Original article

Investigation of expert rule bases, logistic regression, and non-linear machine learning techniques for predicting response to antiretroviral treatment

Mattia CF Prosperi, Andre Altmann, Michal Rosen-Zvi, Ehud Aharoni, Gabor Borgulya, Fulop Bazso, Anders Sönnerborg, Eugen Schülter, Daniel Struck, Giovanni Ulivi, Anne-Mieke Vandamme, Jurgen Vercauteren and Maurizio Zazzi on behalf of the EuResist and Virolab study groups

Antiviral Therapy **14:**433–442

Supplementary material

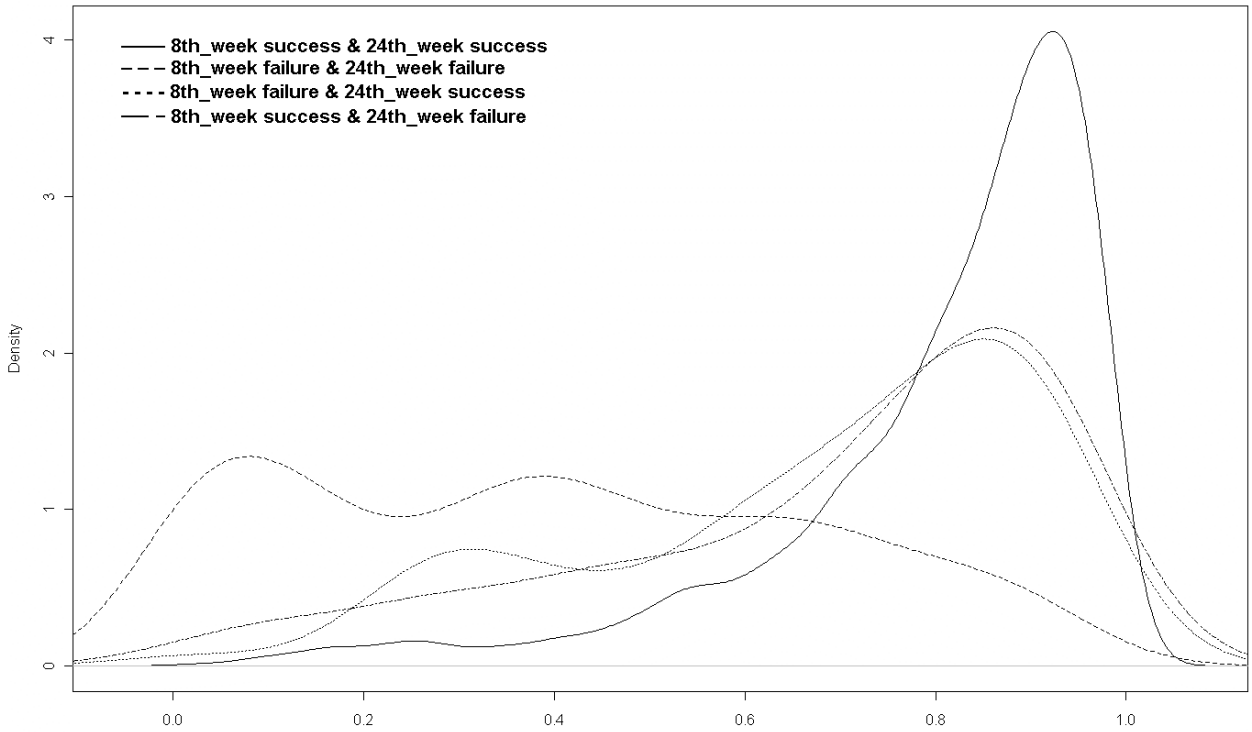
Methods and results for 8-week viral load change regression

We also investigated the problem of actual viral load regression. The study design and variable coding was the same as for the 8-week response classification, except that we substituted the binary outcome variable with the 8th week follow up viral load change from baseline (ΔLog_{10} copies/ml, original values cut to [2.69, 8]). The input space was the same as the one used for the Logistic Regression model (vi), but we replaced LR with Multiple Linear Regression. As a feature selection technique we used the univariable filter based on chi-square test, plus the M5 method (Quinlan, JR. Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence*, Singapore, 343–348, 1992). Regression results in 10-fold CV yielded a correlation coefficient ρ of 0.7525, Root Mean Squared Error (RMSE) of 0.6805, and Mean Absolute Error (MAE) of 0.4954. Evaluation on the test set was concordant, yielding ρ=0.7628, RMSE=0.7006, MAE=0.5062. Results are lower than the model described in the previous study of *Larder et al* [14] where Artificial Neural Network modelling yielded an extra sample correlation ρ=0.831, using a similar TCE selection policy. However, direct comparison of methods is not possible due to different data sets used.

Additional plots for medium-term prediction

When analysing performances of the LR model (vi) in predicting medium-term outcomes, using cross-validated predictions and independent test set analysis, we found that the density estimation for outcome distributions varied significantly when stratifying by outcome cross-tabulations. In particular, distributions of predictions for incongruent outcomes (i.e., success at 8 weeks and failure at 24 weeks or vice-versa) differ significantly from distributions obtained for concordant outcomes (i.e., successes both at 8 and 24 weeks or failures at both 8 and 24 weeks), as depicted in Figure A1.

Figure A1: *Density estimation plots of CV predictions for medium-term and short-term outcome stratification*



Sensitivity analysis for success definition using HIV-1 RNA below 50 copies/ml

We tested performance of LR model (vi) changing the definition of 8-week success from the 500 HIV-1 RNA copies/ml cut-off to the 50 HIV-1 RNA copies/ml cut-off, leaving all the other definitions for ESD extraction as they are. LR model (vi) was trained using the 500 copies/ml cut-off and then we analysed CV predictions against the new success definition. This may underestimate the performances of an engine trained using the new cut-off, but we did not investigate theoretical methods to account for data truncated at different levels. However, such a performance evaluation may be a starting point for further refinements of our model. In detail, changing the success threshold, 900 (31.8%) instances were both failures using a 50 or 500 copies/ml cut-off, 437 (15.4%) were failures at 50 and successes at a 500 copies/ml cut-off, and 1494 (52.8%) were both successes at 50 and 500 copies/ml cut-off. We observed a relevant loss in accuracy (decreasing to 66% of correct classifications) and AUC (decreasing to 0.69), as depicted in Figure A2. Upon analysing the reasons for such a decrease in more depth, we observed that the main loss is in specificity, due to the fact that many of the instances classified as successes using the 500 copies/ml threshold are not anymore using the new 50 copies/ml threshold (Figure A3). However, when analysing the outcome distribution stratifying by concordant/discordant outcomes (i.e., success or failure both at 500 and 50 copies/ml, success at 50 copies/ml, and failure at 500 copies/ml cut-off), we found that indeed the distribution of probability values for discordant stratum (failure/success) is significantly different ($p<0.03$, Wilcoxon rank-sum test) from the corresponding concordant stratum (success/success), suggesting that the model is able to recognise such TCEs as “less successful” (Figure A4).

Figure A2: *ROC comparison for LR model (vi) performances changing the 8-week success cut-off at HIV-1 RNA <50 copies/ml*

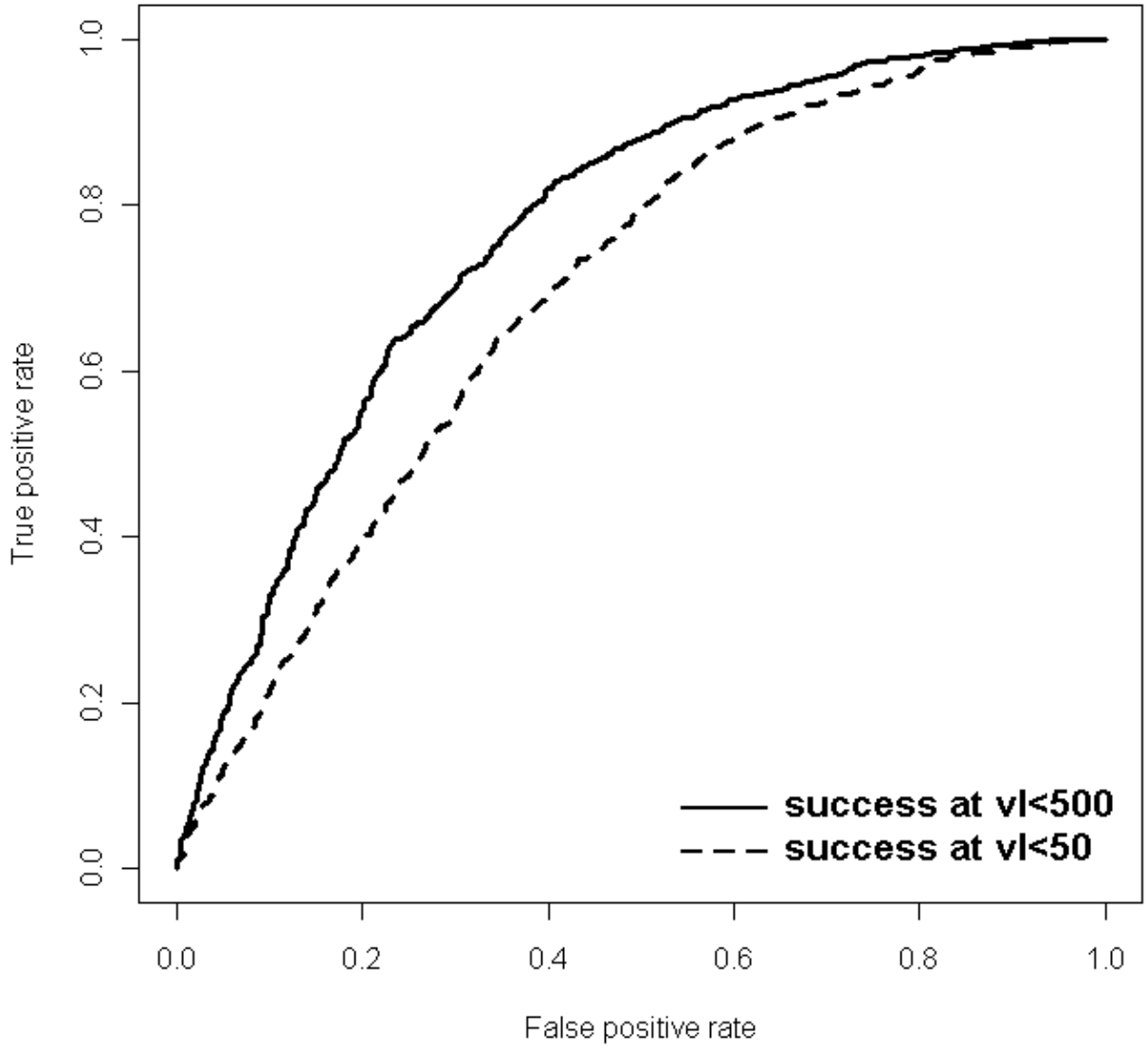


Figure A3: *Sensitivity/specificity comparison for LR model (vi) performances changing the 8-week success cut-off at HIV-1 RNA <50 copies/ml*

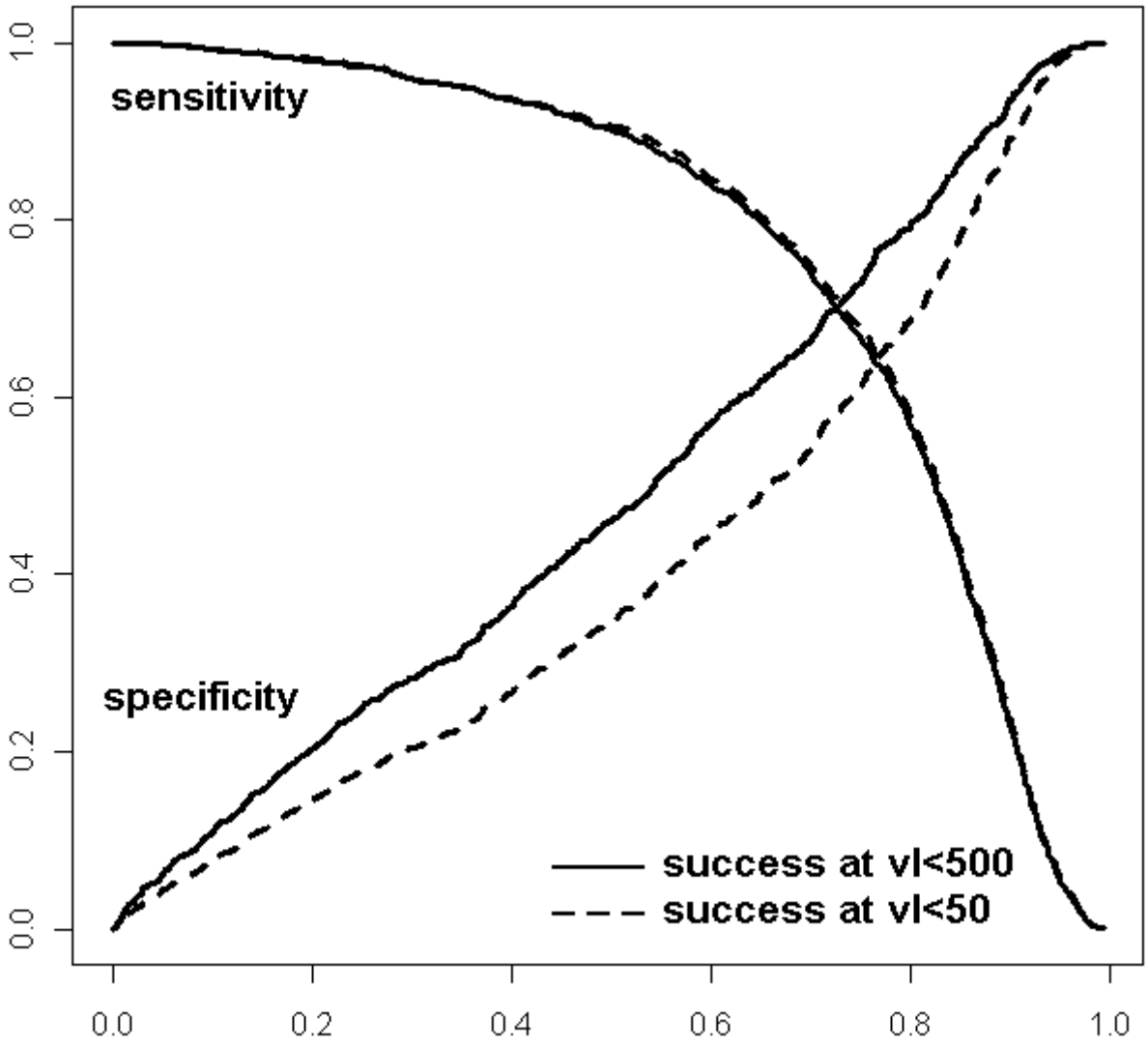


Figure A4: *Density estimation of LR model (vi) predicted success probability stratifying by cut-off outcome cross-tabulations*

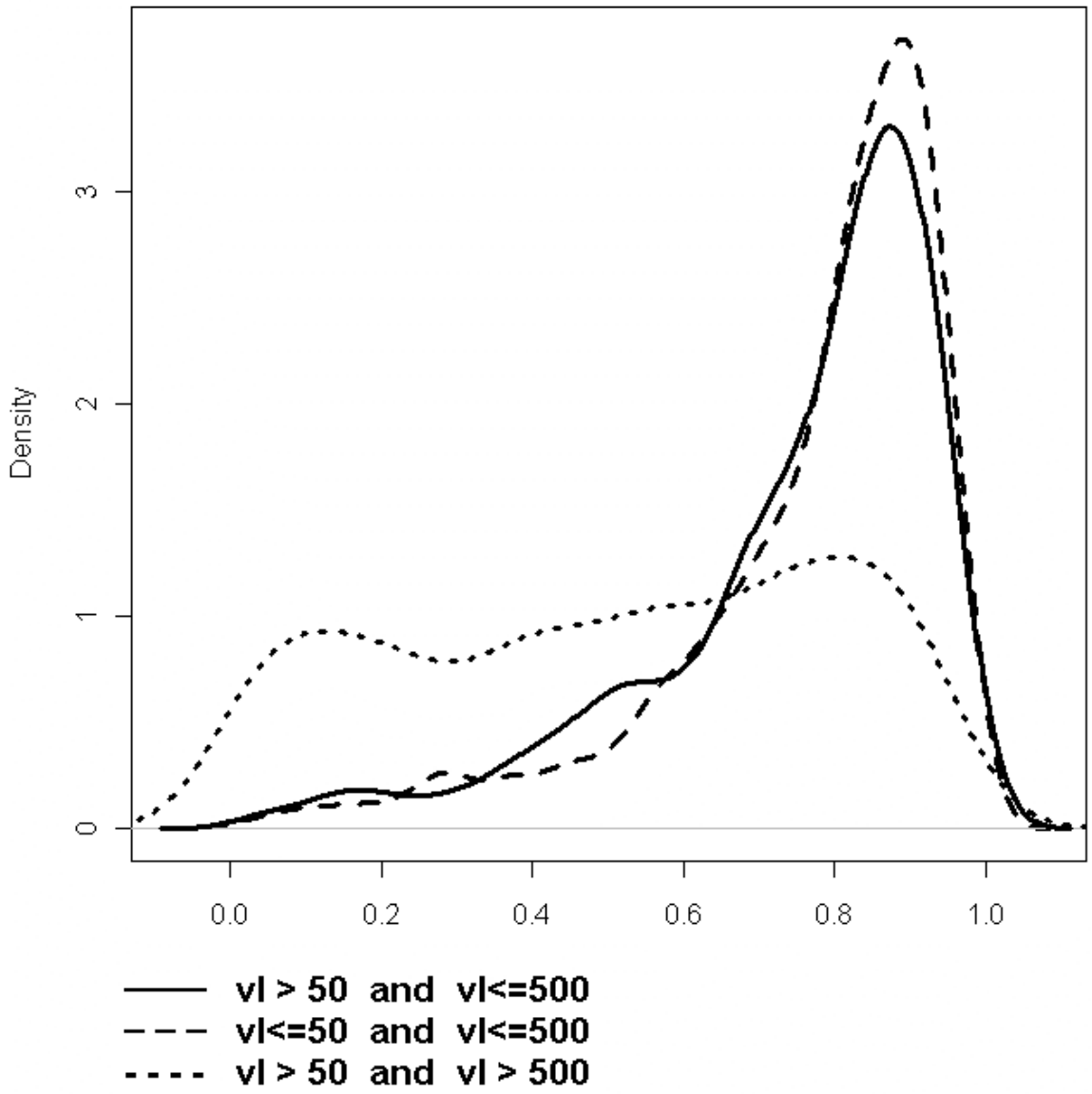


Table A1. *Summary of Logistic Regression model on complete Standard Datum input set, including higher-order variable interactions, trained on EuResist database (coefficients for class 1 = success)*

variable	*odds ratio*	*lower 95 CI*	*upper 95 CI*	*p-value*
Number of drugs in cART	1.91554	1.76827	2.07508	2.00E-16
HIV RNA baseline Log10 copies/ml	0.63763	0.60050	0.67706	2.66E-12
PR_54_V	0.24171	0.18087	0.32303	1.31E-06
EFV and EFV exper.	0.24171	0.17377	0.33622	0.00002
RT_184_V and 3TC	0.51171	0.43605	0.60050	0.00003
SQV and AZT exper.	0.44043	0.35701	0.54335	0.00015
NFV and PI exper.	0.44933	0.36059	0.55990	0.00022
RT_184_V and NVP	0.39852	0.30728	0.51685	0.00034
RT_39_A and RT_211_K	0.43605	0.34646	0.54881	0.00038

(Intercept)	4.75882	3.06485	7.38906	0.00040
RT_67_N and RT_184_V	1.99372	1.63232	2.43513	0.00056
RTV exper.	0.50662	0.41478	0.61878	0.00061
TDF and EFV exper.	0.48191	0.39063	0.59452	0.00063
PR_63_P and PR_90_M	0.55990	0.47237	0.66365	0.00082
PR_89_M and PR_93_L	3.85743	2.55998	5.81244	0.00087
PR_20_M	0.20805	0.12873	0.33622	0.00115
EFV	1.80399	1.50682	2.15977	0.00122
PR_10_I	0.58275	0.48675	0.69768	0.00252
RT_177_E and RT_207_A	2.33965	1.69893	3.22199	0.00754
PR_54_L	0.20190	0.11080	0.36788	0.00758
APV exper.	0.46767	0.34994	0.62500	0.00858
LPV and DDC exper.	1.95424	1.52196	2.50929	0.00870
booster and LPV exper.	0.50158	0.38289	0.65705	0.00940
PR_82_F	0.15724	0.07654	0.32303	0.01039
PR_63_P and IDV	0.52729	0.41066	0.67706	0.01059
PR_63_P and 69_K	0.61878	0.51171	0.74826	0.01156
EFV exper.	1.53726	1.29693	1.82212	0.01194
RT_207_E and RT_41_L	0.51171	0.39063	0.67032	0.01232
D4T and DDI	0.62500	0.51685	0.75578	0.01280
RT_118_I and RT_122_E	0.52729	0.40657	0.68386	0.01311
RT_83_K and RT_135_T	2.09594	1.55271	2.82922	0.01410
D4T and IDV exper.	0.65051	0.54881	0.77105	0.01468
booster and RTV exper.	2.15977	1.56831	2.97427	0.01469
PR_35_D and PR_10_I	1.95424	1.47698	2.58571	0.01521
PR_54_V and PR_93_L	2.58571	1.75067	3.81904	0.01616
risk vertical transmission	0.32956	0.20805	0.52205	0.01683
RT_118_I and DDI	0.48675	0.36059	0.65705	0.01701
PR_46_L	2.13828	1.55271	2.94468	0.01768
AZT and DDC exper.	0.55433	0.43171	0.71177	0.01889
DDI and DDC exper.	0.60653	0.48675	0.75578	0.01969
RT_293_V and DDI	0.65051	0.53259	0.79453	0.02954
RT_41_L and AZT	0.50158	0.36422	0.69073	0.03120
RT_122_E and 181_C	0.55433	0.41895	0.73345	0.03137
RT_122_E and 123_N	0.57695	0.44486	0.74826	0.03474
DDC	0.37158	0.23224	0.59452	0.03545
RT_219_Q and D4T	0.54881	0.41066	0.73345	0.03977
PR_10_F	0.47711	0.32956	0.69073	0.04336
NRTI exper.	0.73345	0.62500	0.86071	0.04618
RT_200_A and RT_67_N	0.57695	0.43605	0.76338	0.05202
RT_123_E and RT_103_N	0.68386	0.55990	0.83527	0.05309
3TC and LPV	1.36343	1.16183	1.59999	0.05370
RT_196_E and RT_41_L	1.56831	1.24608	1.97388	0.05675
RT_135_T and RT_210_W	0.63763	0.50158	0.81058	0.06385
ABC and DDI	0.58860	0.44043	0.78663	0.06645
RT_293_V and EFV	1.84043	1.30996	2.58571	0.06735
RT_211_K and AZT	1.37713	1.15027	1.64872	0.07204
PR_20_R and PR_36_I	0.58860	0.43605	0.79453	0.07982
RT_207_E and AZT	1.53726	1.18530	1.99372	0.10197
LPV and LPV exper.	0.62500	0.46767	0.83527	0.10659
PR_20_R and booster	0.53259	0.35701	0.79453	0.11273
RT_286_A and RT_67_N	0.60050	0.43605	0.82696	0.11273
PR_24_I	0.46301	0.28365	0.75578	0.11811
RT_116_Y	0.44043	0.25666	0.75578	0.13228
RT_245_M	1.52196	1.13883	2.03399	0.14093
CD4 baseline percent	1.01005	1.00000	1.02020	0.15220

Figure A5. *Random Forest feature importance evaluation: mean decrease in accuracy (left panel) and Gini index (right panel)*