Rapid Diagnosis of HIV-1 virus by Near Infrared Spectroscopy: based on Partial least squares regression

Xiaohong He¹, Zhihong Song¹, Haifei Shang¹, Silang Yang¹, Lujing Wu¹, Yan Ma¹, Yu Sun¹, Xuejiao Xia¹, Quangui Wang^{2*}, Bei Luo^{3*}, and Chaojiang Gu^{1*}

¹. College of Life Science and Health, Wuhan University of Science and Technology, Wuhan, Hubei, 430065, China,

². College of Information Science and Engineering, Hunan University, Changsha, Hunan, 410082, China,

³. Wuhan Qiangui Technology Co., Ltd. Wuhan, Hubei, 430074, China.

Abstract: Currently, the laboratory diagnostic tests available for HIV-1 viral infection are mainly based on serological testing which relies on enzyme-linked immunosorbent assay (ELISA) for blood HIV antigen detection and reverse transcription polymerase chain reaction (RT-PCR) for HIV specific RNA sequence identification. However, these methods are expensive and time-consuming, and suffer from false positive and/or false negative results. Thus, there is an urgent need for developing a cost effective, rapid and accurate diagnostic method for HIV-1 infection. In order to reduce the barriers for effective diagnosis, a near-infrared spectroscopy (NIR) method was used to detect the HIV-1 virus in human serum, specifically, three absorption peaks with dose-dependent at 1582nm, 1810nm and 2363nm were found by multiple FBiPLSR test analysis for HIV-nano and HIV-EGFP, but not for MLV. Therefore, we recommend the use of 1582nm, 1810nm and 2363nm as the characteristic spectrum peak, for early screening and rapid diagnosis of serum HIV.

1 introduction :

The p24 antigen test and HIV-1 RNA RT-PCR are very effective means to detect HIV-1. HIV-1 nucleic acid test (NAT), in addition, is recommended for HIV-1 screening and diagnosis, especially for acute HIV-1 Infected and advanced AIDS patients. However, due to the cost and technical barriers, their use as general screening and diagnostic tools is limited[1]. The level of viral replication determined by the HIV-1 p24 capsid protein is quantified using the ELISA antigen capture assay, which is time consuming and produces false positive and/or negative results. Real-time reverse transcription polymerase chain reaction (RT-PCR) detection of HIV virus nucleic acid is currently the most reliable detection method, but it is also very time-consuming and has a high incidence of false negative results. In view of the fact that early diagnosis of HIV is essential for controlling infection and establishing early and effective antiviral therapy, there is an urgent need to develop a high-precision, low-cost, and easy-to-operate HIV-1 virus diagnosis method.

Near-infrared spectroscopy (NIRS) technology has been used for non-invasive detection of hemodynamic changes. It has the advantages of being real-time, continuous, low-cost and portable, and it has been used clinically for many years[2].

Near-infrared spectroscopy (NIRS) with multivariate statistical methods (such as principal component analysis (PCA)) has great potential in HIV diagnostic analysis.

Previous research has demonstrated that near infrared (NIRS) spectroscopy combined with partial least squares (PLS) regression can be used to detect HIV infection in plasma samples[3]. In addition, study showed that plasma Vis-NIR spectroscopy combined with principal component analysis (PCA) and analogous soft independent modeling (SIMCA) can distinguish samples from untreated HIV patients and uninfected controls[4]. However, there is no characteristic spectral peak of NIRS that can be directly used for HIV-1 diagnosis in their research[5], which brings great trouble to the promotion and application of near-infrared spectroscopy detection and analysis for HIV-1 diagnosis.

The focus of this research is to develop a method using NIR spectroscopy and chemometrics to screen known and potential HIV infections in serum. The pre-processed spectral data was divided into 100 segments, using partial least square regression (PLSR) algorithm to perform multiple linear fitting on each segment of spectral data, and further using forward and backward least square (FBiPLS) to screen for the best screening frequency bands and corresponding PLS principal component numbers. This method can quickly diagnose HIV-1 virus in serum and effectively distinguish it from MLV virus without use of any reagent.

* Corresponding author: Guchaojiang@wust.edu.cn; * Co-corresponding author: wdefu@yeah.net;

* Co-corresponding author: qgkeji@yeah.net

2 Materials and method

2.1 Plasmids

All the three plasmids used in our study including HIV-1 NL4-3-nano, HIV-1 NL4-3-EGFP and pNCS (Mo-MuLV)[6] were described in previous studies, and referred to as HIV-nano, HIV-GFP and MLV, respectively. The structure diagrams of HIV-nano, HIV-GFP and MLV are shown in Figure 1.

Fig 1. The genomes of HIV-1-NL4-3-nanoLuc, HIV-1-NL4-3-EGFP and MLV.

HIV-nano was constructed by introducing a luciferase gene (GB: AAA89084, from the vector pGL3 Basic) and an internal ribosomal entry site (IRES) into the Nef locus of HIV-1 NL4-3 backbone. The HIV-EGFP was constructed from HIV-1 NL4-3 by inserting an enhanced GFP (EGFP) gene and an internal ribosome entry site (IRES) sequence between gp41 and the nef sequence by PCR-based subcloning. The constructed plasmids were purified using Endo-Free Plasmid Maxi Kit (AxyGEN, USA) and verified through sequencing from Sangon Biotech (Shanghai, China).

2.2 Virus packaging and quantification

130.0~140.0x10⁶ 293T cells were first plated in two multi-layer cell culture flasks (Hyperflask) and cultured in DMEM complete medium in an incubator (37 °C, 5% CO₂) to reach ~60% cell confluence. Then, 960 µg PEI and 320 µg plasmid (HIV-nano/HIV-EGFP) were added to two centrifuge tubes with 16 mL DMEM basic medium (without FBS), respectively. Then, the DMEM basic medium mixed with PEI was blown with electric pipette, and the DMEM mixed with plasmid was added to PEI drop by drop and mixed with vortex. Finally, after incubating 20min at room temperature, the mixture of plasmid / PEI was added to the cell culture bottle and cultured in the 5%CO₂ incubator at 37 °C for 72 hours. The cell culture supernatants were collected and centrifuged at 4,000 x g for 30 min at room temperature to remove cells and cell debris. Then it was filtered through a 0.22 µm filter membrane and collected in a 250 mL centrifuge bottle, ultra-centrifuged at 4 °C, 30000 g for 2.5 h. The supernatant was removed by vacuum pump in a biological safety cabinet, and the virus pellets were resuspended in T cell medium, and aliquot into 20 µL/tube and 100 µL/tube.

Viral stocks concentration was normalized by measuring p24 content using ELISA (Lenti-X™ p24 Rapid Titer Kit, Takara, Japan).

The MLV virus was packaged and purified following the same protocol as we did for HIV-nano and HIV-EGFP described above. The collected MLV virus solution was quantified using quantitative PCR (qPCR) with primer detecting MLV gp80 gene[7]

HIV-nano, HIV-GFP and MLV viruses were diluted with human serum to 10⁶, 10⁵, 10⁴, 10³, 10², and 10 pg/ml, respectively, in the same way as normal saline. The concentration of each viruses was measured three times.

2.3 NIR spectra scanning

Near-infrared spectroscopy (NIRS) is a non-destructive and almost instantaneous technique that allows high throughput differentiation of biological samples[8]. The LAMBDA 750 is equipped with true double-beam and double-monochromator, designing to provide the highest possible stability and accuracy coupled with lowest stray-light performance.

All samples were allowed to equilibrate to room temperature (25°C) before NIR spectra scanning to ensure uniform sample temperature. The samples were measured for absorbances in the range of 780-2500 nm wavelength at 1 nm resolution, by a LAMBDA 750 UV/Vis/NIR (Perkin Elmer, USA) under standard laboratory conditions. In order to verify the effectiveness of the method, the same batch of samples were scanned for three days.

2.4 Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is a multivariate data analysis technique which generalizes and combines features from principal component analysis (PCA) and multiple linear regression (MLR)[9].

In order to improve the prediction accuracy, the raw spectrum data was first denoised using the background spectrum data and then smoothed by Moving Average (MA) data smoothing technology. First derivative and second derivative were processed for spectrum peak and pattern analysis.

The processed data going through noise reduction and smoothing was then performed with PLSR (Partial Least Squares Regression). By analyzing the RMSE (root mean square error) and R² (coefficient of determination) the fitting result of PLSR was assessed. FPLS (longitudinal spacing Partial Least Squares) was used for selection best fitting characteristic band (wavelength), for each of the HIV-nano and HIV-EGFP serum samples.

3 Results

The genomic sequence structures of HIV-1 viruses were constructed with reporter gene, named HIV-nano and HIV-EGFP respectively (figure 1). The cloned provirus DNA was used to transfect HEK293T cells to prepare HIV-nano and HIV-EGFP virus primordial seeds. The virus titer was determined by Lenti-X p24 rapid titration

kit (Takara, Japan)[10]. The results of virus titer are listed in Table 1.

Table 1 Characteristics of plasma samples examined in the present study.

Sample name	Estimated by ELISA [pg/ml]
HIV-1-NL4-nano	3.44×10^7
HIV-1-NL4-EGFP	4.05×10^7

200ul HIV-nano virus solution collected previously was accurately diluted with human serum to obtain a concentration of 1×10^6 pg/ml HIV-nano virus stock

solution. The stock solution was then serial diluted with serum to obtain standard viral concentration of 1×10^5 , 1×10^4 , 1×10^3 , 1×10^2 , 1×10^1 and 1×10^0 pg/ml. Same method was applied to prepare HIV-EGFP and MLV standard solutions.

Various factors might contribute to the pattern of spectrum, including chemical composition, chemical bond strength, reflection and scattering resulting from multi-compounds interactions. Due to the complexity of viral serum solution, it is necessary to preprocess the raw data before further analysis. In this experiment, the raw data was imported into OMNIC software to perform the Moving Average method for data smoothing and background noises reduction. The result of the smoothing process is shown in figure 2.

Figure 2. Near-infrared spectrum of HIV virus in serum.

Next, we use the partial least squares algorithm (PLS) to process the spectra. In PLS, the input spectrum was reduced by selecting the best principal component fractions, following a fitting by Multivariate Linear Fitting. RMSE (Root Mean Square Error) and R^2 (Coefficient of Determination) were used as parameters for evaluation of better fitting, as we were aiming a smaller RMSE and R^2 closer to 1.

Finally, the optimal feature band is filtered by forward and backward partial least squares (FBiPLS) filtering [11]. According to the joint xy distance, the sample set was divided into 100 segments, and PLS fitting is performed to obtain 100 local regression models. Forward interval partial least squares method FiPLS and backward interval partial least squares method BiPLS were performed respectively for each of the 100 segments, and the first

selected model band selected by FiPLS to combine the remaining bands one by one, and select the band with the smallest RMSE in the combined band as the second selected Model, while using BiPLS to remove one segment, until there is one segment or none at all.

Through multiple FBiPLSR test analysis, HIV-nano and HIV-EGFP were found to have common peaks at 1582nm, 1810nm and 2363nm, while MLV virus and serum have no obvious peaks in these wavelengths. Therefore, these three wavelengths are recommended as the characteristic wavelengths of the HIV virus. Since the HIV virus has a low absorption at 1200~1300nm and weakly interferes with water molecules, 1250nm is chosen as its reference wavelength. Figure 3 shows the characteristic peaks identified by analysis.

Fig 3: The three best characteristic bands of HIV-nano and HIV-EGFP except MLV virus.

4 Discussion

Related studies in Japan has shown the application of NIR spectrum in the diagnosis of HIV-1 virus[3, 4]. In their study, the results obtained by the NIR spectroscopic model exhibited a good correlation with those obtained with the

reference method (HIV-1 p24 ELISA). However, there was no obvious characteristic spectrum peak of NIR that could be directly applied in the diagnosis of HIV-1, which caused great trouble for the promotion and application of HIV-1 near-infrared spectroscopy.

The HIV genome has approximately 9,800 bases and is located in the middle of the virus particle. Its genome

contains 9 genes, and encodes 15 proteins. Since the protein contains multiple CN bonds, C-C bonds, C-O bonds, OH---OH bonds, NH---O=C bonds, and OH---O=C bonds that can be absorbed by near-infrared light, in theory, HIV virus has more than 80 near-infrared characteristic peaks, and the experimental data also verifies this inference and judgment[12]. In order to achieve rapid screening and diagnosis of HIV virus, it is necessary to select 3-5 typical characteristic peaks.

HIV-nano and HIV-EGFP are the two progeny of the B subtype HIV international virus strain, and their coded viral proteins are exactly the same. They have a common peak at 1582, 1810, and 2363nm. MLV has no obvious peak in this waveband, so It can be considered that 1582, 1810, and 2363nm are the characteristic peaks of HIV international strain B subtype. At the same time, this characteristic peak can detect virus particles at 1 pg/ml with high sensitivity. After four repeated experiments, the test sample concentration covered $1-1 \times 10^6$ pg/ml, and a total of 147 samples were verified. These samples all have a common peak at 1582, 1810, and 2363nm, which once again proves the reliability of the virus test and analysis method.

Partial least squares regression (PLSR)[13] is a multivariate data analysis technique which generalizes and combines features from principal component analysis (PCA) and multiple linear regression (MLR). Both it and the principal component analysis method try to extract the maximum information reflecting the variation of the data, but the principal component analysis method only considers a matrix of independent variables, while the partial least square method also has a "response" matrix, so it has a predictive function.

We have shown that HIV-1 virus in serum may possess their own unique spectrum from NIRS measurement. Moreover, when we establish the best mathematical model based on the HIV-1 virus characteristic spectrum, NIRS only needs to measure a few limited wavelengths, the speed will be further improved, and no reagents are needed for analysis[14]. Overall, this study demonstrates Near infrared spectroscopy (NIRS) as a new diagnostic tool with the potential to develop into a quicker detection of HIV virus.

Fig 4: The data algorithm model diagram of HIV-1 virus detection by near infrared spectroscopy.

5 Conclusions

This study is limited by a few factors. For one, all samples are derived from the classic virus strain HIV-1-NL4-3, which are diluted with water, saline or serum, and cannot completely mimic the state of human body infected with HIV virus. Second, larger sample size is needed to validate the preliminary findings of this study in future research. third, the variability of human immunodeficiency virus type 1 (HIV-1) has given rise to multiple subtypes and recombinant strains. The characteristic peaks of HIV-1 NIRS of different subtypes may be slightly different, which is another focal point for expansion into future research.

We hope that these researches will provide new perspectives for a better understanding of diagnose HIV-1 virus by NIR spectrum; a larger follow-up researches will help to gain a deeper understanding of the potential role of NIR spectrum as a credible technology for diagnosis AIDS patients. Blood-based NIR spectrum tests may aid clinical assessments for the effective and accurate differential diagnosis of HIV, decrease the labor, time and cost of diagnosis. More effort should be focused on the sensitivity and specificity of NIRS in differential diagnosis of different subtypes of HIV in clinical diagnosis[15].

Acknowledgments

This work was supported by the National Science and Technology Major Project (2017ZX10202102-007 and 61972299) and by grants from the Applied Basic Research Project of Wuhan City (No. 2017060201010193) and Hubei Provincial Plan of Natural Science Foundation of China (No. ZRMS2019000926) at Wuhan University of Science and Technology. Finally, we would like to thank all the authors for their contributions in this Research Topic.

References:

1. Kong, W., Liu, P., Tang, L., Zhu, Z., *et al.*, Estimation of the Seroconversion Duration of HIV-1 Antibodies in Individuals With Recent Infection in China. *FRONT MICROBIOL* 2019, *10*.
2. von Lümann, A., Li, X., Müller, K., Boas, D. A., Yücel, M. A., Improved physiological noise regression in fNIRS: A multimodal extension of the General Linear Model using temporally embedded Canonical Correlation Analysis. *NeuroImage (Orlando, Fla.)* 2020, *208*, 116472.
3. Sakudo, A., Tsenkova, R., Onozuka, T., Morita, K., *et al.*, A Novel Diagnostic Method for Human Immunodeficiency Virus Type-1 in Plasma by Near-Infrared Spectroscopy. *MICROBIOL IMMUNOL* 2005, *49*, 695-701.
4. Bahmani, M. K., Khosravi, A., Miri, R., Iwabu, Y., *et al.*, Spectroscopic characterization of human immunodeficiency virus type-1-infected plasma by principal component analysis and soft independent

- modeling of class analogy of visible and near-infrared spectra. *MOL MED REP* 2009, *2*, 805-809.
5. Sakudo, A., Suganuma, Y., Sakima, R., Ikuta, K., Diagnosis of HIV-1 infection by near-infrared spectroscopy: Analysis using molecular clones of various HIV-1 subtypes. *CLIN CHIM ACTA* 2012, *413*, 467-472.
6. Sindberg, G. M., Sindberg, G. M., Sharma, U., Sharma, U., *et al.*, An Infectious Murine Model for Studying the Systemic Effects of Opioids on Early HIV Pathogenesis in the Gut. *J NEUROIMMUNE PHARM* 2015, *10*, 74-87.
7. Gu, C. J., Borjabad, A., Hadas, E., Kelschenbach, J., *et al.*, EcoHIV infection of mice establishes latent viral reservoirs in T cells and active viral reservoirs in macrophages that are sufficient for induction of neurocognitive impairment. *PLOS PATHOG* 2018, *14*, e1007061.
8. Volino-Souza, M., de Oliveira, G. V., Barros-Santos, E., Pinheiro, V., *et al.*, Near-infrared spectroscopy-derived muscle oxygen saturation during exercise recovery and flow-mediated dilation are impaired in HIV-infected patients. *MICROVASC RES* 2020, *130*, 104004.
9. Han, D., Tan, J., Men, J., Li, C., Zhang, X., Quantitative Structure Activity/Pharmacokinetics Relationship Studies of HIV-1 Protease Inhibitors Using Three Modelling Methods. *MED CHEM* 2021, *17*, 396-406.
10. Meshesha, M., Esadze, A., Cui, J., Churgulia, N., *et al.*, Deficient uracil base excision repair leads to persistent dUMP in HIV proviruses during infection of monocytes and macrophages. *PLOS ONE* 2020, *15*, e235012.
11. Hitchman, S., Johnson, P., Bain, W., Craigie, C. R., Reis, M. M., Short communication: Long term performance of near infrared spectroscopy to predict intramuscular fat content in New Zealand lamb. *MEAT SCI* 2020, 108376.
12. Tarancon-Diez, L., Rodríguez-Gallego, E., Rull, A., Peraire, J., *et al.*, Immunometabolism is a key factor for the persistent spontaneous elite control of HIV-1 infection. *EBIOMEDICINE* 2019, *42*, 86-96.
13. Rafiq, S., Majeed, M. I., Nawaz, H., Rashid, N., *et al.*, Surface-enhanced Raman spectroscopy for analysis of PCR products of viral RNA of hepatitis C patients. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* 2021, *259*, 119908.
14. Gamage, S., Howard, M., Makita, H., Cross, B., *et al.*, Probing structural changes in single enveloped virus particles using nano-infrared spectroscopic imaging. *PLOS ONE* 2018, *13*, e199112.
15. Alibayeva, K. O., Saparbekov, M. K., Baiserkin, B. S., Abishev, A. T., *et al.*, <p>Study of the possibility of introduction of Kazakhstan NGO-based rapid HIV testing procedures</p>. *HIV/AIDS (Auckland, N.Z.)* 2019, *Volume 11*, 219-227.