## 1 Title: The role of Late Presenters on HIV-1 transmission clusters in Europe

### 3 Running title: HIV-1 TC in LP in Europe

4

Mafalda N. S. Miranda<sup>1\*</sup>, Marta Pingarilho<sup>1</sup>, Victor Pimentel<sup>1</sup>, Perpétua Gomes<sup>2,3</sup>,
Maria do Rosário O. Martins<sup>1</sup>, Sofia G. Seabra<sup>1</sup>, Rolf Kaiser<sup>4,5</sup>, Michael Böhm<sup>4,5</sup>,
Carole Seguin-Devaux<sup>6</sup>, Roger Paredes<sup>7</sup>, Marina Bobkova<sup>8</sup>, Maurizio Zazzi<sup>9</sup>, Francesca
Incardona<sup>10,11</sup> and Ana B. Abecasis<sup>1</sup>

9

### 10 Affiliations:

- 1- Global Health and Tropical Medicine (GHTM), Institute of Hygiene and 11 12 Tropical Medicine, New University of Lisbon (IHMT/UNL), 1349-008 Lisbon, 13 Portugal; martapingarilho@ihmt.unl.pt (M.P.); victor.pimentel@ihmt.unl.pt (V.P.); mrfom@ihmt.unl.pt (M.d.R.O.M.); sgseabra@ihmt.unl.pt (S.G.S); 14 15 ana.abecasis@ihmt.unl.pt (A.A.) 16 2- Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, 17 Portugal; pcrsilva@chlo.min-saude.pt (P.G) 18 3- Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Instituto 19 Universitário Egas Moniz, Costa da Caparica, Portugal. 20 4- Institute of Virology, Faculty of Medicine and University Hospital of Cologne, 21 University of Cologne, Germany; rolf.Kaiser@uk-22 koeln.de (R.K.); michael.boehm@uk-koeln.de (M.B.) 23 5- DZIF Deutsches Zentrum für Infektionsforschung 24 6- Laboratory of Retrovirology, Department of Infection and Immunity, 25 Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg; 26 carole.devaux@lih.lu (C.S.-D.) 27 7- Infectious Diseases Department and IrsiCaixa AIDS Research Institute, Hospital 28 Universitari Germans Trias i Pujol, Badalona, Spain; rparedes@irsicaixa.es 29 (R.P.) 30 8- Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia; mrbobkova@mail.ru (M.B.) 31 32 9- Department of Medical Biotechnologies, University of Siena, Siena, Italy; 33 maurizio.zazzi@unisi.it (M.Z.) 34 10- IPRO—InformaPRO S.r.l., Rome, Italy; f.incardona@informa.pro (F.I.) 35 11- EuResist Network, Rome, Italy 36 37 **Correspondence:** 38 **Corresponding Author**
- 39 Mafalda N.S. Miranda (MNSM)
- 40 mafalda\_nsm@hotmail.com
- 41

42 Keywords: HIV-1 infection, Late presenters, Non-Late Presenters, Transmission 43 clusters

- 44
- 45 Abstract

46 Background: Investigating the role of late presenters (LP) on HIV-1 transmission is 47 important, as they can contribute to the onward spread of HIV-1 virus in the long period 48 before diagnosis, when they are not aware of their HIV status

49 **Objective:** To describe the clinical and socio-demographic characteristics of HIV-1 50 infected individuals followed in Europe, to characterize patients in clusters and to 51 compare transmission clusters (TC) in LP vs non-late presenters (NLP) populations.

52 Methods: Clinical, socio-demographic and genotypic information from 38531 HIV-1

53 infected patients was collected from the EuResist Integrated Database (EIDB) between 54 1981 and 2019. Sequences were aligned using VIRULIGN. Maximum likelihood (ML) 55 phylogenies were constructed using FastTree. Putative transmission clusters were 56 identified using ClusterPicker v1.332. Statistical analyses were performed using R.

57 **Results:** 32652 (84.7%) sequences were from subtype B, 3603 (9.4%) were from 58 subtype G, and 2276 (5.9%) were from subtype A. The median age was 33 (IQR: 26.0-59 (41.0) years old and (75.5%) of patients were males. The main transmission route was 60 through homosexual (MSM) contact (36.9%) and 86.4% were originated from Western

61 Europe. Most patients were native (84.2%), 59.6% had a chronic infection, and 73.4% 62 had acquired drug resistance (ADR). CD4 count and viral load at diagnosis (log10) 63 presented a median of 341 cells/mm3, and of log10 4·3 copies/mL, respectively. 51·4%

64 of patients were classified as LP and 21.6% patients were inside TCs. Most patients

from subtype B (85.6%) were in clusters, compared to subtypes A (5.2%) and G (9.2%). 65

66 Phylogenetic analyses showed consistent clustering of MSM individuals. In subtype A,

67 patients in TCs were more frequently MSM patients and with a recent infection. For 68 subtype B, patients in TCs were more frequently those with older age (≥ 56), MSM 69 transmission route, originating from Western Europe, migrants, and with a recent 70 infection. For subtype G, patients in TC were more frequently patients with recent 71 infection and migrants. When analysing cluster size, we found that NLP more 72

frequently belonged to large clusters (>8 patients) when compared to LP.

Conclusion: While late presentation is still a threat to HIV-1 transmission, LP 73 74 individuals are more present either outside or in small clusters, indicating a limited role

75 of late presentation to HIV-1 transmission.

### 76 Introduction

77 At the end of 2020, there were 37.7 million people living with HIV (1) and it is known 78 that in HIV epidemics, certain risk groups contribute to the spread of HIV 79 disproportionately more than others. This can be due to specific demographic, clinical 80 or behavioral factors or to factors related to the infecting strain of the virus (2,3). On the one hand, the literature suggests that a recent infection, without diagnosis, could be 81 82 associated with transmission and spread of HIV-1 (4). On the other hand, late 83 presentation to diagnosis has been increasing over the years and in Europe, late 84 presenters (LP) account for around 50% of HIV new diagnosis (5). Late presenters (LP), 85 based on a definition consensus, are HIV-1 infected individuals defined by a baseline 86 CD4 count lower than 350 cells/mm3 or with an AIDS-defining event, regardless of 87 CD4 cell count (6). Late presentation is associated with high morbidity and mortality, at 88 an individual level, and increased health costs (7). Besides those consequences, LP can 89 also contribute to the onward spread of HIV-1 virus at the population level, as these 90 individuals are not aware of their HIV status and could also spread the virus without 91 knowing their infection status (8).

92 The use of powerful tools as phylogenetic trees and transmission clusters (TC) are 93 essential to understand the dynamics of viral transmission and to identify groups of 94 individuals connected to each other (2,9).

95 In this study, we aim to describe the clinical and socio-demographic characteristics of

96 HIV-1 infected individuals followed in Europe according to subtype and to understand

97 the determinants associated with clustering on each of the more prevalent subtypes. The

analysis of transmission clusters for the most prevalent HIV-1 subtypes, B, A, and G,

and comparison of the patterns of transmission clusters in late presenters (LP) vs non-

100 late presenters (NLP) populations were included in this study.

101

### 102 Methods

### 103 Study Group

104 Clinical, socio-demographic and genomic information from 38531 HIV-1 infected 105 patients from the EuResist Integrated Database (EIDB) between 1981 and 2019 were 106 included in this study. The EuResist integrated database (EIDB) is one of the largest 107 existing datasets which integrate clinical, socio-demographic and viral genotypic 108 information from HIV-1 patients. It integrates longitudinal, periodically updated data 109 mainly from Italy (ARCA database), Germany (AREVIR database) Spain (CoRIS and 110 IRSICAIXA), Sweden, Belgium, Portugal, and Luxembourg databases (10).

In this study, information from the ARCA, AREVIR, Luxembourg, IRSICAIXA,Portugal and Russia databases were used.

113

### 114 Institutional Review Board Statement

All procedures performed in this study were in accordance with the ethical standards of the ethical committee of the Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa and with the 1964 Helsinki declaration. The database enrolled anonymized patients' information, including demographic, clinical and genomic data from patients from the EuResist Integrated Database (Date of approval: 15 January 2021).

121

### 122 Drug Resistance Analysis and Subtyping

HIV *pol* sequences were derived from existing routine clinical genotypic resistance tests
(Sanger method). The size of reverse transcriptase (RT) and protease (PR) fragments
used for this analysis was between 500 and 1000 nucleotides. Only the first HIV

126 genomic sequence per patient was analyzed. Transmitted Drug Resistance (TDR) was 127 defined as the presence of one or more surveillance drug resistance mutations in a 128 sequence, according to the WHO 2009 surveillance list (11). The sequences were 129 submitted to the Calibrated Population Resistance tool version 8.0. Clinical resistance to 130 ARV drugs was calculated through the Standford HIVdb version 9.0.

131 HIV-1 subtyping was performed using the consensus of the result obtained based on 132 three different subtyping tools: REGA HIV Subtyping Tool version 3.46 (https://www.genomedetective.com/app/typingtool/hiv), COMET: adaptive context-133 134 HIV-1 (https://comet.lih.lu) based modeling for and SCUEAL 135 ("http://classic.datamonkey.org/dataupload\_scueal.php").

136 137

### 138 Transmission cluster (TC) identification

For the analysis of transmission clusters and construction of phylogenetic trees, the 139 140 database was divided in three separate datasets, subtype B, A, and G. Control sequences were retrieved from the Los Alamos database and all HIV-1 pol sequences from 141 142 subtypes B, A, and G from Europe, South America and Africa were included 143 (http://www.hiv.lanl.gov) (12). We used as the outgroup reference three subtype B and 144 C references retrieved from the Los Alamos database. For each subtype, sequences were 145 aligned against the control sequences dataset using VIRULIGN (13). The HIV-1 146 K03455.1 (HXB2) pol nucleotide sequence (nt) was used as reference for codon 147 corrected alignment. The dataset was then manually edited to exclude sequences with 148 low quality, duplicates and clones using MEGA7 software. The final datasets of 149 subtypes B, A, and G consisted of 62543, 10122, and 5547 sequences, respectively, 150 with a length of 948. Maximum likelihood (ML) phylogenies were constructed in FastTree with the generalized time reversible model. Statistical support of clades was 151 152 assessed using the Shimodaira-Hasegawa-like test (SH-test). Putative transmission 153 clusters were identified using ClusterPicker v1.332 (14) and defined a threshold that 154 included a genetic distance of 0.030 and a branch support  $\geq 0.90$  aLRT. For analyses of 155 cluster size, we defined clusters with 8 patients or more as large clusters and cluster with less than 8 patients as small clusters. The visual configuration of the phylogenetic 156 157 trees was possible through the iTOL programme.

158

### 159 Study Variables

160 New variables were created according to:

- Migration Status- Based on Country of Origin and Country of Follow-up (if
   country of origin and country of follow-up is the same, then patient was
   classified as native; otherwise as migrant)
- Age at Resistance Test Based on the difference between Year of Birth and
   Date of the first drug resistance test;
- Region of Origin- Based on Country of Origin;
- Treatment Status at date of first Drug Resistance Test based on the difference
   between sample collection date for first drug resistance test and date of start of
   first therapy:
- ART-naïve→ patients who had a sample collection date for first drug resistance
   test before the date of start of first therapy
- 172ART-experienced  $\rightarrow$  patients who had a sample collection date for first drug173resistance test after the date of start of first therapy

# Recentness of infection - Based on ambiguity rate of genomic sequences. We defined Chronic Infection if the ambiguity rate was higher than 0.45% otherwise Recent infection was defined, as previously described (15).

- 177
- 178 179

• LP vs NLP at HIV diagnosis- Based on CD4 count, LP were defined as patients with a baseline CD4 count =< 350 cells/mm3 and NLP were defined as patients with baseline CD4 count > 350 cells/mm3 (6).

180

### 181 Statistical analysis

The proportion and median (interguartile range, IOR) were calculated for every 182 categorical and continuous variable, respectively. The treatment status variable was 183 compared with the categorical variables with Chi-square test, and continuous variables 184 with Mann-Whitney U test. Logistic regression was used to analyze the association 185 186 between demographic, clinical factors, clustering status and the subtypes. First, we 187 presented the logistic regression with the unadjusted odds ratios (uOR) and confidence intervals at 95% (95% CI), then we included only the variables with a p-value <0.05 in 188 189 the final model. The final model was adjusted for sex, this variable was forced into the 190 model regardless of its significance. Data was analyzed using RStudio (Version 191 1.2.5033).

192

### 193 **Results**

### 194 Characteristics of European Population

195 Among the 38531 HIV-1 infected patients from the EIDB included in the analysis, 196 32652 (84.7%) were from subtype B, 3603 (9.4%) were from subtype G and 2276 197 (5.9%) were from subtype A. The median age at resistance test was 40.0 (34.0-47.0) years old and 75.5% of HIV-1 infected patients were males. The main transmission 198 199 route was through homosexual (men who have sex with men- MSM) contact (36.9%). 200 For subtype B, the MSM route (40.7%) was also the most prevalent route whereas was 201 the heterosexual route was the predominant route for subtype A and G (43.3% and 202 49.9%, respectively) (Data not shown). 86.4% of patients were originated from Western 203 Europe and according to subtypes, South America was the most prevalent region of 204 origin for subtype B, whereas subtype A-infected patients were mainly from Eastern 205 Europe region and for G it was Africa region the most prevalent. Most patients included 206 in this study were native (84.2%) and according to subtype, in subtype B natives were 207 more prevalent (94.8%) and in subtype A and G migrants were more prevalent (13.8% 208 and 16.6%). Based on the ambiguity rate of the first genomic sequence, most patients 209 were classified as presenting a chronic infection (59.6%). Most patients were ART-210 experienced (59%) and 73.4% had acquired drug resistance (ADR). CD4 count at 211 diagnosis and viral load at diagnosis (log10) presented a median of 341 cells/mm3 (IQR 212 170-540) and log10 4.3 copies/mL (IQR 3.4-5.0), respectively. 21.6% of patients were 213 represented within transmission clusters and 51.4% of patients were classified as LP 214 (CD4<350 cells/mm3). Most patients from subtype B (85.6%) were located in clusters 215 in contrast to subtypes A (5.2%) and G (9.2%) (Table S1).

216

### 217 Dynamics of subtype A HIV-1 epidemic in Europe

Based on the sequences from our database and the control sequences retrieved, we could observe that the majority of the subtype A population had its origin in Africa, and the major route of transmission was heterosexual. There were some individuals with IDU transmission. The phylogenetic analyses indicates that most EuResist patients cluster in two different parts of the tree, indicated with arrows A and B in figure 1, suggesting two parallel epidemics of subtype A in Europe. The first cluster was related

224 to patients originating from Africa and Eastern Europe through heterosexual and IDU transmission (cluster A) and the other was linked to patients originating from Western 225 226 Europe with MSM transmission (cluster B). LP individuals are mostly concentrated in

cluster A, where the majority of individuals are also migrants (Figure 1). 227



228 229

Figure 1. Phylogenetic tree for Subtype A. This image shows a visual phylogenetic tree of the subtype 230 A population. The region A, figure 1.a), shows a clustering of individuals originating from Africa and 231 Eastern Europe with a heterosexual and IDU transmission, the region B, figure 1.b), shows a clustering of 232 individuals originating from Western Europe and MSM transmission. C- Clusters; RO- Region of origin; 233 T- Transmission; LP- Late presenters vs non-late presenters; MS- Migration status

234

#### 235 **Dynamics of subtype B HIV-1 epidemic in Europe**

236 Based on the sequences from our database and the control sequences retrieved, we 237 could observe that most subtype B patients originated from Western Europe, are native 238 and MSM. Individuals with IDU transmission originating from Western Europe 239 dominate one cluster of the tree (indicated with an arrow). LP and NLP individuals are 240 distributed evenly in the tree. Based on the configuration of the phylogenetic tree and 241 apart from the cluster dominated by IDUs, there seems to be no major 242 compartmentalization patterns in the subtype B epidemic in Europe (Figure 2).



Figura 2. Phylogenetic tree for Subtype B. This image shows a visual phylogenetic tree of the subtype
B population. The region highlighted with an arrow shows a clustering of individuals with an IDU
transmission and originating from Western Europe. C- Clusters; RO- Region of origin; T- Transmission;
LP- Late presenters vs non-late presenters; MS- Migration status

249

### 250 Dynamics of subtype G HIV-1 epidemic in Europe

Based on the sequences from our database and the control sequences retrieved, we could observe two major regions of origin - Western Europe and Africa – compose the subtype G epidemic of HIV-1 in Europe. These are largely divided in two major clusters indicated with arrows A and B in the Figure 3. Most individuals are heterosexuals and LP dominate. The tree configuration indicates lack of compartmentalization of subtype G epidemic in Europe and suggests frequent importations of subtype G.



258 259

Figure 3. Phylogenetic tree for subtype G. This image shows a visual phylogenetic tree of the subtype
 G population. The regions highlighted with arrows shows a clustering of individuals originating from
 Western Europe and Africa. C- Clusters; RO- Region of origin; T- Transmission; LP- Late presenters vs
 non-late presenters; MS- Migration status

### 263 Determinants associated with transmission clusters of HIV-1 in Europe for 264 different subtypes

In the first unadjusted logistic regression model for Subtype A, the variables associated with a patient being in clusters from subtype A were male individuals (p<0.001), MSM (p<0.001) route of transmission, having a recent infection (p<0.001) and being NLP

268 (p=0.004) (Table S2).

In the final logistic regression model for subtype A, we adjusted the model to the variable sex, and individuals with a MSM transmission route were more likely to be in clusters when compared to heterosexual route (OR:2.65, p=0.001). Patients with a recent infection were more likely to be in clusters when compared to individuals with a chronic infection (OR:2.70, p<0.001) (Table 1.).

In the subtype B unadjusted logistic regression model, the variables associated with a patient being in clusters were male individuals (p<0.001), individuals with an age at resistance test between 19-30 years (p<0.001), route of transmission of MSM (p<0.001), patients originating from Eastern Europe (p<0.001), being migrant (p<0.001) and having a recent infection (p<0.001). Also, not having ADR (p<0.001), being NLP (p<0.001) and higher levels of viral load (p<0.001) were also associated with being in clusters from subtype B (Table S2).

281 The final logistic regression model was adjusted to sex and the determinants associated 282 with a patient being in clusters from subtype B were males (OR:1.18, p=0.037), age at 283 resistance testing, individuals with an age between 19-30 years were more likely to be 284 within clusters (OR:1.49, p=0.002) when compared to older age (>56 years old). 285 Individuals with a MSM transmission route were more likely to be in clusters when 286 compared to heterosexual route (OR:1.74, p<0.001), while individuals with a IDU 287 transmission route were less likely to be in clusters when compared to heterosexuals 288 (OR:0.58, p<0.001). Patients originated from South America had a lower probability of 289 being in clusters when compared to patients originated from Western Europe (OR:0.30, 290 p<0.001). Native individuals were less likely to be in clusters when compared to 291 migrants (OR:0.60, p<0.001) and individuals with a recent infection were more likely to 292 be in clusters when compared to individuals with a chronic infection (OR:1.88, 293 p<0.001) (Table 1.).

In the subtype G unadjusted logistic regression model, the variables associated with a patient being in clusters were female individuals (p=0.012), being native (p=0.015), having a recent infection (p<0.001), not having ADR (p<0.001) and being NLP (p=0.037) (Table S2).

The final logistic regression model was adjusted to sex and individuals from subtype G and in clusters were more likely to be native when compared to migrants (OR:1.55, p=0.021). Other factor associated with a patient being in clusters from subtype G was to have a recent infection when compared to those individuals with a chronic infection (OR: 1.99, p<0.001) (Table 1.).

		Subtype A		Subtype B		Subtype G		
In clusters/Out-of-	clusters	Final Model (Stepwise)		Final Model (Stepwise)		Final Model (Stepwise)		
		aOR (95%CI)	p-value	aOR (95%CI)	p-value	aOR (95% CI)	p-value	
Sex	Female	Ref	Ref	Ref	Ref	Ref	Ref	
	Male	1.44 (0.95-2.18)	0.083	1.18 (1.01-1.38)	0.037	0.80 (0.56-1.16)	0.237	
Age at resistance test	<18			1.32 (0.70-2.50)	0.391			
	19-30			1.49 (1.20-1.86)	<0.001			
	31-55			1.05 (0.87-1.28)	0.617			
	>56			Ref	Ref			
Transmission Route	Heterosexua 1	Ref	Ref	Ref	Ref			
	MSM	2.65 (1.50-4.69)	0.001	1.74 (1.52-2.00)	<0.001			
	IDU	0.79 (0.50-1.25)	0.308	0.58 (0.49-0.68)	<0.001			
	Other	0.91 (0.51-1.61)	0.747	0.73 (0.53-1.01)	0.057			
<b>Region of Origin</b>	Western	· · · · · · · · · · · · · · · · · · ·		Ref	Ref			
0 0	Europe							
	Eastern			1.00 (0.69-1.46)	0.997			
	Europe				0.040			
	Africa			0.75 (0.46-1.22)	0.249			
	South			0.30(0.21-0.44)	<0.001			
	America				0.000			
	Other			0.61 (0.39-0.95)	0.028			
Migration Status	Migrant			Ref	Ref	Ref	Ref	
	Native			0.60 (0.49-0.72)	<0.001	1.55 (1.07-2.25)	0.021	
Recentness of Infection	Chronic	Ref	Ref	Ref	Ref	Ref	Ref	
	Recent	2.70 (1.88-3.88)	<0.001	1.88 (1.70-2.07)	<0.001	1.99 (1.37-2.87)	<0.001	

### Table 1. Determinants associated with belonging to a transmission cluster according to Subtype A, B and G.

307

### 308 Transmission Clusters analysis:

Of the 38531 patients, 8335 were in clusters (21.6%). The minimum clusters size was 2
and the maximum clusters size was 52 (data not shown).

311 The proportion of late presenters and non-late presenters in clusters was analyzed and

312 LP were more in small clusters (95.7%) than NLP (92.4%). Also, LP within the small

313 clusters were more in dual clusters (65.6%) (2 patients per clusters) (Figure 4.A).

We also analyzed the proportion of migrants vs natives in clusters, and migrants were more in large clusters (6.6%) than natives (4.8%) (Figure 4.A).

316 According to subtypes, subtype A had a higher proportion of patients in dual clusters

(61.9%), subtype B had a higher proportion of patients in clusters of 9 or more (6.1%)

and subtype G has a higher proportion of patients in clusters between 3-5 and 6-8

319 clusters (36.4% and 8.1%, respectively) (Figure 4.B).



Figure 4. General clusters size characterization (A) and clusters size according to subtypes (B).

323

### 324 Transmission Clusters in LP vs NLP:

325 Here we studied the associated characteristics to LP vs NLP in clusters. Although, 326 within the clinical and socio-demographic characteristics, there were specific 327 characteristics of LP in clusters and NLP in clusters. LP were mainly out-of-clusters in 328 all subtypes. In subtype A the variables associated with being LP vs NLP in clusters 329 were age at resistance test (p=0.029) and recentness of infection (p=0.017), where LP 330 individuals in clusters were older than  $31y_0$  (p=0.001; p=0.030), while NLP in clusters 331 were males (p=0.002), with an age between 19-30yo (p=0.027) and a recent infection 332 (p<0.001).

333 LP vs NLP with subtype B in clusters were associated with age at resistance testing 334 (p<0.001), treatment status (p<0.001), transmission route (p=0.033), recentness of 335 infection (p < 0.001), viral load (p < 0.001) and ADR (p = 0.010). Where LP in clusters 336 were older than 31yo (p<0.001; p<0.001), ART-experienced (p<0.001), originating 337 from Eastern Europe (p<0.001), viral load higher than 5.1 copies/mL (p<0.001) and 338 having ADR (p < 0.001). While NLP in clusters were males (p < 0.001), younger than 339 30yo (p<0.001;p<0.001), ART-naïves (p<0.001), with a MSM and IDU route of 340 transmission (p < 0.001 and p = 0.047, respectively), from Western Europe (p < 0.001), 341 with a recent infection (p < 0.001) and a viral load lower than 4.0 copies/mL (p < 0.001).

LP vs NLP with subtype G in clusters were associated with the variables treatment status (p=0.015), recentness of infection (p=0.001), and transmission route (p=0.035). LP in clusters were mainly females (p=0.002), ART-experienced (p<0.001), with a chronic infection (p<0.001) and an IDU transmission route (p=0.011). For this subtype, MSM route in clusters was 100% related to NLP, but without being statistically significant (Table 2.)

348

	Inc	lusters		
	LP	NLP	p-value	p-value compared
Subtype A				-
Sex			0.052	
Male (n=109)	43 (39.4%)	66 (60.6%)		0.002
Female (n=64)	35 (54.7%)	29 (45.3%)		0.288
Age at resistance test			0.029	
<18 (n=4)	1 (25%)	3 (75%)		0.157
19-30 (n=33)	12 (36.4%)	21 (63.6%)		0.027
31-55 (n= 62)	40 (64.5%)	22 (35.5%)		0.001
>56 (n=12)	8 (66.7%)	4 (33.3%)		0.030
<b>Recentness of infection</b>			0·017	
Chronic (n=84)	45 (53.6%)	39 (46.4%)		0.351
Recent (n=95)	34 (35.8%)	61 (64.2%)		<0.001
Subtype B				
Sex			0.879	
Male (n=3421)	1632 (47.7%)	1789 (52.3%)		<0.001
Female (n=629)	298 (47.4%)	331 (52.6%)		0.065
Age at resistance test			<0.001	
<18 (n=16)	2 (12.5%)	14 (87.5%)		<0.001
19-30 (n=608)	216 (35.5%)	392 (64.5%)		<0.001
31-55 (n=2307)	1215 (52.7%)	1092 (47.3%)		<0.001
>56 (n=245)	154 (62.9%)	91 (37.1%)		<0.001
Treatment Status			<0.001	
ART-naive (n=2121)	1014 (47.8%)	1107 (52.2%)		0.004
ART-experienced (n=1284)	707 (55.1%)	577 (44.9%)		<0.001
Transmission Route			0.033	
Heterosexual (n=712)	373 (52.4%)	339 (47.6%)		0.070
MSM (n=1626)	760 (46.7%)	866 (53.3%)		<0.001
IDU (n=427)	199 (46.6%)	228 (53.4%)		0.047
Region of Origin			0.256	
Western Europe (n=2795)	1332 (47.7%)	1463 (52.3%)		<0.001
Eastern Europe (n=81)	47 (58%)	34 (42%)		0.042
Africa (n=53)	26 (49.1%)	27 (50.9%)		0.853
South America (n=88)	46 (52.3%)	42 (47.7%)		0.542
Recentness of infection			<0.001	
Chronic (n=1942)	1126 (58%)	816 (42%)		<0.001
Recent (n=2181)	834 (38.2%)	1347 (61.8%)		<0.001
Viral load (log10)			<0.001	
≤ 4.0 (n=598)	201 (33.6%)	397 (66.4%)		<0.001
4.1-5.0 (n=601)	265 (44.1%)	336 (55.9%)		<0.001
≥ 5.1 (n=537)	342 (63.7%)	195 (36.3%)		<0.001
ADR			0.010	
Yes (n=240)	296 (61.3%)	187 (38.7%)		<0.001
No (n=483)	123 (51.2%)	117 (48.8%)		0.456
Subtype G				
Sex			0.460	
Male (n=142)	78 (54.9%)	64 (45.1%)		0.099
Female (n=140)	83 (59.3%)	57 (40.7%)		0.002
Treatment Status	· · · · /		0.015	
ART-naive (n=73)	39 (53.4%)	34 (46.6%)		0.075

### 350 Table 2. Characteristics of LP and NLP in clusters according to subtypes

	In	In clusters		
	LP	NLP	p-value	p-value compared
ART-experienced (n=92)	66 (71.7%)	26 (28.3%)		<0.001
Recentness of infection			0.003	
Chronic (n=152)	99 (65.1%)	53 (34.9%)		<0.001
Recent (n=130)	62 (47.7%)	68 (52.3%)		0.458
Transmission Route			0.035	
Heterosexual (n=42)	22 (52.4%)	20 (47.6%)		0.660
MSM (n=4)	0	4 (100%)		0.516
IDU (n=25)	17 (68%)	8 (32%)		0.011

351

### 352 **Discussion:**

In the current work, the use of the genomic sequences from the EuResist database combined with genomic sequences collected from public databases provides a comprehensive sample to study and characterize HIV-1 transmission clusters in Europe. On the one hand, HIV-1 transmission investigations are possible through reconstruction of transmission clusters. Information collected through such studies in a large scale can be highly useful for public health purposes, to fine-grain transmission patterns with higher resolution compared to classical epidemiology.

360 We reconstructed transmission clusters of HIV-1 in Europe with the specific objective

of understanding the role of LP on transmission of infection. Specifically, we aimed to understand HIV-1 transmission clusters and determinants associated to transmission in clusters, taking into account the independent pandemics of the most prevalent subtypes in our population, A, B and C and to understand clustering patterns of LP and NLP in each subtype.

In our population the majority of patients were from subtype B, males, with MSM route of transmission and the region of origin of was Western Europe. These results are in accordance with a previous study conducted in Europe to analyse the distribution of subtypes (16).

We decided to study the transmission patterns of HIV-1 mainly according to subtypes, since there has been some discussion regarding the biological differences between them and mostly because of the higher prevalence of subtype B among Western Europe individuals (17). There were more patients outside TCs (78.4%) compared to those inside TCs (21.6%), in agreement with our study population based on sequences isolated at the first resistance test (18).

376 For subtype B, it was expected that one of the factors identified as associated with being 377 inside a cluster was indeed the MSM transmission route (19). For subtype A, we also 378 found MSM transmission route as a factor associated with being inside clusters. The 379 fact that the MSM route of transmission is being associated with clustering in other non-380 B subtypes is in accordance with some studies that report an increase of non-B subtypes 381 associated with MSM route (19–21). Nevertheless, we expected an association of IDU 382 route of transmission and subtype A since both subtype and route of transmission are 383 highly prevalent in Eastern Europe where this type of transmission route is also 384 prevalent (22). On the other hand, in subtype G, being inside a cluster was associated 385 with heterosexual transmission, as expected (23,24). We also found that, for subtype B, 386 the age at resistance test was associated with being in cluster, with a higher probability 387 among individuals with younger age. This is in accordance with some other studies 388 (18,25). Migration status was also associated with being in clusters: migrants infected 389 with subtype B were mainly in a cluster, and migrants infected with subtype G were less 390 likely to be in a cluster. These results were in accordance with a recent study focusing

on migration and HIV-1 in Portugal (26). These results could be explained by the region
of origin of migrants. Subtype B had higher prevalence of migrants from Brazil, and
Brazil has a concentrated HIV epidemic among MSM population (27,28).

394 Regarding the potential association between transmission clusters and late presentation,

395 we found that both LP and NLP were mainly outside clusters. As for the differences 396 between the populations of LP and NLP inside clusters, these patterns were consistent 397 between subtypes A and B: concerning sex, there were more NLP males inside cluster; 398 concerning age, there were older LP, with a higher and growing proportion as age 399 increases. Subtype G had the most different patterns of all. Furthermore, for subtypes B 400 and G individuals, there were more treated patients among LP than among NLP inside 401 clusters. As for transmission route, for subtype G, we found more LP with an IDU 402 transmission route inside clusters. Finally, for subtype B, it was interesting to observe 403 that LP located inside clusters had higher viral loads than NLP. These results could not 404 be compared by LP and NLP populations, nevertheless our results are overall in 405 accordance with some studies (18,26,28).

Finally, we found that LP were more frequently present in small clusters or outside clusters compared to NLP which can indicate a limited role of this population on HIV-1 transmission, given the less frequent presence of these patients in TCs. However higher viral loads were observed in LP located inside clusters that can indicate higher transmissibility of infection within individuals from the TCs.

Finally, there is still scarce to none information regarding transmission clusters and late presentation. We studied here the association of transmission clusters according to subtypes in LP and NLP, and our results showed that the patterns of LP vs NLP in TCs presented similar characteristics in subtypes A and B, but not in subtype Gdominated by

415 LP.

416

### 417 Limitations:

In our study we did not used the time and place of the most recent common ancester
(tMRCA), instead we used a total number of sequences from a specific region. This
methodology can cause some sampling bias since sequences can artificially be in cluster
due to their shared region of origin.

### 422

### 423 **Conclusion:**

424 In conclusion, our study presented an updated description of the socio-demographic and 425 clinical characteristics of HIV-1 infected individuals followed in Europe according to 426 subtype. Our study also highlights the patterns of transmission clusters in LP vs NLP populations selected in the european dataset of EuResist. We conclude that late 427 428 presentation could have a limited role on HIV-1 transmission. However further 429 investigation should be considered to exclude LP classification bias, and to better 430 estimate the time of infection based on phylogenetic trees reconstruction and molecular 431 clock analysis.

432

433

Author Contributions: Conceptualization, M.N.S.M., M.P. and A.A.; Methodology,
M.N.S.M., M.P., V.P., M.d.R.O.M. and A.A.; Software, M.N.S.M., V.P., S.G.S.;
Validation, M.N.S.M., M.P., F.I. and A.A.; Formal Analysis, M.N.S.M., V.P., S.G.S.,
M.P. and A.A.; Investigation, M.N.S.M., M.P., V.P., and M.d.R.O.M.; Resources, C.S.D., R.P., R.K., M.B. (Marina Bobkova), M.B (Michael Böhm), M.Z., P.G and F.I.; Data
Curation, C.S.-D., R.P., R.K., M.B. (Marina Bobkova), M.B (Michael Böhm), M.Z., P.G
and F.I.; Writing—Original Draft Preparation, M.N.S.M., M.P. and A.A.; Writing—

441 Review and Editing, M.N.S.M., M.P., C.S.-D., F.I. and A.A.; Visualization, M.N.S.M., 442 M.P., V.P., M.d.R.O.M. and A.A.; Supervision, A.A.; Project Administration, A.A.; 443 Funding Acquisition, A.A. 444 Funding: This study was financed by FCT through the following projects: GHTM-445 UID/04413/2020, INTEGRIV (PTDC/SAU-INF/31990/2017) and the scholarship 446 PD/BD/135714/2018 and COVID/BD/152613/2022 and Gilead Génese 447 HIVLatePresenters. 448 Conflicts of Interest: Author Francesca Incardona is employed by InformaPRO S.r.l. 449 The remaining authors declare that the research was conducted in the absence of any 450 commercial or financial relationships that could be construed as a potential conflict of 451 interest. 452 **References:** 453 1. Global HIV & AIDS statistics — Fact sheet | UNAIDS [Internet]. [cited 2022 Jan 454 25]. Available from: https://www.unaids.org/en/resources/fact-sheet 455 2. Grabowski MK, Herbeck JT, Poon AFY. Genetic Cluster Analysis for HIV 456 Prevention. Curr HIV/AIDS Rep [Internet]. 2018 Apr 19;15(2):182–9. Available 457 from: https://doi.org/10.1007/s11904-018-0384-1 Abecasis AB, Wensing AMJ, Paraskevis D, Vercauteren J, Theys K, Van de 458 3. 459 Vijver DAMC, et al. HIV-1 subtype distribution and its demographic 460 determinants in newly diagnosed patients in Europe suggest highly 461 compartmentalized epidemics. Retrovirology [Internet]. 2013 Dec 14;10(1):7. Available from: https://retrovirology.biomedcentral.com/articles/10.1186/1742-462 463 4690-10-7 464 4. Brenner BG, Roger M, Routy J, Moisi D, Ntemgwa M, Matte C, et al. High 465 Rates of Forward Transmission Events after Acute/Early HIV 1 Infection. J 466 Infect Dis [Internet]. 2007 Apr:195(7):951–9. Available from: 467 https://academic.oup.com/jid/article-lookup/doi/10.1086/512088 468 5. Miranda MNS, Pingarilho M, Pimentel V, Martins MDRO, Vandamme A-M, Bobkova M, et al. Determinants of HIV-1 Late Presentation in Patients Followed 469 470 in Europe. Pathogens [Internet]. 2021 Jul 2;10(7):835. Available from: 471 https://www.mdpi.com/2076-0817/10/7/835 472 6. Antinori A, Coenen T, Costagiola D, Dedes N, Ellefson M, Gatell J, et al. Late 473 presentation of HIV infection: a consensus definition. HIV Med [Internet]. 2011 474 Jan;12(1):61–4. Available from: 475 https://onlinelibrary.wiley.com/doi/10.1111/j.1468-1293.2010.00857.x 7. Late Presentation Working Groups in EuroSIDA and COHERE. Estimating the 476 477 burden of HIV late presentation and its attributable morbidity and mortality across Europe 2010-2016. BMC Infect Dis [Internet]. 2020 Oct 7;20(1):728. 478 479 Available from: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-480 020-05261-7 481 8. Rava M, Domínguez-Domínguez L, Bisbal O, López-Cortés LF, Busca C, Antela A, et al. Late presentation for HIV remains a major health issue in Spain: Results 482 483 from a multicenter cohort study, 2004–2018. Andrei G, editor. PLoS One 484 [Internet]. 2021 Apr 21;16(4):e0249864. Available from: 485 https://dx.plos.org/10.1371/journal.pone.0249864 9. Pineda-Peña A-C, Pingarilho M, Li G, Vrancken B, Libin P, Gomes P, et al. 486 487 Drivers of HIV-1 transmission: The Portuguese case. Blackard J, editor. PLoS One [Internet]. 2019 Sep 30:14(9):e0218226. Available from: 488 489 https://dx.plos.org/10.1371/journal.pone.0218226 490 10. Euresist Data Analysis - database [Internet]. [cited 2021 Jan 4]. Available from:

101		http://angina.aurosist.org/databasa/
491	11	Bannatt DE Camacho PL Otalaa D Kuritzkas DP Elaury H Kiuchi M at al
492	11.	Demicti DE, Califactio KJ, Olecca D, Kultizkes DK, Ficuly H, Kluchi W, et al.
495		Drug Resistance Mutations for Surveinance of Transmitted HTV-1 Drug-
494		Resistance: 2009 Update. Nixon DF, editor. PLoS One [Internet]. 2009 Mar $(14/2)_{12} 4724$ Assolution between (10.1271/jacramel.neuro.0004724
495	10	6;4(5):e4724. Available from: https://dx.pios.org/10.15/1/journal.pone.0004724
496	12.	Stoesser G, Griffith M, Griffith OL. HIV Sequence Database. In: Dictionary of
497		Bioinformatics and Computational Biology [Internet]. Chichester, UK: John
498		Wiley & Sons, Ltd; 2004. p. 52–61. Available from:
499		https://onlinelibrary.wiley.com/doi/10.1002/9780471650126.dob0322.pub2
500	13.	Libin PJK, Deforche K, Abecasis AB, Theys K. VIRULIGN: fast codon-correct
501		alignment and annotation of viral genomes. Hancock J, editor. Bioinformatics
502		[Internet]. 2019 May 15;35(10):1763–5. Available from:
503		https://academic.oup.com/bioinformatics/article/35/10/1763/5123354
504	14.	Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJL, et
505		al. Automated analysis of phylogenetic clusters. BMC Bioinformatics [Internet].
506		2013 Dec 6;14(1):317. Available from:
507		https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-
508		317
509	15.	Andersson E, Shao W, Bontell I, Cham F, Cuong DD, Wondwossen A, et al.
510		Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify
511		recent HIV-1 infection in transmitted drug resistance surveys. Infect Genet Evol
512		[Internet], 2013 Aug:18(1):125–31. Available from:
513		https://linkinghub.elsevier.com/retrieve/pii/\$156713481300141X
514	16	Abecasis AB Wensing AM Paraskevis D Vercauteren I Theys K Van de
515	10.	Viver DA et al HIV-1 subtype distribution and its demographic determinants in
516		newly diagnosed nations in Furone suggest highly compartmentalized
517		enidemics Retrovirology [Internet] 2013 Dec 14:10(1):7 Available from:
510		http://retrovirology biomedcentral com/articles/10/1186/17/2/4600/10/7
510	17	Kaplei DI Hamal DI Sankalá I Hajah C. Thior I Parin F. at al Human
519	17.	Immunodeficiency Virus Type 1 Subtypes Differ in Disease Progression I Infect
520		Dis [Internet] 1000 June 170(1):68, 72 Available from:
521		bttns://acadamia.oup.com/jid/articla.loclaup/doi/10.1086/21/557
522	10	Dereskavis D. Delaylas A. Stasings K. Dertaris N. de Mandaza C. Dernaut N. et
523	18.	Paraskevis D, Beloukas A, Stasinos K, Pantazis N, de Mendoza C, Bannert N, et
524		al. HIV-1 molecular transmission clusters in nine European countries and
525		Canada: association with demographic and clinical factors. BMC Med [Internet].
526		2019 Dec 8;17(1):4. Available from: $(10, 110, 110, 120, 120, 120, 120, 120, 1$
527	10	https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1241-1
528	19.	Petersen A, Cowan SA, Nielsen J, Fischer TK, Fonager J. Characterisation of
529		HIV-1 transmission clusters and drug-resistant mutations in Denmark, 2004 to
530		2016. Eurosurveillance [Internet]. 2018 Nov 1;23(44):1–9. Available from:
531		https://www.eurosurveillance.org/content/10.2807/1560-
532		7917.ES.2018.23.44.1700633
533	20.	Delgado E, Benito S, Montero V, Cuevas MT, Fernández-García A, Sánchez-
534		Martínez M, et al. Diverse Large HIV-1 Non-subtype B Clusters Are Spreading
535		Among Men Who Have Sex With Men in Spain. Front Microbiol [Internet]. 2019
536		Apr 3;10(APR):1–19. Available from:
537		https://www.frontiersin.org/article/10.3389/fmicb.2019.00655/full
538	21.	Hanke K, Faria NR, Kühnert D, Yousef KP, Hauser A, Meixenberger K, et al.
539		Reconstruction of the Genetic History and the Current Spread of HIV-1 Subtype
540		A in Germany. Kirchhoff F, editor. J Virol [Internet]. 2019 Jun 15;93(12).

541		Available from: https://journals.asm.org/doi/10.1128/JVI.02238-18
542	22.	Lai A, Bozzi G, Franzetti M, Binda F, Simonetti FR, De Luca A, et al. HIV-1 A1
543		Subtype Epidemic in Italy Originated from Africa and Eastern Europe and Shows
544		a High Frequency of Transmission Chains Involving Intravenous Drug Users.
545		Tornesello ML, editor. PLoS One [Internet]. 2016 Jan 11;11(1):e0146097.
546		Available from: https://dx.plos.org/10.1371/journal.pone.0146097
547	23.	Yebra G, Holguín Á, Pillay D, Hué S. Phylogenetic and demographic
548		characterization of HIV-1 transmission in Madrid, Spain. Infect Genet Evol
549		[Internet]. 2013;14(1):232–9. Available from:
550		http://dx.doi.org/10.1016/j.meegid.2012.12.006
551	24.	Lorenzin G, Gargiulo F, Caruso A, Caccuri F, Focà E, Celotti A, et al.
552		Prevalence of Non-B HIV-1 Subtypes in North Italy and Analysis of
553		Transmission Clusters Based on Sequence Data Analysis. Microorganisms
554		[Internet]. 2019 Dec 23;8(1):36. Available from: https://www.mdpi.com/2076-
555		2607/8/1/36
556	25.	Pingarilho M, Pimentel V, Miranda MNS, Silva AR, Diniz A, Ascenção BB, et
557		al. HIV-1-Transmitted Drug Resistance and Transmission Clusters in Newly
558		Diagnosed Patients in Portugal Between 2014 and 2019. Front Microbiol
559		[Internet]. 2022;13. Available from:
560		https://www.frontiersin.org/article/10.3389/fmicb.2022.823208
561	26.	Pimentel VF, Pingarilho M, Sole G, Alves D, Miranda M, Diogo I, et al.
562		Differential patterns of postmigration HIV-1 infection acquisition among
563		Portuguese immigrants of different geographical origins. AIDS [Internet]. 2022
564		Jun 1;36(7):997–1005. Available from:
565		https://journals.lww.com/aidsonline/Fulltext/9000/Differential_patterns_of_post_
566		migration_HIV_1.96205.aspx
567	27.	MacCarthy S, Brignol S, Reddy M, Nunn A, Dourado I. Late presentation to
568		HIV/AIDS care in Brazil among men who self-identify as heterosexual. Rev
569		Saude Publica [Internet]. 2016;50:54. Available from:
570		http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-
571		89102016000100233&lng=en&tlng=en
572	28.	Kostaki EG, Gova M, Adamis G, Xylomenos G, Chini M, Mangafas N, et al. A
573		Nationwide Study about the Dispersal Patterns of the Predominant HIV-1
574		Subtypes A1 and B in Greece: Inference of the Molecular Transmission Clusters.
575		Viruses [Internet]. 2020 Oct 19;12(10):1183. Available from:
576		https://www.mdpi.com/1999-4915/12/10/1183
577		
578		

### 579 Table Supplementary 1. Patients socio-demographic and clinical characteristics

### 580

Patient Characteristics	Total	Subtype A	Subtype B	Subtype G	n-value
Total	38531 (100)	2276 (5·9)	32652 (84.7)	3603 (9.4)	p-value
Sex, n (%)	36699 (95.2)	2011 (5.5)	31200 (85.0)	3488 (9.5)	
Male	27715 (75.5)	1156 (4.2)	24570 (88.6)	1989 (7.2)	<0.001
Female	8984 (24.5)	855 (9.5)	6630 (73.8)	1499 (16•7)	-
Median age at resistance	18421 (47.8)	631 (3.4)	17149 (93.1)	641 (3.5)	
teste in years IQR, n (%)	40.0 (34.0-47.0)	37.0 (30.0-45.0)	41.0 (35.0-47.0)	34.0 (28.0-41.0)	<0.001
< 18					
<u>&gt; 18</u>	233 (1·3)	32 (13.7)	143 (61•4)	58 (24.9)	<0.001
21.55	2529 (13.7)	150 (5.9)	2209 (87•4)	170 (6•7)	
> 56	14199 (77.1)	390 (2.7)	13430 (94•6)	379 (2.7)	-
$\leq 50$	1460 (7•9)	59 (4.0)	1367 (93.6)	34 (2·4)	
I ransmission Koute, n (%)	18140 (47.1)	1230 (6.8)	16129 (88•9)	781 (4·3)	-
MSM	5568 (30.7)	532 (9.6)	4646 (83•4)	390 (7.0)	<0.001
IDU	6692 (36•9)	97 (1•4)	6566 (98.1)	29 (0.4)	-
	4883 (26•9)	433 (8.9)	4269 (87.4)	181 (3.7)	-
Other	997 (5.5)	168 (16•9)	648 (65.0)	181 (18·2)	
Region of origin, n (%)	23647 (61.4)	1277 (5•4)	20595 (87.1)	1775 (7.5)	-
Western Europe	20440 (86•4)	482 (2·4)	18836 (92·2)	1122 (5.5)	-
Eastern Europe	933 (3.9)	468 (50.2)	365 (39.1)	100 (10.7)	<0.001
Africa	1162 (4.9)	294 (25·3)	343 (29.5)	525 (45.2)	-
South America	787 (3·3)	11 (1·4)	754 (95.8)	22 (2.8)	-
Other	325 (1.4)	22 (6.8)	297 (91.4)	6 (1.8)	
Migration Status	16101 (41.8)	574 (3.6)	14629 (90•9)	898 (5.6)	-0-001
Migrant	2543 (15.8)	351 (13.8)	1771 (69•6)	421 (16.6)	<0.001
Native	13558 (84.2)	223 (1.6)	12858 (94.8)	477 (3.6)	
Clusters	38531 (100)	2276 (5•9)	32652 (84.7)	3603 (9•4)	0.000
In Clusters	8335 (21.6)	433 (5•2)	7136 (85.6)	766 (9•2)	0.006
Out-of-Clusters	30196 (78•4)	1843 (6•1)	25516 (84.5)	2837 (9•4)	
Treatment Status	21687 (56•3)	709 (3.3)	19387 (89•4)	1591 (7.3)	0.001
ART-naive	8887 (41.0)	443 (5.0)	8046 (90.5)	398 (4.5)	<0.001
ART-experienced	12800 (59•0)	266 (2.1)	11341 (88.6)	1193 (9•3)	
Recentness of infection	38531 (100)	2276 (5•9)	32652 (84.7)	3603 (9•4)	0.001
Recent	15571 (40•4)	929 (6•0)	13322 (85.6)	1320 (8.5)	<0.001
Chronic	22960 (59.6)	1347 (5•9)	19330 (84·2)	2283 (9.9)	
TDR	7727 (20.1)	393 (5.1)	7035 (91.0)	299 (3.9)	_
Yes	967 (12.5)	17 (1.8)	918 (94•9)	32 (3·3)	<0.001
No	6760 (87.5)	376 (5.6)	6117 (90.5)	267 (3.9)	
ADR	6184 (16•0)	133 (2·2)	5320 (86.0)	731 (11.8)	_
Yes	4542 (73•4)	73 (1.6)	3958 (87.1)	511 (11·3)	<0.001
No	1642 (26•6)	60 (3.7)	1362 (82.9)	220 (13·4)	
Median CD4 count at	24321 (63·1)	1011 (4·2)	21583 (88.7)	1727 (7.1)	_
(%)	341.0 (170.0-540.0)	328.0 (174.0-510.0)	349•0 (172•0-547•0)	273.0 (139.0-445.0)	<0.001
LP	12501 (51.4)	545 (4.4)	10875 (87.0)	1081 (8.6)	
NLP	11820 (48.6)	466 (3•9)	10708 (90.6)	646 (5.5)	<0.001
Viral Load at diagnosis	15670 (40•7)	614 (3.9)	13105 (83.6)	1951 (12.5)	
(log <sub>10</sub> copies/mL) IQR, n (%)	4.3 (3.4-5.0)	4•3 (3•4-5•1)	4.3 (3.3-5.0)	4.5 (3.7-5.1)	<0.001
≤ 4.0	6210 (39.6)	246 (4.0)	5315 (85.6)	649 (10.5)	
4.1-5.0	5361 (34·2)	197 (3.7)	4436 (82•7)	728 (13.6)	<0.001
≥ 5.1	4099 (26·2)	171 (4·2)	3354 (81.8)	574 (14.0)	

582	TableSupplementary2.	Unadjusted	analysis	for	determinants	associated	with	belonging	to	а
583	transmission cluster according	ng to Subtype	A, B and	G						

		Subtype A		Subtype B		Subtype G		
In clusters/Out-of-cl	usters	Unadjusted Model		Unadjusted Model	Unadjusted Model		Unadjusted Model	
		uOR (95%CI)	p-value	uOR (95%CI)	p-value	uOR (95% CI)	p-value	
Sex	Female	Ref	Ref	Ref	Ref	Ref	Ref	
	Male	1.84 (1.46-2.32)	<0.001	1.49 (1.38-1.59)	<0.001	0.81 (0.69-0.96)	0.012	
Age at resistance test	<18	0.56 (0.16-1.90)	0.353	0.72 (0.45-1.17)	0.187	1.85 (0.35-9.71)	0.469	
	19-30	1.19 (0.57-2.49)	0.641	1.72 (1.47-2.03)	<0.001	2.89 (0.65-12.80)	0.162	
	31-55	0.76 (0.38-1.50)	0•423	0.91 (0.80-1.06)	0.223	2.16 (0.50-9.30)	0.303	
	>56	Ref	Ref	Ref	Ref	Ref	Ref	
Transmission Route	Heterosexual	Ref	Ref	Ref	Ref	Ref	Ref	
	MSM	3.46 (2.12-5.65)	<0.001	1.84 (1.67-2.02)	<0.001	1.51 (0.62-3.67)	0.367	
	IDU	0.76 (0.51-1.13)	0.178	0.60 (0.53-0.68)	<0.001	1.52 (0.99-2.34)	0.055	
	Other	1.06 (0.64-1.77)	0.810	0•76 (0•60-0•97)	0.025	0.76 (0.46-1.25)	0.276	
Region of Origin	Western Europe	Ref	Ref	Ref	Ref	Ref	Ref	
	Eastern Europe	0.63 (0.46-0.87)	0.002	1.58 (1.26-1.99)	<0.001	0.67 (0.40-1.14)	0.141	
	Africa	0.55 (0.37-0.80)	0.002	1.11 (0.86-1.43)	0.424	0.57 (0.43-0.75)	<0.001	
	South America	1.20 (0.31-4.59)	0.793	0.94 (0.78-1.12)	0.478	2.12 (0.90-5.02)	0.087	
	Other	1.20 (0.46-3.13)	0.714	1.02 (0.77-1.34)	0.917	0.61 (0.07-5.27)	0.656	
Migration Status	Migrant	Ref	Ref	Ref	Ref	Ref	Ref	
	Native	1.48 (0.97-2.27)	0.070	0.72 (0.64-0.81)	<0.001	1.57 (1.09-2.26)	0.015	
Recentness of Infection	Chronic	Ref	Ref	Ref	Ref	Ref	Ref	
	Recent	2.70 (2.18-3.35)	<0.001	2.22 (2.10-2.34)	<0.001	2•31 (1•96-2•72)	<0.001	
ADR	Yes	Ref	Ref	Ref	Ref	Ref	Ref	
	No	2.01 (0.87-4.64)	0.102	1.54 (1.30-1.82)	<0.001	2.35 (1.47-3.74)	<0.001	
LP/NLP	LP	Ref	Ref	Ref	Ref	1	1	
	NLP	1.61 (1.16-2.23)	0.004	1.15 (1.08-1.23)	<0.001	1.32 (1.02-1.71)	0.037	
Viral load groups	<4.0	Ref	Ref	Ref	Ref	1	1	
	4.1-5.0	1.15 (0.67-1.98)	0.603	1.27 (1.14-1.42)	<0.001	1.16 (0.88-1.53)	0.297	
	>5.1	1.66 (0.98-2.82)	0.060	1.55 (1.38-1.73)	<0.001	1.24 (0.93-1.66)	0.144	