



**Figure S1.** *tat* gene and functional domains of Tat protein. Tat is encoded by two exons. The first five domains are encoded by the first exon. I - first domain, proline-rich region: in 11 position - conserved tryptophan residue (W); II - second domain, cysteine-rich region: 7 highly conserved cysteine residues – in 22, 25, 27, 30, 31, 34 and 37 positions; III - third domain: contains hydrophobic core sequence <sup>43</sup>LGISYG<sup>48</sup>; IV - fourth domain, arginine-rich region, binds to the TAR element of HIV RNA; V - fifth domain, glutamine-rich region. The sixth domain is encoded by the second exon: contains <sup>78</sup>RGB<sup>80</sup> and <sup>86</sup>ESKKKVE<sup>92</sup> functionally significant motifs.

**Table S1.** The amino acid positions in Tat1-A6 with a statistically significant difference in the frequency of conservation between the sequences in the group of virus variants circulating in the Moscow Region and in the reference group (the sequences downloaded from the Los Alamos database).

Amino acid*	Reference** group N=141 (100%)	Moscow Region N=250 (100%)	<i>p</i>
2D	140 (99.3%)	225 (90%)	<b>0.0004</b>
7N	132 (93.6%)	216 (86.4%)	0.0285
21A	140 (99.3%)	238 (95.2%)	0.0303
23S	121 (85.8%)	181 (72.4%)	<b>0.0024</b>
31C	129 (91.5%)	197 (78.8%)	<b>0.0012</b>
47Y	141 (100%)	239 (95.6%)	0.0115
54H	89 (63.1%)	108 (43.2%)	<b>0.0002</b>
57G	130 (92.2%)	202 (80.8%)	<b>0.0025</b>
61S	119 (84.4%)	172 (68.8%)	<b>0.0007</b>
62S	121 (85.8%)	167 (66.8%)	<b>&lt;0.0001</b>
68P	84 (59.6%)	118 (47.2%)	0.0187

\*- the amino acid in the reference sequence. Consensus sequence in reference group was used as a reference sequence; \*\* – the group of Tat1-A6 sequences downloaded from the HIV-1 Los Alamos Database ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)). In bold: significant in the  $\chi^2$  test after Bonferroni correction  $p < 0.0045$

**Table S2.** Natural polymorphisms into Tat2 functionally significant motifs, <sup>78</sup>RGD<sup>80</sup> and <sup>86</sup>ESKKKVE<sup>92</sup>, in different clades\*.

Domain	Substitution	A6 N=205	A1 N=236	B N=25 0	C N=250	<i>p</i> A6-A1	<i>p</i> A6-B	<i>p</i> A6-C
RGD	R78Q	186	190	8	50	<b>0.0025</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>
	R78P	3	22	2	5	<b>0.0004</b>	-	-
	R78H	1	8	4	0	0.0316	-	-
	R78G	0	0	23	0	-	<b>&lt;0.0001</b>	-
	G79R	194	38	4	4	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>
	D80N	45	25	16	74	<b>0.0011</b>	<b>&lt;0.0001</b>	-
	D80E	6	3	0	0	-	0.0065	0.0065
	D80V	1	94	0	2	<b>&lt;0.0001</b>	-	-
	D80I	0	57	0	2	<b>&lt;0.0001</b>	-	-
ESKKKVE	E86K	15	10	16	5	-	-	<b>0.0059</b>
	S87P	7	10	35	6	-	<b>0.0001</b>	-
	S87Q	0	8	31	3	0.0078	<b>&lt;0.0001</b>	-
	K88Q	0	0	5	1	-	0.0418	-
	K88T	30	8	3	1	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>
	K88A	5	2	0	0	-	0.0130	0.0130
	K88E	16	7	17	1	0.0226	-	<b>&lt;0.0001</b>
	K89E	4	13	21	1	-	<b>0.0027</b>	-
	K90E	5	17	18	13	0.0219	0.0211	-
	K90T	1	1	11	1	-	0.0096	-
	V91M	9	30	2	2	<b>0.0021</b>	0.0131	0.0131

\***Note.** Amino acid substitutions were defined as the substitutions in the indicated positions in comparison with the consensus sequence of clade B. *p* values are given for the positions with *p* < 0.05, the positions with *p* ≥ 0.05 are marked “-”. In bold: significant in the  $\chi^2$  test with Bonferroni correction: A6-A1 *p* < 0.0042; A6-B *p* < 0.0036; A6-C *p* < 0.0062

**Table S3.** The amino acid positions in Tat2-A6 with statistically significant differences in the frequency of conservation between the sequences in the group of virus variants circulating in the Moscow Region and in the reference group (the sequences downloaded from the Los Alamos database).

<b>Amino acid*</b>	<b>Reference group** N=205 (100%)</b>	<b>Moscow Region N=189 (100%)</b>	<b><i>p</i></b>
73P	187 (91.2%)	186 (98.4%)	<b>0.0015</b>
86E	184 (89.8%)	180 (95.2%)	0.0404
97T	157 (76.6%)	128 (67.7%)	0.0495
101D	157 (76.6%)	160 (84.6%)	0.0436

\*- the amino acid in the reference sequence. Consensus sequence in reference group was used as a reference sequence; \*\* – the group of Tat2-A6 sequences downloaded from the HIV-1 Los Alamos Database ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)). In bold: significant in the  $\chi^2$  test with Bonferroni correction  $p < 0.0125$