**S1 Text. Methods of statistical analysis**

To identify factors associated with the presence of mutation a logistic regression model was used. Variables were included in the multivariate model when univariate comparisons yielded a level of significance of $p<0.15$. The following variables were tested: age, sex, CMV pre-transplant donor/recipient serology (D/R), type of transplant, prophylaxis and induction and maintenance therapy received, number of days after transplantation, if the patient was receiving prophylaxis, preemptive or disease treatment, viral load, CMV disease, and the dose and number of days of GCV or VGCV received prior to the suspicion. Finally, a forward stepwise selection ($p_{in}<0.05$, $p_{out}>0.10$) was used to determine factors associated to the presence of mutations. The odds ratio (OR) and 95% confidence interval (CI) were calculated. For continuous variables with non-normal distribution a logarithmic transformation was used. To identify the problem of collinearity, the r coefficient of two variables were calculated. When two independent variables were highly correlated ($r >|\pm 0.30|$), the variable with the largest variance was excluded from the multivariate analysis [1]. The Hosmer-Lemeshow goodness-of-fit test was performed to assess the overall fit of the models [2]. Internal validation of the prediction model was conducted using ordinary nonparametric bootstrapping with 1,000 bootstrap samples and bias-corrected (S1Table A), accelerated 95% confidence intervals (CIs) [3]. The area under the receiver operating characteristic (ROC) curve was constructed for the ability to predict patients with mutations, using variables derived from the multivariate logistic regression model.

## S2 Text. Results of statistical analysis

Internal validation of the logistic regression model was conducted using bootstrapping with 1,000 samples and demonstrated robust results with small 95% CI around the original coefficient (S2Table) (S1 Fig.).

## S3 Text. References of statistical analysis


1.- Healey JF. Statistics: a tool for social research. Belmont, CA: Wadsworth Publishing Company,2016.

2.- Hosmer D, Lemeshow S. Applied logistic regression. New York: Wiley, 1989.

3.- Efron B, Tibshirani R. An introduction to the bootstrap (Monographs on statistics and applied probability. New York: Chapman and Hall, 1993.


## S1 Table. Internal validation of prediction model for mutations using nonparametric bootstrap technique

| Variable | Original | Bias | SE | P value | 95% BCa CI |
|---|---|---|---|---|---|
| Log treatment duration before suspicion of resistance | 0.806 | 0.171 | 0.662 | 0.072 | 0.111 to 2.487 |

Abbreviations: BCa indicates adjusted bootstrap; CI, confidence interval; SE, standard error.

## S2 Table. Significant univariate and multivariable logistic regression analyses for mutations

| Variable | Univariate[a] | | | Multivariable[bc] | | |
|---|---|---|---|---|---|---|
| | OR | 95% CI | P-value | OR | 95% CI | P-value |
| **Type of transplant: lung** | 9.00 | 0.91 to 89.27 | 0.061 | - | - | - |
| **Induction therapy: basiliximab** | 0.34 | 0.09 to 1.24 | 0.10 | - | - | - |
| **Initial maintance therapy: cyclosporine** | 4.33 | 0.70 to 27.01 | 0.12 | - | - | - |
| **CMV prophylaxis[d]** | | | 0.093 | - | - | - |
| No | 1.00 | - | - | - | - | - |
| 1-3 months | 3.20 | 0.76 to 13.46 | 0.11 | - | - | - |
| 6-12 months | 8.00 | 1.06 to 60.32 | 0.044 | - | - | - |
| Log Time interval between transplantation and suspicion (+1 day)[e] | 1.83 | 0.91 to 3.72 | 0.092 | - | - | - |
| Log Treatment duration | 2.24 | 1.03 to 4.87 | 0.042 | 2.24 | 1.03 to 4.87 | 0.042 |

| Variable | Univariate[a] | | | Multivariable[bc] | | |
|---|---|---|---|---|---|---|
| | OR | 95% CI | P-value | OR | 95% CI | P-value |
| before suspicion of resistance (+1 day) | | | | | | |

Abbreviations: CI indicates confidence interval; COPD, chronic obstructive pulmonary disease; OR, odds ratio. Data are shown as estimated ORs (95% CIs) of the explanatory variables in the mutations group.

The OR is defined as the probability of membership of the group with mutations divided by the probability of membership of the non-mutations group. The P value is based on the null hypothesis that all ORs relating to an explanatory variable equal unity (no effect).

[a] The variables analyzed in the univariate analysis were: age, sex, CMV pre-transplant donor/recipient serology (D/R), type of transplant, prophylaxis and induction and maintenance therapy received, number of days after transplantation, if the patient was receiving prophylaxis, preemptive or disease treatment, viral load, CMV disease, and the dose and number of days of GCV or VGCV received prior to the suspicion.

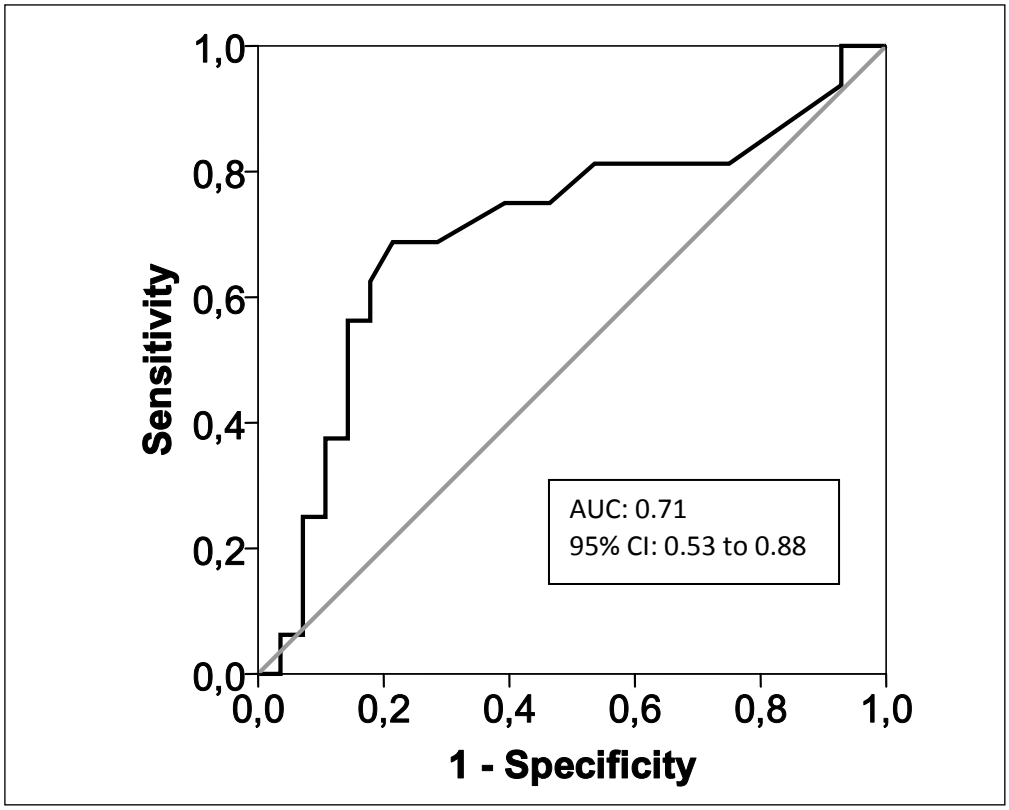
[b] Hosmer-Lemeshow goodness-of-fit test, p=0.33.

[c] Predictors from the model can be used to calculate the probability of mutations by the following formula: Exp(β) / (1 + Exp(β)), where β = -3.265 + 0.806 x Log treatment duration before suspicion of resistance.

[d] The p-value corresponds to differences between the three groups (no, 1-3 months or 3-6 months).

[e] +1 means a one-unit increase on the scale in the predictor variable (i.e., going from 1 to 2, 2 to 3, etc.).

**S1 Fig. Receiver operating characteristic curve analysis of significant variables derived from the logistic regression model for the ability to predict mutations**




Abbreviations: AUC indicates area under the curve; CI, confidence interval.