CROI 2014
Poster 600
Session P-L3

M184I and M230I Minority Variants in ART-naïve Patients are Linked to APOBEC3G/F Activity

Noguera-Julian M[1], Cozzi-Lepri A[2], di Giallonardo F[3], Schuurman R[4], Däumer M[5], Aitken S[6], Günthard HF[3], Brun-Vezinet F[6], Paredes R[1], Metzner KJ[3], and the CHAIN Minority HIV-1 Variants Working Group

[1]Institut de Recerca de la SIDA IrsiCaixa, Badalona, Catalonia, Spain; [2]Royal Free Hospital, London, UK; [3]University Hospital Zurich, Zurich, Switzerland; [4]Utrecht Medical Centre, Utrecht, Netherlands; [5]Institut für Immunologie und Genetik, Kaiserslautern, Germany; [6]Association de Recherche en Virologie et Hematologie, France

M. Noguera-Julian
Irsicaixa – AIDS Research Lab
Hospital Germans Trias i Pujol
Ctra. de Canyet s/n, 08916
Badalona, Catalonia, Spain
mnoguera@irsicaixa.es

Cumulative evidence indicates that minority (or low-frequency) drug-resistant HIV-1 variants impair virological responses to first-line antiretroviral therapy (ART) containing non-nucleoside reverse transcriptase inhibitors (NNRTIs). The origin of such variants remains unclear: whereas drug-resistant variants can be transmitted, they could also be generated spontaneously through a variety of mechanisms. In a previous pan-European case-control study, we showed that the pre-existence of minority variants harbouring resistance to reverse transcriptase inhibitors more than doubled the risk for virological failure in ART-naïve patients subjects starting efavirenz (EFV) or nevirapine (NVP) containing regimens. Interestingly, the six most prevalent mutations detected in the study corresponded to G-to-A nucleotide mutations, which resulted in Valine to Isoleucine (V90I, V106I, and V108I), Methionine to Isoleucine (M184I and M230I) or Glutamate to Lysine (E138K) mutations. This observation prompted us to investigate whether such minority variants could have been generated through APOBEC3G/F editing.
APOBEC3G/F (Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3G/F) proteins are cellular proteins within the APOBEC family that inhibit replication of Human Immunodeficiency Virus type 1(HIV-1) and other lentiviruses through cytidine deamination of their genomes. APOBEC3G/F molecules are packaged into nascent virions and mutate the DNA produced by the viral Reverse Transcriptase upon desencapsulation. Cytidines in the newly synthetized minus DNA strand are deaminated and turned into uracils, thus, producing G-to-A mutations in the viral plus strand. APOBEC3G/F proteins show a high specificity on the upstream genetic context in the minus strand.

Additional studies indicate the preference of APOBEC3G and APOBEC3G/F for 5'-GG-3' and 5'-GA-3', respectively. Of note, 5'-TGGG-3' is an important hotspot for APOBEC3G activity. The main effect of the APOBEC3G/F proteins on HIV-1 is hypermutating its genome, thus preventing viral replication through the introduction of in-frame STOP codons, but also of other mutations, which impair viral fitness. The effect of APOBEC3G/F is counteracted by the HIV-1 *vif* protein, which binds to APOBEC3G/F and recruits an ubiquitine ligase protein complex that targets APOBEC3G/F for degradation. Since the strength of this specific interaction and subsequent APOBEC3G/F degradation varies on a small set of *vif* and APOBEC3G/F mutations *vif*-mediated inhibition of APOBEC3G/F activity may be total or partial; the latter can generate sublethal levels of viral genetic diversity. Therefore, these virus-host coevolutionary dynamics through APOBEC3G/F-*vif* interactions, provides HIV-1 with new paths for adaptation to selective pressure and can lead to the generation of new resistance mutations depending on the background viral genomic context. Particularly, previous studies using different modified cell-lines and virus either expressing/not expressing APOBEC3G/F or with defective *vif* proteins have shown that mutations generated by this host defense mechanism are used by HIV-1 to generate viable and, in some cases, mutations in RT that provide resistance to several antiretroviral drugs through well defined mutations. Here, we sought to elucidate if minority drug-resistant variants detected in ART-naïve subjects could originate from APOBEC3G/F editing.

Methods

- ***Study design.*** This is a sub analysis of HIV-1 sequences obtained by 454 pyrosequencing during a large pan-European case-control study aimed at determining the effect of pre-existing drug-resistant minority variants on the virological outcomes of first-line ART including NNRTIs. The details of the parent study have been described elsewhere (2) Importantly, original case-control design and sample selection was disregarded and all available datasets were used, since we were not interested in the clinical evolution of patients but on the association of distinct virological characteristics.
- ***Plasma samples processing, RT-PCR, and 454 pyrosequencing.*** Samples collected from six different European cohorts (Swiss HIV Cohort Study, EuroSIDA Study in EuroCOORD (EuroSIDA), Italian Cohort of Antiretroviral-Naïve Patients (ICONA), German Competence Network (KompNet), AREVIR Platform, ANRS CO3 Aquitaine Cohort (Aquitaine), and Royal Free Hospital (RFH) Database) were shipped to the Utrecht Medical Centre, Utrecht, Netherlands, where cDNA was generated. cDNAs were then shipped to the Institut für Immunologie und Genetik, Kaiserslautern, Germany, where 454 amplicons were immediately prepared and pyrosequenced in each in a 454 FLX Genome Sequencer using the Titanium chemistry. Amplicons used for this analysis included amplicons A (protease), B, C, and D (reverse transcriptase) from the 454/Roche HIV-1 Genotyping kit v2, plus one additional amplicon designed specifically for this study (amplicon E),which covered the main NNRTI mutations (Figure 1). Of note, amplicon A was amplified, but was not sequenced.



Figure 1. Amplicon design for HIV-1 drug resistance testing using 454 pyrosequencing. Amplicons A (protease), B, C, and D (reverse transcriptase) were derived from the 454/Roche HIV-1 Genotyping kit v2· Amplicon E was designed specifically for this study covering the main NNRTI mutations.

Data Analysis.(Figure2)
•Aligned error-corrected unique sequences were obtained from AVA software (v 2.7.0) for all samples. Sequence alignments were produced by *Mosaik* Aligner Software (v.2.1.73).(19) Perl scripting language, R,(20) and Samtools(21) libraries were used for sequence data manipulation and statistical analysis.
•For each of the samples, a reference-guided pairwise alignment using HXB2 genome reference (Genbank ID: K03455.1) was produced. Each sequence was analysed individually in order to obtain an amplicon-sized haplotype with the corresponding amino acid polymorphisms and resistance associated mutations. Additionally, from each alignment a sample-specific consensus sequence was obtained which covered the sequenced region.
•Sequence reads for each sample were re-aligned to the corresponding sample-specific consensus sequence. Using these alignments, APOBEC3G/F target (GRD) and APOBEC3G/F control (GYN|GRC), were located and counted either as mutated or non-mutated when comparing aligned sequence read to sample specific consensus reference sequence mimicking Hypermut2 tool (http:// www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html). Using these counts, a Fisher's Exact test was realized on each sequence and an APOBEC3G/F-edited status flag was linked to each sequence if p<0.05.
•Statistical tests were run at two different levels: first, sample level co-occurrence of presence of each MV and APOBEC3G/F status within a sample was used to create a contingency table. Second, a global population analysis was performed by analysing the presence of one or more MV and APOBEC3G/F in the same clonal sequence. Fisher's exact test was used to check for statistical significant differences at both levels.
•Multiple sequence alignments were created using MAFFT software.(22) Maximum likelihood phylogenetic analysis in selected samples were performed using PhyML(v3.1)(23) Best evolution model was selected for each sample/amplicon and a maximum likelihood phylogenetic tree was calculated using the Subtree Pruning and Regrafting (SPR) algorithm and a 1000 bootstrap value.

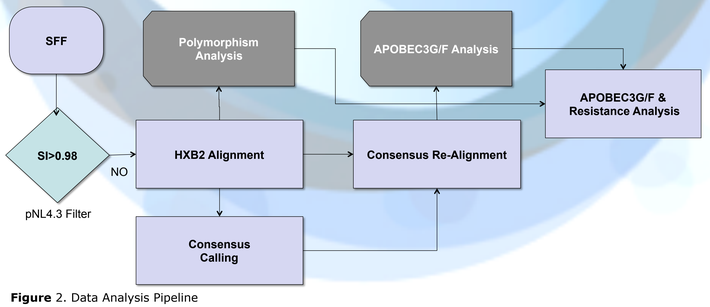

Figure 2. Data Analysis Pipeline

Results

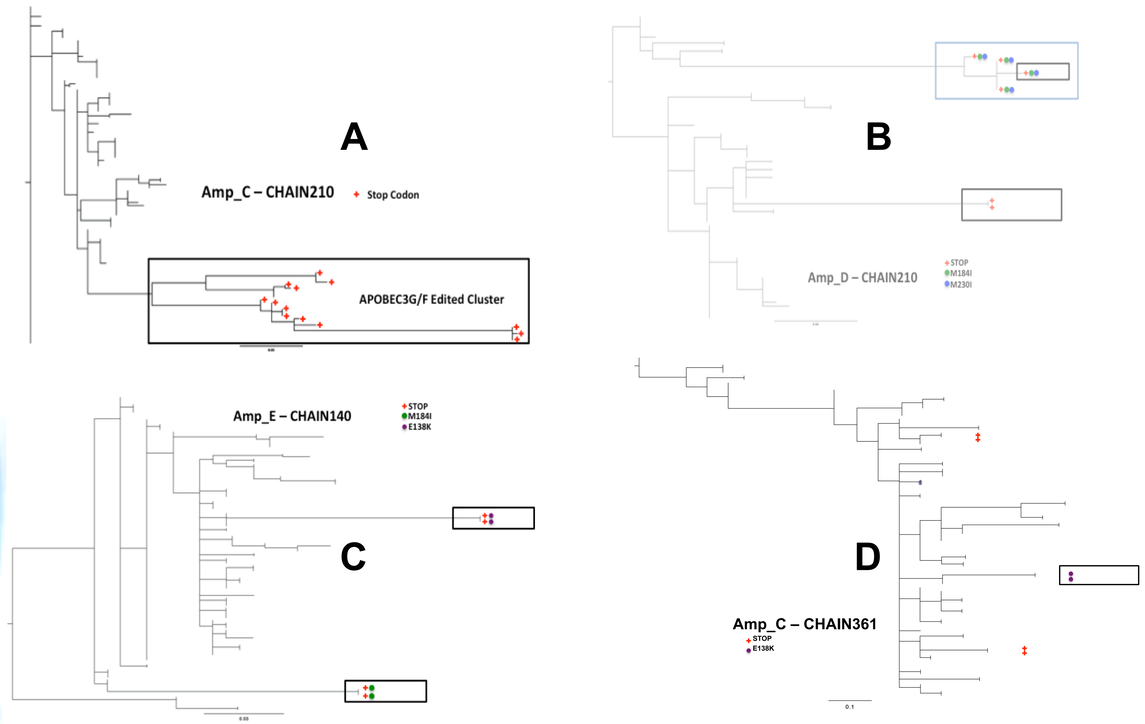

Figure 2 Maximum Likelihood phylogenetic tree for amplicons C and D sequences from sample CHAIN210, amplicon E for CHAIN140 and amplicon C for CHAIN361. Solid black squares indicate APOBEC3G/F edited clusters (Fisher Test p<0.05. Red cross: Sequence contains in-frame STOP codon/s) Green dot: Sequence contains M184I. Blue dot: Sequence contains M230I.. Purple dot: Sequence contains E138K)

- Data of 332 samples was available. 326 (93%) samples provided sequence data suitable for analysis for one or more amplicons, while 185 (55.7%) provided a high number (>400) of sequences for all amplicons. (Table 1) Data from 6 samples was not evaluable.
- APOBEC3G/F edition signature was detected in one or more sequences of 32 (9.8%) samples. Within those, APOBEC3G/F edition signature was detected in 1% or more of the sequences in 8 of the samples. The percentage of sequences within each sample for which APOBEC3G/F edition was detected ranged from 0% to 12%. (Table 1)
- M184I, M230I showed strong statistical association (p<0.0001) with same sample co-occurrence of APOBEC3G/F hypermutation while E138K association was marginally significant (p=0.043). (Table 3)
- V106I, and V108I, V90I did not appear more frequently in APOBEC3G/F edited samples.
- Clonally, M184I, M230I and E138K did not only tend to occur in the same samples that show APOBEC3G/F edition, but also on the same sequence viral clone sequences, also when different amplicons covering each position of interest were interrogated. Conversely, V90I, V108I or V106I, tend to appear independently of APOBEC3G/F edition. (Table 4)
- All clonal sequences with an APOBEC3G/F footprint and either M184I or M230I also harboured STOP codons, while two samples showed a non-STOP linked APOBEC3G/F mutated sequences. 12 more samples harboured some amplicon-length sequence which was APOBEC3G/Fhypermutated and did not contain a STOP codon. Globally, these sequences accounted for a 16% of APOBEC3G/F edited sequences.
- Phylogenetic Analysis for three representative APOBEC3G/F edited samples/amplicons shows supported clustering(>80%), of in-frame STOP codons and co-occurrence of APOBEC3G/F edition sequences and either M184I, M230I and E138K for amplicons covering these positions. (Figure 2A, 2B and 2C). Sample CHAIN361 APOBEC3G/F edited sequences clustering in a similar way to previous cases sequences.(Figure 2D) However, for this sample, no in-frame STOP codons are found, although E138K is indeed found.

Table 1. Sequencing results and APOBEC3G/F detection for the four different amplicons used. For each amplicon number of target APOBEC3G/F sites and proportion of sequences with APOBEC3G/F detection are given(Fisher's Exact Test p<0.05).

	Amp_B	Amp_C	Amp_D	Amp_E
Samples With Sequence Data (>400)	294	252	236	284
Mean Number of Sequences	1525	1712	1744	1560
Median number of APOBEC target sites*	29	37	37	38
APOBEC3G/F edited sequences (% p<0.05)	0.02%	0.26%	0.03%	0.11%

Table 2. Intrasample proportion of sequences showing APOBEC3G/F edition (p≤0.05 or p≤0.1) ad specific G-to-A IAS-USA listed resistance mutation. Only samples with ≥1% of APOBEC3G/F are shown.

	%Sequences							
SampleID	p≤0.05	p≤0.1	M184I	M230I	E138K	V108I	V106I	V90I
CHAIN130	12.74	12.74	20.45	N/C	0.00	0.00	0.00	0.00
CHAIN247	5.97	5.97	0	8.68	0.00	0.00	0.00	0.00
CHAIN210	4.87	11.44	13.28	17.87	0.00	0.00	0.00	0.00
CHAIN083	2.94	2.94	4.53	0	0.32	0.00	0.00	0.00
CHAIN151	1.68	1.68	1.77	2.60	0.00	0.00	3.27	0.00
CHAIN061	1.67	2.12	0.00	0.00	0.00	1.86	0.00	0.00
CHAIN146	1.13	6.90	0.00	0.00	0.00	0.00	0.00	0.00
CHAIN338	1.00	1.00	0.61	1.00%	0.00	0.00	0.00	0.00

*No Sequence data at this position(Not Covered)

Table 3. Statistical association of the detection of specific mutations and APOBEC3G/F hypermutations.

		APOBEC3G/F associated hypermutation Yes(31)	No(294)	Fisher Test (p)
M184I	Yes	9	13	0.00005
M230I	Yes	6	6	0.00034
E138K	Yes	5	16	0.043
V90I	Yes	3	39	0.78
V108I	Yes	4	24	0.50
V106I	Yes	4	25	0.51

Table 4. Association of drug resistance mutations and APOBEC3G/F edited status at the clonal level. Co-occurrence of resistance mutation and APOBEC3G/F detection for each unique sequence was used to create a contingency table. Fisher's test p-value are shown.

Mutation	Fisher Test(p-value)			
	Amp_B	Amp_C	Amp_D	Amp_E
E138K	N/C	2.74E-25	N/C	4.48E-10
M184I	N/C	N/C	9.78E-24	1.64E-46
M230I	N/C	N/C	2.70E-29	N/C
V106I	N/C	1	N/C	3.08E-06
V108I	N/C	1	N/C	1
V90I	N/C	1	N/C	1

Discussion

Several studies have detected the presence of M184I, M230I and E138K as a result of APOBEC3 family hypermutation on HIV genome using either DNA from Peripheral Blood Mononuclear Cells (PBMC), viral reservoirs and tissue, and also from different in vitro designs with vif defective viral constructs. In this dataset, APOBEC3G/F shows a continuous and gradual impact on HIV-1 circulating genome that results in the generation of spontaneous mutations, both resistant and not resistant, and also frequently in the generation of STOP codons which render the virus non-infective. This is reinforced by the fact that 16% of the sequences that are APOBEC3G/F edited sequences do not contain STOP codons. Interestingly, *in vitro* experiments suggested that clones harbouring STOP codons along with resistance mutations may serve as a source of ART resistance through recombination.
In all, results here presented support a role for APOBEC3G/F proteins in the generation, not only of HIV genetic diversity, but also of resistance, that could facilitate viral escape from ART pressure through particular mutations. Reasons for the greater impact of APOBEC3G/F detection among different samples may include different host genetics, viral genomic background, presence of vif polymorphisms which affect APOBEC3G/F degradation efficacy and to determine if the detection of APOBEC3G/F in plasma viral RNA is linked to an increased/decreased risk of viral failure.

This study was supported through an unrestricted research grant from Merck Sharpe & Dohme, 'CHAIN, Collaborative HIV and Anti-HIV Drug Resistance Network', Integrated Project no. 223131, funded by the European Commission Framework 7 Program.