22nd ECCMID London, United Kingdom 31 March – 3 April 2012
ESCMID EUROPEAN SOCIETY OF CLINICAL MICROBIOLOGY AND INFECTIOUS DISEASES

Bioinformatic application to facilitate the genotypic determination of HIV-1 tropism.

Parra-Sánchez M*, Zakariya-Yousef I, Córdoba J, Ferrero I, Martín-Mazuelos E, Palomares JC.

Clinical Unit of Infectious Diseases and Microbiology. Valme University Hospital, Seville, Spain.

*email contact: manuel.parra.exts@juntadeandalucia.es

Servicio Andaluz de Salud CONSEJERÍA DE SALUD Área hospitalaria de Valme

P2129

Abstract

OBJETIVES: We have developed a bioinformatic tool in order to simplify the computer analysis of the genotypic study of HIV tropism by means of the V3 loop of the *gp120* protein. To get this, we have to analyze the FASTA sequence obtained after sequencing the V3 region by Geno2pheno (G2p) and/or WebPSSM (WP) algorithms individually. With this tool, we can obtain simultaneously the interpretation of both algorithms. Furthermore, with G2p analysis you can obtain results with two false positive ratios (% FPR). The first one is the analysis from MOTIVATE clinical data: (2 and 5.75% FPR). The second one is the Recommendations from the European Consensus Group (10%FPR).

METHODS: We have analyzed with G2p and WP a total of 653 protein sequences of V3 regions of HIV-1 obtained from Los Alamos database and sequences analyzed in the laboratory of Molecular Microbiology of the Valme University Hospital. These sequences correspond to 443 sequences with CCR5 (R5) genotype and 123 sequences with CXCR4 (X4) genotype (X4 or R5X4). 87 sequences with discordant results (9 sequences R5 by G2p but X4 by WP, and 78 sequences X4 by G2p but R5 by WP) were also included. These sequences belong to subtypes A (80 sequences, including A1 and A2 subtypes), B (501 sequences) and C (72 sequences) of HIV-1. (The next update will include approximately 1600 sequences with different subtypes of HIV-1 and HIV-2 and several CRFs)

Results: At running a new sequence, the result will show a % of similarity to any of the sequences of our database that will be used as reference sequence, and three tropism predictions of the model sequence: 1. the result obtained by analyzing the sequence of G2p with a FPR = 10%. 2. the result obtained by analyzing the sequence by G2P with a FPR = 2.5 and 5.75% and 3. the result obtained by analyzing the sequence with PSSM matrix using the "subtype B: X4/R5" (for C subtype we used the subtype C SI/NSI matrix). Besides, this application gets an automatically full expansion of your sequence. To test this application ten sequences randomly selected were used obtaining the same tropism interpretation in 9 cases.

CONCLUSION: 1.We oversimplified the methods for tropism analysis unifying the bioinformatics tools used for determining it. 2. We had obtained excellent results using this application, but it is necessary to increase the number of sequences in our database to optimize results and minimize discordant results that are generated after entering the sequence of study.

Background

Maraviroc (MVC, Celcentri ® Pfizer) is the first antiretroviral drug available which has as a target an essential cellular factor for the entrance of HIV; the C-C chemokine receptor type 5 (CCR5). The binding to this protein blocks the binding of the HIV-1 gp120 protein not allowing the entrance of the virus in the lymphocytes and monocytes. For this process there are two coreceptors (CXCR4 y CCR5) able to act as HIV coreceptors. The preferential use of CCR5 or CXCR4, known as viral tropism, is mainly determined by the aminoacid sequence of the V3 region, codified by the gp120 protein.

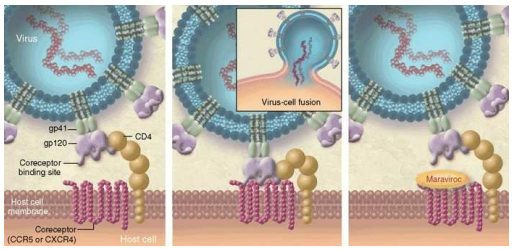

Figure1. Blocking of the HIV-1 entrance by the binding of Maraviroc to the coreceptor.

In this context, the viral tropism determination is an essential pre-requirement before using Maraviroc, so the only patients that could be treated with this drug would be those having a CCR5 tropism determination.

Nowadays, genotypic methods seem the fastest, easiest and cheapest option to determine the viral tropism. In order to do it, we have to interpret the FASTA sequence obtained after sequencing the V3 region. This process can be tedious if ,as the European consensus group and the Spanish group recommend, we must use the Geno2pheno and/or WebPSSM algorithms for the genotypic determination of the HIV tropism in clinical practice.

Aims

This tool allows us to obtain simultaneously the interpretation of both algorithms (Geno2pheno coreceptor and WebPSSM). Furthermore, with G2p analysis we can obtain results with two false positive ratios (% FPR). The first one is the analysis from MOTIVATE clinical data: (2 and 5.75% FPR). The second one is the Recommendations from the European Consensus Group (10%FPR).

Methods

In this new version of our database, we have analyzed with Geno2pheno coreceptor (G2p) and WebPSSM (PSSM) a total of 2793 protein sequences of V3 regions of HIV-1 obtained from Los Alamos database and sequences analyzed in the laboratory of Molecular Microbiology of Valme University Hospital. These sequences correspond to 1151 HIV-1 B subtypes and 1642 HIV-1 non-B subtypes, including circular recombinant forms (CRFs). Of them, 65.5% had a CCR5 genotype, 16% a non CCR5 genotype (CXCR4 or R5X4) and 14,5% had discordant results between G2p and PSSM.
Following updates will include more non-B subtypes HIV-1 and HIV-2 sequences analyzed by both methods.

Results

At running a new sequence, the result will show a % of similarity to any of the sequences in our database that will be used as a reference sequence, and three tropism predictions of the model sequence: 1. the result obtained by analyzing the sequence of G2p with a FPR = 10%. 2. the result obtained by analyzing the sequence by G2P with a FPR = 2.5 and 5.75% and 3. the result obtained by analyzing the sequence with PSSM matrix using the "subtype B: X4/R5" (for C subtype we used the subtype C SI/NSI matrix).

Besides, this application gets an automatically full expansion of your sequence.

R→ A o G	S→ G o C
M→ A o C	K→ G o T
W→ A o T	Y→ C o T

To test this application ten sequences randomly selected were used obtaining the same tropism interpretation in 9 cases.

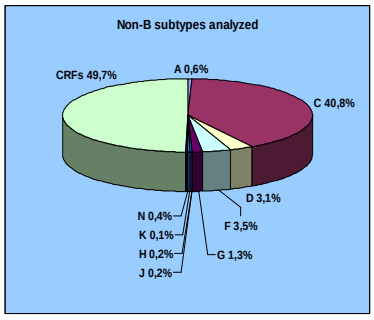

Figure 2.
HIV-1 non-B subtypes analyzed and included in the database

Figure 3.
Example of analysis of your sequence.

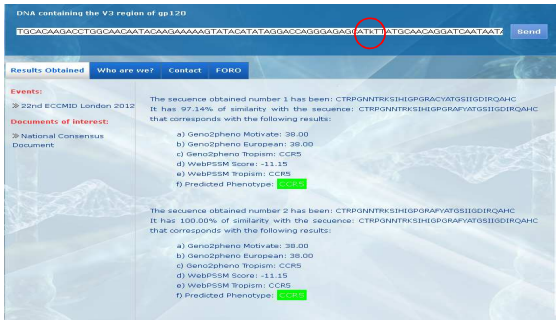

Figure 4.
Results obtained after expasion of the inserted degenerate sequence.

Conclusions

1. We oversimplified the methods for tropism genotyping determination unifying the bioinformatics tools used in order to determine it.
2. This tool allows us to expand the degenerate protein sequence obtained after sequencing so that we can analyze all possibilities by G2p (at this moment only offers an average score of the introduced sequence).
3. Mismacth results obtained by WebPSSM with a different length sequence (≠ 35aa**)** have been solved.
4. We have obtained excellent results using this application, but it is necessary to increase the number of sequences in our database to optimize results and minimize discordant results that are generated after entering the sequence of study.

Acknowledgements

ViiV Healthcare

For more information visit www.vih-tropismo.com or email to v3tropism@gmail.com