331

IrsiCaixa
Institut de Recerca de la Sida

22nd CROI
Seattle
February 23-26, 2015

FLS
FUNDACIÓ LLUITA CONTRA LA SIDA
www.flsida.org

A random forest approach to define immunological thresholds for CD4 T-cell recovery in HIV-treated individuals

Josué Pérez-Santiago[1], Marta Massanella[1], Dan Ouchi[2], Elisabet Gomez[2], Cecilia Cabrera[2], Bonaventura Clotet[2,3,4], Eugenia Negredo[3,4], Julià Blanco[2,4],

[1]University of California San Diego (UCSD), La Jolla, CA, US; [2]IrsiCaixa -AIDS Research Institute, Badalona, Spain; [3]Fundació Lluita contra la SIDA, Badalona, Spain; [4]UVIC-UCC (Universitat de Vic- Universitat Central de Catalunya), Vic, Spain

UNIVERSITY OF CALIFORNIA SAN DIEGO

Julià Blanco
e-mail: jblanco@irsicaixa.es
Marta Massanella
e-mail: mmassanellaluna@ucsd.edu
Josué Pérez-Santiago
e-mail: joperez@ucsd.edu

BACKGROUND

► Although ART reverts HIV-induced immune damage, some HIV-infected patients achieving viral suppression present unsatisfactory CD4 T-cell recovery (discordant subjects) that results in higher risk of clinical progression or death[1].

► The failure to increase CD4 T-cell counts in some ART-suppressed patients has been related to low CD4 T-cell production, high activation and cell death[2-4].

► In our previous studies of discordant patients, we identified CD4 T-cell hyperactivation and increased CD4 T-cell death (intrinsic pathway) as the determinant mechanisms for the unsatisfactory immune recovery[4,5].

► While some clinical data indicate that low CD4 threshold (i.e. 200 CD4 cells/µL) may have strong impact on survival, the lack of an accepted standard definition of immune recovery hamper a proper clinical follow up and has led to discrepant results among different studies[6].

OBJECTIVE

To use a supervised machine learning approach to evaluate different clinical definitions of immune recovery.

METHODS

► Subjects: A cross-sectional, descriptive, comparative study was performed to analyze the immune status of patients with paradoxical immune response. The Institutional Review Board approved the study (EO code: EO-07-024). We included 230 patients on stably suppressive ART (Viral Load <50 copies/mL) for at least 2 years.

► Immune Activation Markers: CD4 T-cell subsets and the expression of activation markers in blood samples was analyzed by six color flow cytometry using the following antibody combinations:

Activation markers: CD95-FITC, PD-1-PE, HLA-DR-PerCP, CD3-APC-Cy7, CD4-APC and CD8-PE-Cy7

HLA-DR-FITC, CD38-PerCP-Cy5.5, CD45RO-APC, CD3-APC-Cy7 and CD8-PE-Cy7

Naive/memory subsets: CD45RA-FITC, CD31-PE, CD38-PerC-Cy5.5,CD3-APC-Cy7, CD4-APC and CD8-PE-Cy7

In addition, for a subset of 50 individuals, frozen PBMCs were used to analyze Treg and cell-proliferation by eight-color flow cytometry using the following antibody combination:

Ki67-FITC, FOXP3-PE, CD25-PE-Cy7,CD5-PerC-Cy5.5, CD127-Alexa fluor 647, CD3-APC-Cy7, CD4-V450 and CD8-V500

► Cell death was measured in cultured fresh PBMC at day 1 by simultaneous staining of CD3, CD4, CD8 antigens, mitochondrial membrane potential (DIOC6) and plasma membrane permeability (Propidium Iodide).

► Plasma levels of soluble CD14, a surrogate marker of bacterial translocation, were quantified by ELISA.

► Random Forest: Participants were divided into 2 groups (discordant vs. concordant) based in either the CD4 counts at time of visit (using cut-off values ranging from 200 to 600 cells/µL) or the change in CD4 counts from CD4 Nadir (ΔCD4, using values ranging from 50 to 500 cell/µL). Patients below or above the indicated values were considered as **Discordant** (Unfavorable immunologic response) or **Concordant** (Favorable immunologic response), respectively. **Random forest** classification was performed using all socio-demographical, clinical and immunological variables collected as predictors using R statistical software (package randomForest). Validation, classification performance and over fitting assessments were evaluated using a 10-fold cross validation. The importance of our predictors for classification was measured using the mean decrease in Gini index.

► Hierarchical Clustering and heat maps: Using the variables with the highest contribution to samples discrimination (Gini's mean >1.2), we performed an unsupervised hierarchical clustering algorithms to visualize specific patterns in our study groups also using R software. Each variable was centered with respect to the mean and scaled by the standard deviation. The distance between samples and variables was measured by means of euclidean distance, and hierarchical clustering was performed using the ward method. Colors in Heatmap represents the values of the variables rangins from minimum values (green/blue) to maximum values (red/yellow).

RESULTS



Figure 1. Study flow chart

Table 1. Patient characteristics

	All subjects	CD4 T-cell counts (cells/µl)					
	n = 197	<200[1] n = 19	201-350[2] n = 64	351-500[3] n = 33	501-750[4] n = 45	>750[5] n = 36	Anova
Age, years , median [IQR]	45 [42-51]	48 [44-51]	47 [44-54]	46 [41-52]	45 [42-51]	44 [38-48]	*0.02*
Male, n (%)	150 (76)	16 (84)	51 (80)	26 (79)	30 (67)	27 (75)	*ns*
Time since HIV diagnosis, year, median [IQR]	13 [8-17]	14 [5-19]	12 [8-19]	13 [10-17]	13 [10-17]	12 [9-17]	*ns*
Time on ART, year, median [IQR]	11 [7-14]	8 [4-11]	9 [5-13]	11 [8-15]	12 [9-14]	11 [7-14]	*0.03*
Nadir CD4 T-cell count, cells/µl	148 [65-272]	47 [16-87]	104 [37-167]	170 [87-242]	239 [80-322]	293 [197-436]	*<0.0001*
Nadir CD4 T-cell count < 200 cells/µl, n(%)	121(61)	19 (100)	53 (83)	19 (58)	20 (44)	10 (28)	*<0.0001*
CD4 T-cells, %	26 [19-33]	14 [10-21]	19 [16-23]	25 [21-30]	34 [26-34]	38 [32-42]	*<0.0001*
Absolute CD4 T-cell count, cells/µl	444 [284-683]	162 [137-186]	286 [237-320]	445 [406-480]	632 [584-695]	909 [799-1084]	*<0.0001*
CD8 T-cells, %	46 [38-53]	55 [48-66]	50 [44-56]	47 [41-52]	41 [34-47]	37 [32-44]	*<0.0001*
Absolute CD8 T-cell count, cells/µl	800 [589-1067]	521 [284-800]	726 [589-949]	804 [542-1070]	809 [636-1105]	973 [736-1326]	*0.0003*
Ratio CD4/CD8	0.5 [0.4-0.9]	0.3 [0.2-0.4]	0.4 [0.3-05]	0.6 [0.4-0.8]	0.8 [0.6-0.9]	1.1 [0.8-1.2]	*<0.0001*
Current ART, n (%)							
PI	89 (45)	12 (63)	34 (53)	10 (30)	19 (42)	14 (39)	*ns*
NNRTI	107 (54)	7 (37)	30 (47)	23 (70)	25 (56)	22 (61)	*ns*
HCV coinfection, %	71 (36)	7 (37)	27 (42)	10 (30)	17 (38)	10 (28)	*ns*
HBV coinfection, %	9 (5)	1 (5)	3 (5)	2 (6)	1 (2)	2 (6)	*ns*

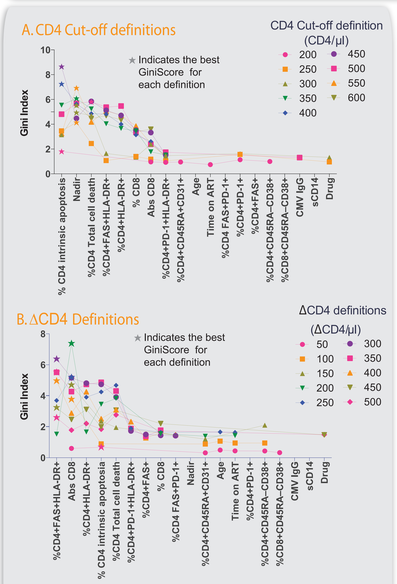


Figure 3. Variables that could contribute to immunodiscordance determined by random forest for (A) CD4 Cut-off and (B) ΔCD4 definitions

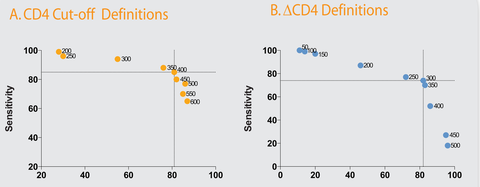
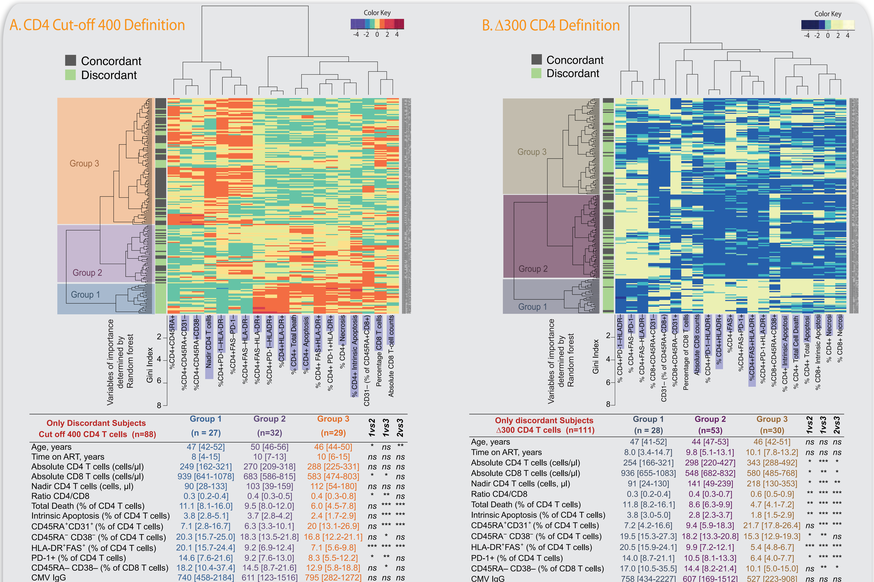


Figure 2. Random forest classification performance: specificity and sensitivity were calculated for each CD4 cutoff (A) and ΔCD4 (B) definitions



Only Discordant Subjects Cut off 400 CD4 T cells (n=88)	Group 1 (n = 27)	Group 2 (n=32)	Group 3 (n=29)	1vs2	1vs3	2vs3
Age, years	47 [42-52]	50 [46-56]	46 [44-50]	*	*ns*	**
Time on ART, years	8 [4-15]	10 [7-13]	10 [6-15]	*ns*	*ns*	*ns*
Absolute CD4 T cells (cells/µl)	249 [162-321]	270 [209-318]	288 [225-331]	*ns*	*ns*	*ns*
Absolute CD8 T cells (cells/µl)	939 [641-1078]	683 [586-815]	583 [474-803]	*	*	*ns*
Nadir CD4 T cells (cells, µl)	90 [28-133]	103 [39-159]	112 [54-180]	*ns*	*ns*	*ns*
Ratio CD4/CD8	0.3 [0.2-0.4]	0.4 [0.3-0.5]	0.4 [0.3-0.8]	*	**	*ns*
Total Death (% of CD4 T cells)	11.1 [8.1-16.0]	9.5 [8.0-12.0]	6.0 [4.5-7.8]	*ns*	***	***
Intrinsic Apoptosis (% of CD4 T cells)	3.8 [2.8-5.1]	3.7 [2.8-4.2]	2.4 [1.7-2.9]	*ns*	***	***
CD45RA$^+$CD31$^+$ (% of CD4 T cells)	7.1 [2.8-16.7]	6.3 [3.3-10.1]	20 [13.1-26.9]	*ns*	**	***
CD45RA$^-$ CD38$^-$ (% of CD4 T cells)	20.3 [15.7-25.0]	18.3 [13.5-21.8]	16.8 [12.2-21.1]	*	*	*ns*
HLA-DR$^+$FAS$^+$ (% of CD4 T cells)	20.1 [15.7-24.4]	9.2 [6.9-12.4]	7.1 [5.6-9.8]	***	***	***
PD-1+ (% of CD4 T cells)	14.6 [7.6-21.6]	9.2 [7.6-13.0]	8.3 [5.5-12.2]	*	**	*ns*
CD45RA– CD38– (% of CD8 T cells)	18.2 [10.4-37.4]	14.5 [8.7-21.6]	12.9 [5.8–18.8]	*ns*	*	*ns*
CMV IgG	740 [458-2184]	611 [123-1516]	795 [282-1272]	*ns*	*ns*	*ns*
sCD14, ng/ml	8.6 [7.7-10.4]	8.9 [7.6-10.1]	8.6 [7.7-10.2]	*ns*	*ns*	*ns*

Only discordant Subjects Δ300 CD4 T cells (n=111)	Group 1 (n = 28)	Group 2 (n=53)	Group 3 (n=30)	1vs2	1vs3	2vs3
Age, years	47 [41-52]	44 [47-53]	46 [42-51]	*ns*	*ns*	*ns*
Time on ART, years	8.0 [3.4-14.7]	9.8 [5.1-13.1]	10.1 [7.8-13.2]	*ns*	*ns*	*ns*
Absolute CD4 T cells (cells/µl)	254 [166-321]	298 [220-427]	343 [288-492]	*	***	*
Absolute CD8 T cells (cells/µl)	936 [655-1083]	548 [682-832]	580 [485-768]	*	***	*
Nadir CD4 T cells (cells, µl)	91 [24-130]	141 [49-239]	218 [130-353]	*	***	***
Ratio CD4/CD8	0.3 [0.2-0.4]	0.4 [0.3-0.7]	0.6 [0.5-0.9]	***	***	***
Total Death (% of CD4 T cells)	11.8 [8.2-16.1]	8.6 [6.3-9.9]	4.7 [4.1-7.2]	*	***	***
Intrinsic Apoptosis (% of CD4 T cells)	3.8 [3.0-5.0]	2.8 [2.3-3.7]	1.8 [1.5-2.9]	**	***	***
CD45RA$^+$CD31$^+$ (% of CD4 T cells)	7.2 [4.2-16.6]	9.4 [5.9-18.3]	21.7 [17.8-26.4]	*ns*	***	***
CD45RA$^-$ CD38$^-$ (% of CD4 T cells)	19.5 [15.3-27.3]	18.2 [13.3-20.8]	15.3 [12.9-19.3]	*	**	*ns*
HLA-DR$^+$FAS$^+$ (% of CD4 T cells)	20.5 [15.9-24.1]	9.9 [7.2-12.1]	5.4 [4.8-6.7]	***	***	***
PD-1+ (% of CD4 T cells)	14.0 [8.7-21.1]	10.5 [8.1-13.3]	6.4 [4.0-7.7]	*	***	***
CD45RA– CD38– (% of CD8 T cells)	17.0 [10.5-35.5]	14.4 [8.2-21.4]	10.1 [5.0-15.0]	*ns*	**	*
CMV IgG	758 [434-2227]	607 [169-1512]	527 [223-908]	*ns*	*ns*	*ns*
sCD14, ng/ml	8.8 [7.8-10.4]	8.4 [7.4-9.9]	8.0 [6.6-9.1]	*ns*	*	*ns*

Figure 4. Hierarchical Clustering for variables with highest Gini Index as determined by random forest for (A) Cut-off 400 CD4 T-cells and (B) Δ300 definitions. In both cases, three main groups of individuals were identified and the main characteristics of Discordant subjects of each group are shown. These data suggest the existence of different mechanisms of discordance.

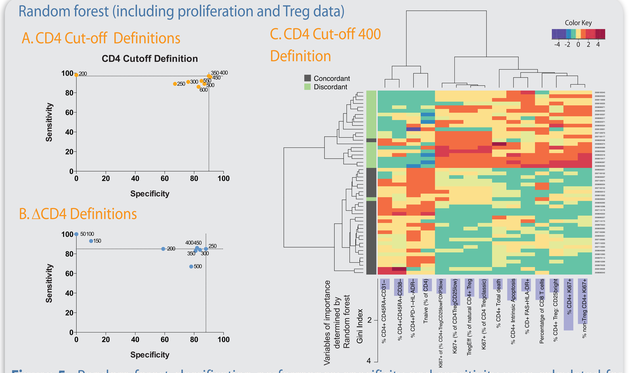


Figure 5. Random forest classification performance: specificity and sensitivity were calculated for each CD4 cutoff (A) and ΔCD4 (B) definitions and Hierarchical clustering for cut-off 400 definition (C)

CONCLUSIONS

► The advantage of using random forest for classification is the inclusion of a high number of immunological and demographical variables as predictors without overfitting, enabling the successful segregation of individuals with good vs. poor immune recovery.

► Using random forest, the 400 CD4 T-cells/µl cut-off definition classified better discordant and concordant patients than ΔCD4 classifications.

► As expected cut-off classification was strongly associated with destruction and activation of CD4 T cells. However, the gain of CD4 T cells is strikingly associated to CD8 absolute counts, corroborating the use of CD4/CD8 ratio as a measure of immune recovery.

► In a substudy, proliferation of Treg and non-Treg CD4 T cells in addition to CD4 T cell death parameter were able to improve the segregation of Concordant and Discordant classes.

► The cluster analysis identified three subclasses of immunodiscordants showing different immune characteristics. Therefore, different interventions might be required to increase CD4 T cells in immunodiscordant individuals.

REFERENCES

1 Gazzola L *et al*, Clin Inf Dis 2009;48:328-37
2 Lederman M *et al*, JID 2011; vol 204 (8): 1217-26
3 Hunt P et al, JID 2008; 197 (1):126-33
4 Massanella M and Negredo, *et al*. AIDS 2010;24(7):959-68
5 Negredo E and Massanella M, *et al*. CID 2010; 50(9): 1300-8
6 COHERE. PLoS Med. 2012; 9(3), e1001194

ACKNOWLEDGEMENTS

This work was supported by the Spanish AIDS network "Red Temática Cooperativa de Investigación en SIDA" (RD12/0017). J.B. is a researcher from Fundació Institut de Recerca en Ciències de la Salut Germans Trias i Pujol supported by the ISCIII and the Health Department of the Catalan Government (Generalitat de Catalunya).