

NON-RANDOM DISTRIBUTION OF CRYPTIC REPEATING TRIPLETS OF PURINES AND PYRIMIDINES (RNY)_n AND RECOMBINATION IN GP120 OF HIV-1

E. De Crignis¹, S. Guglietta¹, B. T. Foley², M. Negroni³, G. Pantaleo¹, C. Graziosi¹

¹Laboratory of AIDS Immunopathogenesis, Division of Immunology and Allergy, Department of Medicine, CHUV, Lausanne, Switzerland

²Theoretical Biology and Biophysics Group, Los Alamos National Laboratories, Los Alamos, NM

³Architecture et Reactivite de l'ARN, Université de Strasbourg, CNRS, IBMC, Strasbourg, France



Introduction

During HIV-1 replication cycle a large number of mutations, including transversions and transitions as well as insertions and deletions, can occur. Events of length polymorphism involving indels spanning multiples of three bases occur in all gp120 variable regions.

We previously demonstrated the presence of elements of misalignment with mutagenic potential in gp120 such as palindromic sequences, long duplications and repeated trinucleotide. RNY motifs such as AAT or AVT have been reported to appear with higher frequencies in the length-variable portions of gp120 and to be related to size variation while being able to encode for glycosylation sites (Bosch et al, J. Virol. 1994; Kitrinis et al., J. Virol. 2003).

We assessed whether patterns of alternations of purines and pyrimidines could be recognized in variable domains of gp120 that could be regarded as hotspots associated to the presence of indels.

Materials and Methods

- Cloning and sequencing of HIV variants from 7 naive patients (Patient 1 underwent also to SGA analysis)
- RNY analysis has been performed using an in-house R script in the R environment
- Recombination analysis was performed using Splitstree package and by the program RDP3Beta30
- Codon usage was calculated using Bioedit, significance in codon usage between different regions were evaluated with a T test

Results

Variable regions of gp120 are enriched in RNY trinucleotides: Major indels affecting the distribution of PNG sites of gp120 can be observed in all variants isolated. Patterns of alternating purines (R) and pyrimidines (Y) can be retrieved in variable regions of the protein. The frequency of stretches of RNY (spanning from four to seventeen repeats) is significantly higher in variable regions than in constant regions (Fig. 1).

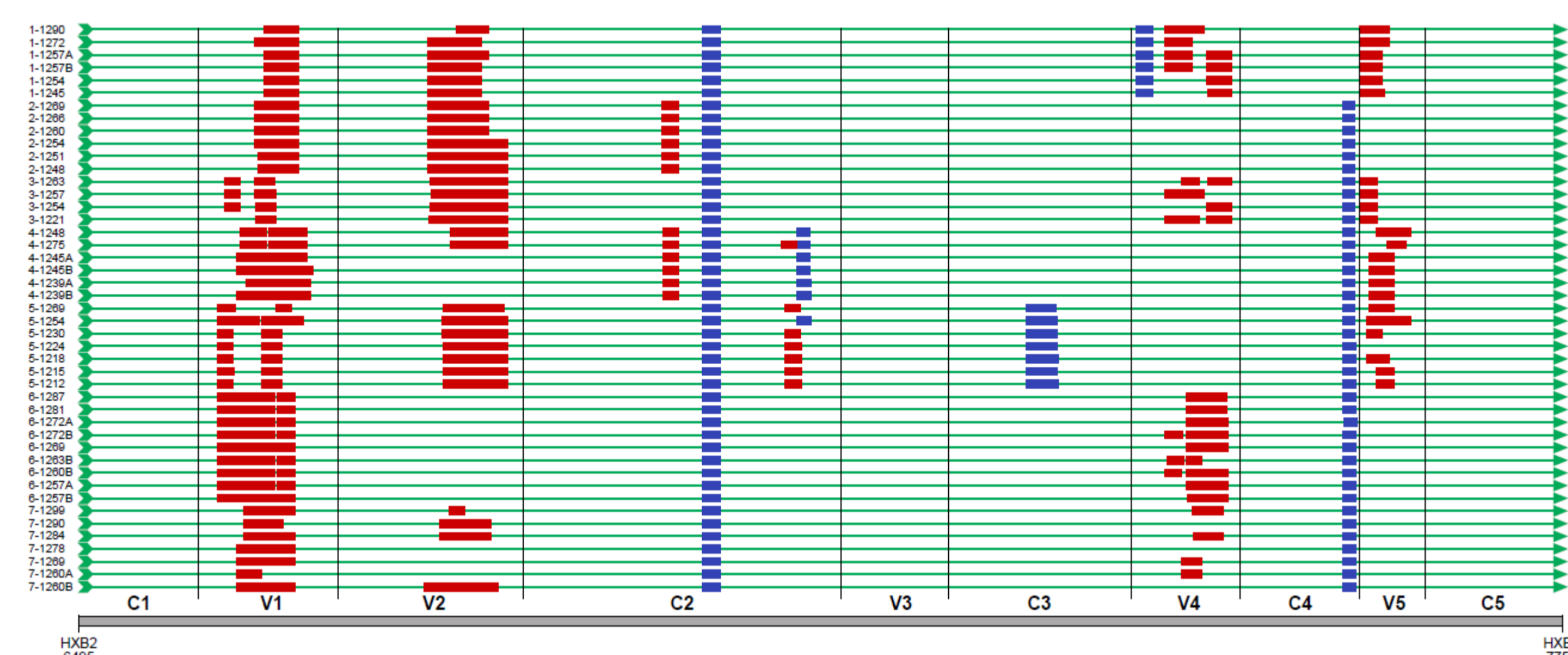
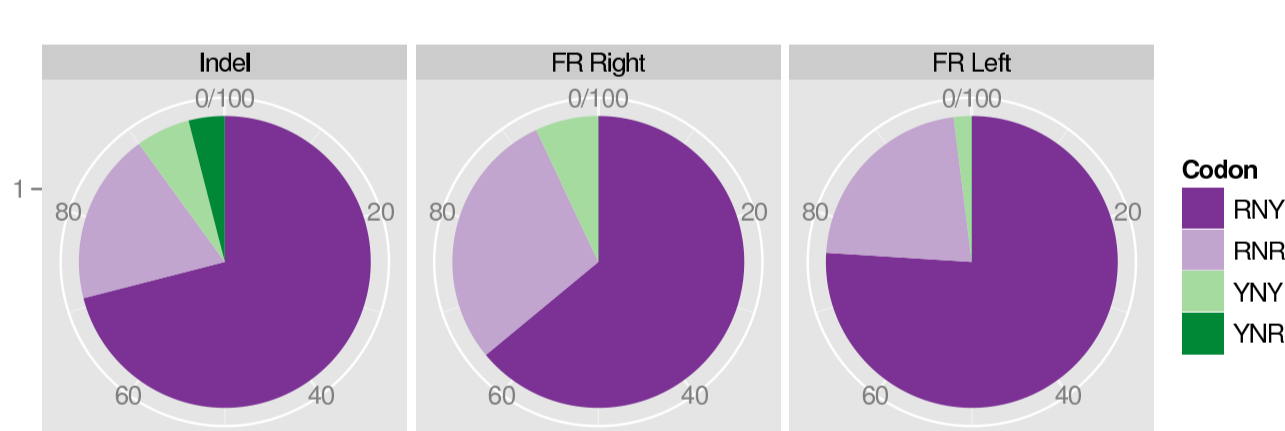


Fig.1: Stretches of contiguous in-frame RNY codons (red bars) along gp120



Sequences comprised within indels consist mainly of RNY codons and the indel events are usually flanked by RNY codons as well (Fig.2).

Fig.2: Codon composition of indels and indels flanking regions (FR)

A linear relationship between length of the indels and number of RNY codons can be identified ($R^2=0.90$). The association is weaker when only asparagine is considered ($R^2=0.55$) and absent when considering other codons (Fig. 3).

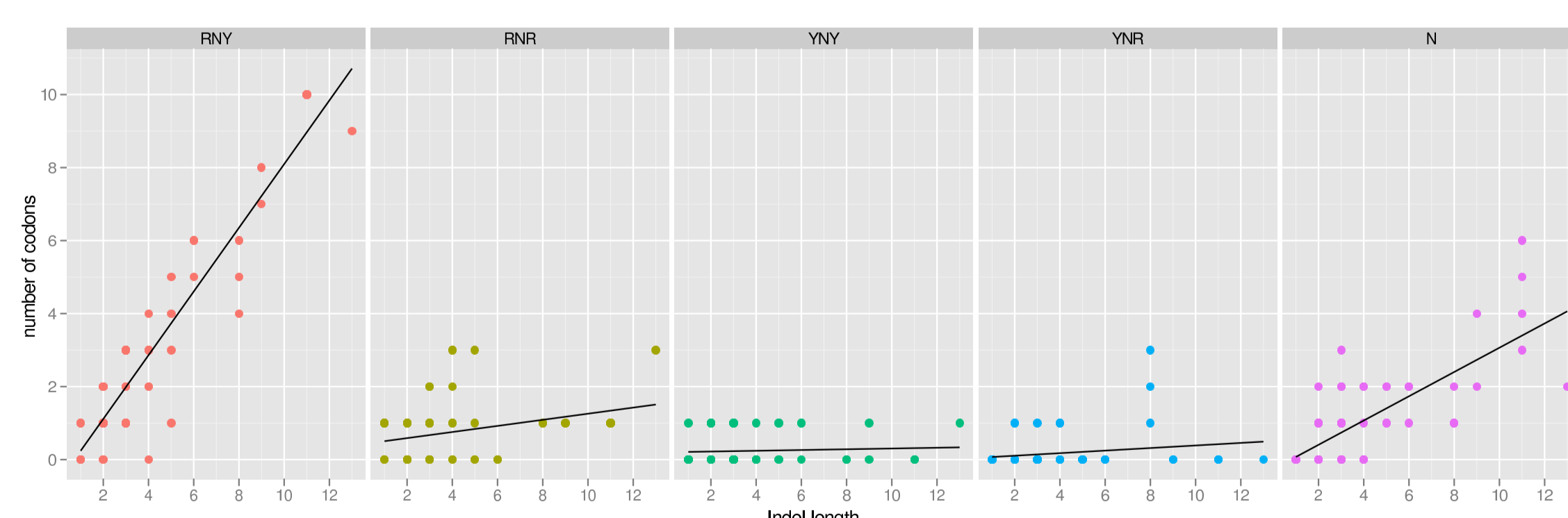


Fig.3: Linear relationship between length of the indels and number of codons. The last panel (N) shows the relationship between length of the indels and the number of asparagine residues.

Codon usage is markedly different among gp120 regions: Aminoacids that can be encoded either by RNY or not-RNY codons are mainly encoded by RNY codons inside indels (Fig. 4).

Fig.4: Frequency of residues encoded by RNY codons in Env, Constant Regions, Variable Regions and inside indels.



The sole codon AVT did not constitute a motif: AVT codons are less than the 50% of RNY codons. Strings made by AVT alone are very short and distributed unevenly among the different regions, masking the recurrent RNY pattern.

Sample	RNY non-AVT gp120	RNY non-AVT Indels	Sample	RNY non-AVT gp120	RNY non-AVT Indels
Pt1	50.7	81.2	Pt4	51.8	46.9
Pt2	60.1	42.9	Pt5	56.9	29.4
Pt3	52.1	82.5	Pt6	58.3	64.8
Pt7	47.5	40.4	Average	54.2	53.9

Table 1: Codon-content (%) of RNY non-AVT in full length gp120 and within inserted/deleted fragments.

RNY distribution differs between HIV-1 genes: Stretches of 4 or more RNY are mainly located in Env. The frequency of aminoacids encoded by RNY is significantly different ($p < 0.01$, Kruskal-Wals test) among HIV regions with a higher prevalence of RNY codons in Env (Fig.5).

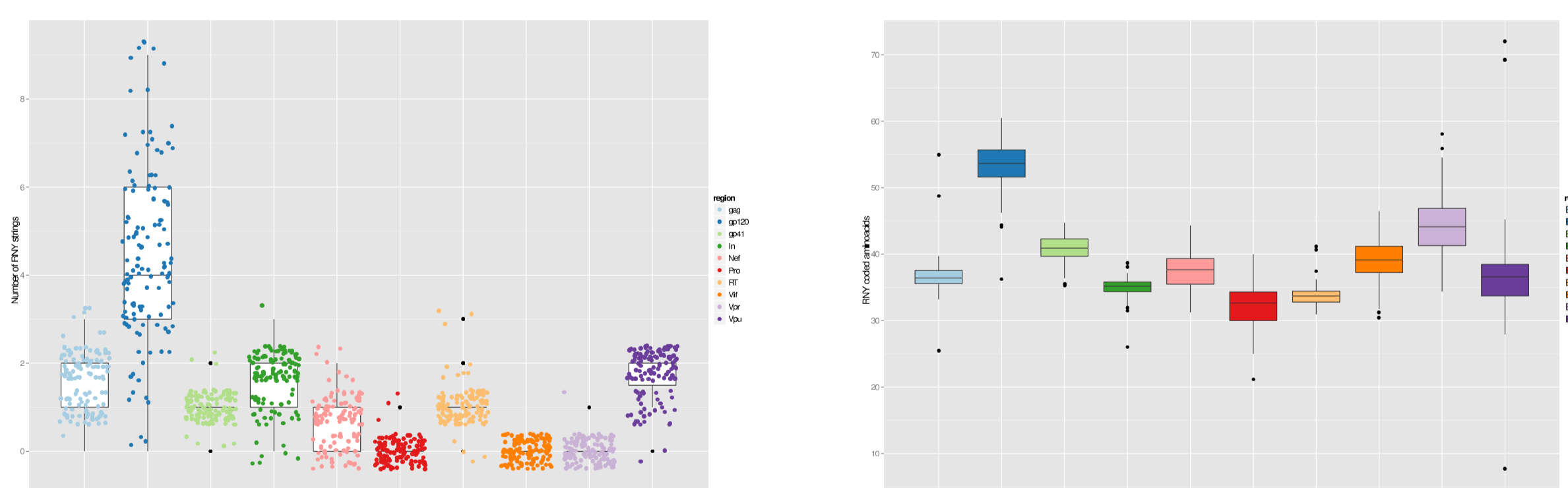


Fig.5: Number of RNY stretches (4 or more codons) along HIV genome (left panel). Percentage of RNY coded aminoacids in different HIV genes (right panel).

Recombination occurs among individual clones in the same patient PHI-test and RDP analysis were significant for all the patients analyzed. Analysis of clone alignments by RDP identified with certainty twelve breakpoints.

Acknowledgements: The authors wish to acknowledge the statistical support and advice of Dr. Antonio Di Narzo, Swiss Institute for Bioinformatics, University of Lausanne.

Contacts: Elisa De Crignis, Department of Medicine, Division of Allergy and Immunology, CHUV BT02-36, 1011 Lausanne-Switzerland. E-mail: elisa.decrignis@gmail.com

Conclusions

- Regions responsible for length variation in gp120 are characterized by the presence of an unusual distribution of RNY with respect to the rest of the molecule.
- Multiple base pair indels in gp120 may be generated by a mechanism of triplet repeat expansion. The onset of new of RNY codons in gp120 would be a random process, due to single base pair mutations caused by the low fidelity of RT and regulated by selection.
- Accumulation of RNY codons increases the probability of occurrence of insertions and deletions, and therefore can constitute a selective advantage for the virus.

Elisa De Crignis, Silvia Guglietta, Brian T. Foley, Matteo Negroni, Antonio Fabio Di Narzo, Vreneli Waelti Da Costa, Matthias Cavassini, Pierre-Alexandre Bart, Giuseppe Pantaleo, Cecilia Graziosi, *Non-random distribution of cryptic repeating triplets of purines and pyrimidines (RNY)_n in gp120 of HIV-1, AIDS Research and Human Retroviruses* - article in press