (   J <u>3 re</u>	G G	т	G	с,	2) A (		A	T .	G	T .	G • •	C	Т .	2 T · ·	210 T ·	G · ·	T T · · ·	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus	A	G . • •	A G	5 T • • •	ст.	G • • •	180 A	A	T .	G	G . • •	A .	T - G (	Γ Α · · · · · ·	· A · ·	190 G	G	G	C (   J3 reg	jion G G J Jion G C	т	G • • • • • • • • • • • • •	с,	2 A ( · · · · · · ·	ρο 5 G   	A	T G	G	T	G • • • •	C • • • • • • • • • • • • • • • • • • •	т .	2 T · ·	210 T	G · ·	т т • • • • • •	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert	A	G .	A G	5 T • • • •	· · · · · · · · · · · · · · · · · · ·	G • • •	180 A · ·	A	т	G	G .	A	T - G (	Γ Α - · · - · ·	· A · · ·	190 G	G	G ·	C (      	gion G G G G G G G G G G G G G G G G G G G	т	G	с,	21 A ( · · · · · ·	ρο 5 G   	A	T .	G	T • • •	G • • • T	C	T .	T -	210 T ·	G · ·	т т • • • • • •	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert	A	G .	A G	5 T	т - - - - - - - - - - - - - - - - - - -	G	180 A	A	т • • • •	G	G .	A .	T -	F A	· A · · ·	1900 G	G	G · · ·	C (    J3 rec T ( 	jion G G G G G G G G G G G G G G G G G G G	т	G • • • • • • • • • • • • • • •	с,	2; A ( · · · · · ·	ро G G G C A C A C A C A C A C A C A C A C C C C C C C C C C C C C	A	T	G	T	G • • • T	C • • 250 C •	T '	2 T · · ·	210 T · ·	G .	T 1   	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472	A	G .	A G	5 T • • • • • •	200	G • • • • •	180 A	A	т • • • •	G .	G .	A · · · · · · · · · · · · · · · · · · ·	T - G (	Γ Α · · · · · ·	· A · · · · ·	190 G	G	G · · · · · · · · · · · · · · · · · · ·	C (C · · · · · · · · · · · · · · ·	jion Gion Gion	т	G	с,	24 A ( 	ρο G G C A C A C A C A C A C A C A C A	A	T	G	T • • • •	G • • • •	C • • • • • •	T *	2 T · · · ·	210 T · ·	G .	T 1 · · · · · ·	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus		G .	A C · · · · · · · · · · · ·	5 T	<ul> <li>T</li> <li>.</li> <li>.</li></ul>	G	180 A · · · · · ·	A	т. 	G	G .		T - · · · · · · · · · · · · · · · · · ·	Γ Α · · · · · · · · · · · · · · ·	· A · · · · · · · · · · · · · · · · · ·	1900 G · · · · ·	G	G · · · · · · · · · · · · · · · · · · ·	C (C	jion Gion Gion Gion	T	G	C ,	22 A (     		A .	T .	G	T	G • • • • •	C · · · · · · · · ·	Т <sup>•</sup> • • • • •	2 T · · ·	210 T · ·	G .	T 1 · · · · · ·	- Α	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus	A	G .	A C         	5 T	<ul> <li>T</li> <li>.</li> <li>.</li></ul>	G	180 A · · · · · · · · · · · · · · · · · ·	A .	т • • • • • • • • •	G	G .	A	T	Γ Α.         	· A · · · · · · ·	1900 G · · · · · · · · · · · · · · · · · ·	G	G .	C (                  	5 G 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0	T .	G	C ,	22 A ( 		A	T	G	T	G	C	T	T · · · · · · · · · · · · · · · · · · ·	210 T · · ·	G .	T 1      	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert	A	G .	A G         	5 T	<ul> <li>T</li> <li>.</li> <li>.&lt;</li></ul>	G	180 A	A	т · · · · · · · · · · · · · · · · · · ·	G	G .	A .	T -	Γ Α ·	· A · · · · · · · · · · · · · ·	1900 G	G	G · · · · · · · · · · · · · · · · · · ·	C ( 	juni juni gion juni gion juni juni juni juni juni juni juni jun	T .	G	C ,	21 A ( · · · · · · · · · · · · · · · ·		A .	T	G	T	G · · · · · · · · · · · · · · · · · · ·	C	T .	T	210 T · · · · ·	G .	T 1      	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 4. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert		G .	A G       	5 T	<ul> <li>T</li> <li>.</li> <li>.</li></ul>	G · · · · · · · · · · · · · · · · · · ·	180 A · · · · · · · · · · · · · · ·	A	T · · · · · · · · · · · · · · · · · · ·	G	G .	A .	T	Г А        	· A · · · · · · · · · · · · · · · · · ·	1900 G · · · · · · · · · · · · · · · · · ·	G		C (C	3 G 3 G 3 G 3 G 3 G 3 G 4 G 4 G 4 G 4 G 4 G 4 G 4 G 4	T	G	C · ·	21 A ( · · · · · · · · · · · · · ·		A	T	G	T	G	C	T .	2 T · · · ·	210 T · · · ·	G .	T 1       	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=978 5. K113_LTR_1=9788 5. K1	A	G .	A G   	5 T	T - 1000	G · · · · · · · · · · · · · · · · · · ·	180 A	A	T · · · · · · · · · · · · · · · · · · ·	G	G .	A .	T		· A · · · · · · · · · · · · · · · · · ·	1990 G · · · · · · · · · · · · · · · · · ·	G	G .	C (C) 	jon G G - - - - - - - - - - - - -	T	G	C · ·	24 A ( · · · · · · · · · · · · · · · ·		A · · · · · · · · · · · · · · · · · · ·	T	G	T	G	C	T .	2 T · · · · · · · · · · ·	210 T · · · ·	G .	T 1          	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus	A	G .	A G	5 T	- T 	G · · · · · · · · · · · · · · · · · · ·	180 A · · · · · · · · · · · ·	A .	т	G	G .		T		· A · · · · · · · · · · · · · · · · · ·	1900 G · · · · · · · · · ·	G • • • • • • • • • • • • • • • • • • •	G	C (C ) 	juli juli juli juli juli juli juli juli	T	G · · · · · · · · · · · · · · · · · · ·				A · · · · · · · · · · · · · · · · · · ·	T	G	T	G	C	T	2 T · · · · · · · · · · · · · · · · · ·	210 T	G .		- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-9472	A	G .	A C · · · · · · ·	5 T	- T	G	180 A · · · · · · · · · · · · · · · · · ·	A	T · · · · · · · · · · · · · · · · · · ·	G	G .	A	T - · · · · · · · · · · · · · · · · · ·		<ul> <li>A</li> <li>·</li> <li>·</li></ul>	1990 G · · · · · · · · · · · · · · · · · ·	G · · · · · · · · · · · · · · · · · · ·	G .	C C C C C C C C C C C C C C C C C C C	<ul> <li>G G</li> <li>G</li> <li>G</li></ul>	T	G · · · · · · · · · · · · ·				A	T	G	T	G	C	Т .	2 T · · · · · · · · · · · · · · · · · ·	210 T · · · ·	G .		- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_8501=9472 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=964 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K113_LTR_1=968 5. K11	A	G .		5 T - - - - - - - - - - - - -	- T	G	180 A · · · · · · · · · · · · · · · · · ·	A	T · · · · · · · · · · · · · · · · · · ·	G	G .	A · · · · · · · · · · · · · · · · · · ·	T		<ul> <li>A</li> <li>·</li> <li>·</li></ul>	1900 G · · · · · · · · · · · · · · · · · ·	G	G	C ( C (	jion G C G C G C G C G C G G C G G G G G G G G G G G G G	T	G				A · · · · · · · · · · · · · · · · · · ·	T	G	T	G	C	T .	2 T · · · · · · · · · · · · · · · · · ·	C C C C C C C C C C C C C C C C C C C	G	T 1 · · · · · · ·	- A	A
Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1-968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 4. K113_LTR_1968 5. K113_LTR_8501-94722 Consensus 1. Patient_sanger_consus 2. HDR_plasmid_insert 3. luc_vector_insert 3. luc_vector_insert	A	G .	A C · · · · · · · · · · · · · · ·	5 T - - - - - - - - - - - - -	- T	G · · · · · · · · · · · · · · · · · · ·	180 A	A	T · · · · · · · · · · · · · · · · · · ·	G	G .	A	T		· A · · · · · · · · · · · · · · · · · ·	1900 G · · · · · · · · · · · · · · · · · ·	G · · · · · · · · · · · · · · · · · · ·	G .	C C C C C C C C C C C C C C C C C C C	juon Gion Gion Gion Gion A A A A A A A A A A A A A	T	G		A ( 	C A A A A A A A A A A A A A A A A A A A	A	T	G	T	G · · · · · · · · · · · · · · · · · · ·	C	T · · · · · · · · · · · · · · · · · · ·	2 T · · · · · · · · · · · · · · · · · ·	C	G .	T 1  	- A 	A

Consensus 1. Patient_sanger_consus	т ·	C	T ·	C	C	350 C	C ·	A	T ·	G	T ·	G	A	T ·	A	360 G	т (	ст 	6	i A	A	A	т ·	A	370 T	G	G	с (	ст 	C	G	т ·	G ·	380 G	G	A	A (	G (	5 G	A
2. HDR_plasmid_insert																																								
3. luc_vector_insert								•							•										•		•													
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:	:	:	:	:	:	:	:	:	•	:	•	:	:		•			U3	regi	: on	:	•	:	:	•			:	:	:	:	•	:	•	:			:
Consensus 1. Patient_sanger_consus	A	A	390 G	A ·	C ·	C ·	т ·	G	A ·	C ·	c ·	G	400 T	С	С	C	с (	C A 	· G	5 C	С	410 C	G ·	A	C ·	A	с	с (	G G	т	420 A	A	А	G ·	G ·	G ·	т (	с т 	Г <del>С</del>	430 T
2. HDR_plasmid_insert	·	•	·	·	•	•	•	•	•	·	·	·	·	•	•	•	•			•	·	•	•	•	•	•	•	•			·	·	·	•	·	·	•		•	•
3. luc_vector_insert	·	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			·	·	•	•	•	•	•	•	•		•		•	·	•	•	•	•		•	·
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				U3	regi	on	:	:	:	:	:	:		:	:	:	:	:	:	•	:			
Consensus 1. Patient_sanger_consus	G	c	т ·	G	A	G	G ·	A	G	440 G	A	T ·	T ·	A	G ·	т. •	A ۱	N A A ·	45 A	50 G	A	G	G	A	A	G ·	G	4 A 7	60 АТ	G	c	C	т ·	C ·	т ·	T ·	4 G (	70 C /	A G	т ·
2. HDR_plasmid_insert																	· ,	Ą.																						
3. luc_vector_insert								•		•								г・							•		•								•					
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:	:	•	•	•	•	:	•	:	:	:	•	:	•	•	: :	T ·		U3	regi	on	•			:	:	:		•			•	•	:	:	:			
Consensus 1. Patient_sanger_consus	т ·	G	A	G	A	C ·	480 A •	A	G ·	A	G	G ·	A ·	A	G ·	G ·	190 C	ат 	- c	: т ,	G	т ·	C ·	т ·	C ·	500 C	T ·	S ( G	с с 	T ·	G	т ·	C ·	C ·	510 Ċ	T ·	G (	G (	5 C	А
2. HDR_plasmid_insert	·		•	•				•			·		•	•	•	•	•			•	·	•			•	•	•	G					•			·	•			·
3. luc_vector_insert	·		·	·	·		·	·	·	·	·	•	·	•	•	•	•						·		·	•	·	С				·		·	·	·	•		•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472			:	:	:	:	:	:	:	:	:	:	:	:	:	:				U3	regi	on	:	:	:	:	:	C C		:		:	:	:	:	:	:			:
Consensus 1. Patient_sanger_consus	A	т ·	G	520 G	A	A	т ·	G	T ·	C ·	T ·	с	G ·	530 G	T ·	A .	T,	A A	A A	A A	C ·	C ·	540 C	G	A	T ·	T ·	G ·	ГА 	T ·	G	550 C	т ·	C ·	C ·	A	т (	ר כ י	Г А 	C ·
2. HDR_plasmid_insert	•	•	•				•	•	•	•	•		•	•	·	•									•	•	•	•					·	•	•	•	•	• •	•	
3. luc_vector_insert			•	•	•			•			•		•													•	•	•								•	•		•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:	:	:	:		:	:	:	:	:	:	:	:	:	:				U3	regi	on	:		:	:	:	:			•	:	:	:	:	:	:			:
Consensus 1. Patient_sanger_consus	560 T	G	A	G	A	т ·	A	G	G	G ·	570 A	A	A	A	A	C .	C I	G (	: c	58 T	γ Τ	A	G	G	G	C ·	T ·	G (	59 G A	G G	G	т ·	G	G	G ·	A	с (	6  C 7	ро Г <mark>С</mark>	C
2. HDR_plasmid_insert		•	•															• •			т					•	•								•					•
3. luc_vector_insert								•		•	·		•		·	•				•	С				•		•	·							•	·	•			
4. K113_LTR_1-968 5. K113_LTR_8501-9472			:	:	:	•	•		:	:	•	:	:	:	•		•				C C R reg	jion	•	•			•			•	•		•	•	:	•	:			
Consensus 1. Patient_sanger_consus	G	G	G	C	А	G ·	C ·	610 A	A ·	т ·	A	с	T ·	G ·	с	т ·	б Т •	70 T C	; т	A	A	A ·	G	C ·	A	Y C	630 T	G,	A G	А	T ·	G	т ·	т ·	т ·	640 A	т (	G 7	F G	т ·
2. HDR_plasmid_insert	•	•					·	•	•	•	•		•	•		•	•			•					•	С	•	•					·	•	•	•	•	• •	• •	
3. luc_vector_insert			•					•			•		•	•	•										•	т	•									•	•			
4. K113_LTR_1-968 5. K113_LTR_8501-9472			:	:	:		:		:		:	:	•	:	:					R	regic	on	:	:		C T	•							•		:	:			
Consensus 1. Patient_sanger_consus	A	т ·	G	C	650 A	т	A	т	C ·	т ·	A	A	A	A (	60 G	с.	A (	C A 	• G	i C	A	C	т ·	670 T	A	A	T ·	с (	ст.	T ·	T ·	A	680 C	A	т ·	T ·	G	т <b>с</b>	ст.	A
2. HDR_plasmid_insert																																								
3. luc_vector_insert																																								
4. K113_LTR_1-968 5. K113_LTR_8501-9472		:	:	•	•	•	•		:	•	:	:	•	•	•					R	regio	n	•				•	:			•				•	:	•			:

Consensus 1. Patient_sanger_consu:	т 5 ·	690 G	A ·	T	G	C	C ·	A	A	G	A	700 C	с •	T ·	T ·	т ·	G	т ·	T ·	C ·	A	710 C	G	T ·	G ·	T ·	т ·	T ·	G	T ·	c -	20 T (	5 C	: т	G	A	C	C ·	C ·	т -	<sup>730</sup> СТ
2. HDR_plasmid_insert			·			F - 2	2 -lik	ke •																						•											
3. luc_vector_insert	•												•							•									•	•						•			•	•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472	2 .	:		:	:	:	:	:	:	:	:		:	:	:		:	:		•	Rr	egio	n	:	:	:	:	:	:	:					:		:	:	:	:	
Consensus 1. Patient_sanger_consus	<b>C</b>	C ·	c ·	C	A	С	A	A	740 T	с т	G	т ·	C ·	т ·	т ·	G	т ·	G	750 A	c ·	c	C ·	т ·	G	A	C ·	A	C ·	760 A	T ·	с (	C (		: c	T ·	c	т ·	770 T	т ·	G.	AG
2. HDR_plasmid_insert	•																			•				Re	ta di	ohir	– lik	e nr		ter						•					
3. luc_vector_insert	•		•						•	•	•	•					•			•	•	•	·	•			•			•		•				•			•	•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472	<u>,</u>	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		:	:	:	Rr	egio	n	:		:	:	:	:	:				:	:	:	:	:	:	•	
Consensus 1. Patient_sanger_consus	A 5 ·	A	A	C ·	A	780 C	° с	С	A	c	R G	G	А	т ·	G	790 A •	т ·	с	A	A	T TA I	A	A like	A	т	800 A •	C ·	т ·	A	A	G (	G (	5 A	A ·	81 C	от	C	A ır	G	A ·	GG.
2. HDR_plasmid_insert	•		·		•		•		•	•	G	·	·	•	•	•	•	•	·	•	·	•	•	•	•	•	•	·	·	•	•	•		•	•	•	•	·	•	•	
3. luc_vector_insert	•	•	·	•	•	•	•	·	•	·	A	·	·	·	•	•	•	·	·	•	·	·	·	•	•	·	·	·	•	•	·	•	• •	•	·	•	·	·	·	•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472	2 .		:				:		:		A A	:	:		:	•	:	:	Rr	egio	n	:	:	•	:	:	:	:	:					:	:				U5 i	egio	: : n
Consensus 1. Patient_sanger_consus	<b>C</b>	т ·	820 G	G G	C	G <4	G	G	A	T ·	C ·	C ·	830 T	C ·	С	A ·	т ·	A ·	т ·	G	C ·	т ·	840 G	A ·	A ·	C ·	G	C ·	T ·	G	G	т <sup>-</sup>	50 F C	: C	С	C	G	G	K G	т	C C
2. HDR_plasmid_insert	•	•	·			·		•	•	·	·	·	·	·	·	•	•	•	·	•	·	•	·	•	•	•	•	·	•	•	·	•		•	·	·	·	·	G	•	
3. luc_vector_insert	•		•	•					•	•	•	•	·		•		•		•	•	•	•	•					·	•	•	•	•				•	•	•	т	•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472	<u>-</u> 		:	:	:	•	:	:	:	:	:	:	:	:	:	:	:	:	:	:	U5 I	egio	on	:	:	:	:	:	:	:	:			:	:	:	:	:	T T	:	
Consensus 1. Patient_sanger_consu:	<b>C</b>	c ·	т ·	т ·	A	т ·	T ·	T ·	С	870 T	) т	т ·	C ·	т ·	C ·	т ·	A ·	т ·	A	880 C	т ·	т ·	т ·	G ·	т ·	C ·	т ·	C ·	т ·	390 G	т (	G -	Г <b>(</b>	: т	T ·	т ·	т ·	т ·	900 C	т .	тт 
2. HDR_plasmid_insert	•								•	•	•		•	•			•	•	•	•	•	•	•	•	•	•	•	·	•	•	•	•		•				•	•		
3. luc_vector_insert	•		•	•	•	•	•		•	•	•	•	•	·	•	•	•	•	•	·	·	•	•	•	•	•	·	·	•	•	•	•		•	•	•	•	·	•	•	• •
4. K113_LTR_1-968 5. K113_LTR_8501-9472	<u>2</u> :		•	:	:	:	:	•	:	:		•	:	:	:	:	:	:	•	:	U5 (	egio	: on	:	:	:	:	•	:	:	:			:	:		•	•	:	:	: :
Consensus 1. Patient_sanger_consus	т 5 ·	C ·	C ·	A	A	A	910 T	° C	т ·	C ·	T ·	C ·	G	T ·	C ·	C ·	920 C	A	C ·	c ·	т ·	т ·	A	C ·	G ·	A	930 G	A	A	A	с,	4 (	с с	: C	A	940 Ċ	A	G	G	т ·	G Т · ·
2. HDR_plasmid_insert	•						•		•		•			·						•	•	•	•					·	•	•	•			•				·	•		
3. luc_vector_insert	•		•			•	•	•	•	·	·	•	•	·	•	•	•	•	·	•	·	•	•	•		•	•	·	•	•	•			•		•	•	•	•	•	
4. K113_LTR_1-968 5. K113_LTR_8501-9472	2 · ·	:	:	:	:	•	:	:	:	:	:	:	:	:	:	:	:	:	:	:	U5 1	egio	: on	:	:	:	:	:	:	:	:			:	:	:	:	:	:	:	
Consensus 1. Patient_sanger_consus	G 5 ·	T ·	A	950 G	G	G	G	C	A	A	C	C	C	960 A	C ·	C ·	C ·	C	т ·	A	c	968 A																			
2. HDR_plasmid_insert																				•		•																			
3. luc_vector_insert	•					•			•				·		•	•	•			•	•																				
4. K113_LTR_1-968 5. K113_LTR_8501-9472	<u>.</u>	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																			
	<u> </u>									(	<u>J5 r</u> €	gloi	1																												
238																																									
229																																									
239 240 241																																									

Supplementary Figure 2: Nucleotide alignment of 16 sanger-sequenced
RASGRF2-int LTR positive patients, 4 from each group: 1-4: GR PWID
(Greek Persons Who Inject Drugs),5-8: GR non-PWID (Greek non-PWID), 912: UK PWID (British Persons Who Inject Drugs),13-16: UK non-PWID. All
samples were found identical.

Consensus	1 т <u>бтббб</u> б	10 SAAAAGCAAG	20 AGAGATCA	30 GATTGTTA	40 TGTGTCTC	TGTAGAAA	0 GAAGTAGA	60 CATAGGAGAG	70 CTCCATTTT	80 GTTATGTA	90 CTAAGAAAAA	100 ATTCTTCTGC
1. 143486 2. 141267 3. 137685			· · · · · · · · ·		· · · · · · · · · · ·				 			
4.137512 5.99583 6.98523	· · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·		· · · · · · · · · · · · ·	· · · · · · · · · · · · · · ·
7.87014 8.79862 9.pl1f2			· · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · ·	· · · · · · · · · ·		· · · · · · · · · · · ·			
10. pl1e10 11. pl1e1 12. pl1d7							· · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·			
14. pl2g5 15. pl2f2 16. pl2e9												
Consensus	CTTGAGA	110 TTCTGTTAA	120 Т <b>СТАТААС</b>	130 CTTACCCC	140 С <mark>ААСССТБ</mark> Т	150 GCTCTCTG	α <mark>ΑΑ</mark> ϹΑΤGΙ	160 F <mark>GCTGTGTCC</mark>	170 ACTCAGAGT	180 TGAATGGA	190 TTAAGGGCGC	200 GTGCAGGATG
1.143486												
2.141267 3.137685												· · · · · · · · · · · ·
4.137512 5.99583			· · · · · · · · ·				 					· · · · · · · · · · · · ·
6.98523 7.87014							 		 			
8.79862 9.pl1f2												
10. pl1e10						• • • • • • • • •						
12. pl1d7												
13. pl2g11 14. pl2g5			•••••				· · · · · · · · · ·					
15. pl2f2 16. pl2e9			· · · · · · · · ·			· · · · · · · · · · ·	 		 			
	210	220	)	230	240	250	260	270	1	280	290	300
Lonsensus			IGCTIGAA	GGCAGCAT	CICCIIA#	GAGICAIC			461ACCCA6	GGACACAA		AGGCCGCAG
2.141267												
4.137512												
6.98523												
7.87014 8.79862			· · · · · · · · ·			· · · · · · · · · · ·	 		 			· · · · · · · · · · · · ·
9. pl1f2 10. pl1e10												
11. pl1e1												
13. pl2g11												
15. pl2f2						• • • • • • • • •						
16. pize9	310	320	33	0	340	350	360	370	380	3	390	400
Consensus	GGACCTC	TGCCTAGGA	AAGCCAGC	STATTGTCC/	AGGTTTCT	CCCCATGT	ЗАТА <u>Б</u> ТСТ	GAAATATGG	сстсбтббб	GAAGGGAAA	GACCTGACCO	STCCCCCAGC
2.141267						• • • • • • • • •						
4.137512												
5.99583 6.98523												
7.87014 8.79862			•••••				 					
9. pl1f2 10. pl1e10			· · · · · · · · ·				 		 			· · · · · · · · · · · · ·
11. pl1e1 12. pl1d7							 		 			
13. pl2g11												
15. pl2f2												
10. pizes	410	420	430	440		450	460	470	480	490	500	510
Consensus	CCGACAC	CCGTAAAGG	бтстбтбс	TGAGGAGG	TTAGTAAA	AGAGGAAG	GAATGCCI	CTTGCAGTTO	GAGACAAGA	GGAAGGCA	тствтстсст	гөсстөтссс
1.143486 2.141267			· · · · · · · · ·				 					· · · · · · · · · · · · ·
3.137685 4.137512							 		 			
5.99583												
7.87014												
6.79862 9.pl1f2												
11. pl1e10												
12. pl1d7 13. pl2g11		· · · · · · · · · · · ·	· · · · · · · · ·			· · · · · · · · · ·	 					
14. pl2g5 15. pl2f2												
248 16. pl2e9												

Consensus	520 TGGGCAATGGAAT	530 GTCTCGGTATA	540 AAACCCGATTO	550 TATGCTCCAT	560 CTACTGAGA1	570 AGGGAAAAAC	580 CGCCTTAGGO	590 CTGGAGGTGG	600 GACCTGCGG	610 GCAGCAAT
1.143486 2.141267 3.137685 4.137512 5.99583 6.98523	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	
7.87014 8.79862 9.pl1f2 10.pl1e10 11.pl1e1 12.pl1d7 13.pl2g11	· · · · · · · · · · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	
14. pl2g5 15. pl2f2 16. pl2e9	6 <u>2</u> 0	 	 					 	700	  710
Consensus	ACTGCTTTGTAAA	.GCACTGAGATG	TTTATGTGTAT	GCATATCTAA	AAGCACAGCA	ΑCTTAATCCTT	ТАСАТТСТСТ	ATGATGCCAA	.GACCTTTGT	TCACGTGT
1. 143486 2. 141267 3. 137685 4. 137512 5. 99583 6. 98523 7. 87014 8. 79862 9. pl1f2 10. pl1e10 11. pl1e1 12. pl1d7 13. pl2g11 14. pl2g5 15. pl2f2 16. pl2f2										
	720	730	740	750 7	60	770 7	780	790 8	300	810
1. 143486 2. 141267 3. 137685 4. 137512 5. 99583 6. 98523										
7.87014 8.79862 9. pl1f2 10. pl1e10 11. pl1e1 12. pl1d7 13. pl2g1 14. pl2g5 15. pl2f2 16. pl2e9										
Consensus	820 8 GCTGGCGGGATCC	30 840 TCCATATGCTG	) 850 AACGCTGGTTC	860 CCCGGGTCCC	870 CTTATTTCTI	880 TCTCTATACT	890 TTGTCTCTGT	900 GTCTTTTTCT	91 TTT <mark>CCAAA</mark> T	0 CTCTCGTC
1. 143486 2. 141267 3. 137685 4. 137512 5. 99583 6. 98523 7. 87014 8. 79862 9. pl1f2 10. pl1e10 11. pl1e1 12. pl1d7 13. pl2g11 14. pl2g5 15. pl2f2 16. pl2e9	920 930	940	950	960	968					
Consensus	CCACCTTACGAGA	AACACCCACAG	GTGTGTAGGGG	CAACCCACCC	СТАСА					
1. 143486 2. 141267 3. 137685 4. 137512 5. 99583 6. 98523 7. 87014 8. 79862 9. pl1f2 10. pl1e10 11. pl1e1 12. pl1d7 13. pl2g1 14. pl2g5 15. pl2f2					· · · · · · · · · · · · · · · · · · ·					
250 <sup>16. pl2e9</sup> 251										

- 252 **Supplementary Figure 3:** Plasmid map of (A) the HDR vector used for the
- transfection of HEK293s and the first attempt of eHAP1s and (B) the
- 254 minicircle construct used for the second transfection attempt of eHAP1s.



Supplementary Figure 4: CRISP\Cas9 editing of eHAP1 cells using (A) the
 same HDR vector as for the HEK293s and (B) the minicircle sub-construct

#### (A) RASGRF2 HERVK Integration (KI) PCR (KAPA2G 4kb) – 1st round

Conditions: 1 ul gDNA per 25 ul PCR – 1.25 ul of each 10 uM primer Performed two different PCRs per sample



Potentially positive clones A7 – A8 – B4 – B6 – B11 – C10 were found negative upon sequential PCR in expanded population.

#### (B) RASGRF2 HERVK Integration (KI) PCR (KAPA2G 4kb) – 2<sup>nd</sup> round (minicircle)

Conditions: 1 ul gDNA (bulk of targeted cells) per 25 ul PCR – 1.25 ul of each 10 uM primer Tm = 60 deg



○ ● ● = expected band size upon successful targeting event

### 266 **References**

- Subramanian RP, Wildschutte JH, Russo C, & Coffin JM (2011)
   Identification, characterization, and comparative genomic distribution of
   the HERV-K (HML-2) group of human endogenous retroviruses.
   *Retrovirology* 8.
- 272 2. Fuchs NV, *et al.* (2011) Expression of the human endogenous retrovirus
  273 (HERV) group HML-2/HERV-K does not depend on canonical promoter
  274 elements but is regulated by transcription factors Sp1 and Sp3. *J Virol*275 85(7):3436-3448.
- Kovalskaya E, Buzdin A, Gogvadze E, Vinogradova T, & Sverdlov E (2006)
   Functional human endogenous retroviral LTR transcription start sites are
   located between the R and U5 regions. *Virology* 346(2):373-378.
- Knossl M, Lower R, & Lower J (1999) Expression of the human
  endogenous retrovirus HTDV/HERV-K is enhanced by cellular
  transcription factor YY1. *J Virol* 73(2):1254-1261.
- 282