Author’s response to reviews

Title: Clinical and genetic markers associated with tuberculosis, HIV-1 infection, and TB/HIV-immune reconstitution inflammatory syndrome outcomes

Authors:

Nathalia Beatriz Ramos De Sá (nathalia.beatriz2008@gmail.com)
Marcelo Ribeiro-Alves (mribalves@gmail.com)
Tatiana Pereira Da Silva (tpsilva2000@gmail.com)
José Henrique Pilotto (jhpilotto@gmail.com)
Valeria Cavalcanti Rolla (valeria.rolla@gmail.com)
Carmem Beatriz Wagner Giacoia-Gripp (carmembg@ioc.fiocruz.br)
Daniel Scott-Algara (daniel.scott-algara@pasteur.fr)
Mariza Morgado (mmorgado@ioc.fiocruz.br; mariza.morgado@gmail.com)
Sylvia Lopes Maia Teixeira (sylvialmt@hotmail.com)

Version: 2 Date: 01 Nov 2019

Author’s response to reviews:

Dr. Andrea Calcagno,

Editor

We thank you very much for allowing us to revise our manuscript once again. We also appreciate the review and critical comments and suggestions from reviewers on our manuscript entitled "Clinical and genetic markers associated with tuberculosis, HIV-1 infection and TB/HIV-immune reconstitution inflammatory syndrome outcomes" (INFD-D-19-00477). We have carefully examined the reviewer's comments and questions, and most of them have been incorporated into the manuscript, as can be seen underlined in the text. In addition, all points raised are answered below.

Thanks again for this opportunity.

Sincerely,

Nathalia Beatriz Ramos de Sá, MSci
Mariza Gonçalves Morgado, PhD

Reviewer reports:
Caroline Tiemessen, PhD (Reviewer 1):
The revised manuscript has adequately addressed the concerns raised by the reviewers. The inclusion of HLA-C genotyping is an important addition and has added additional findings.

The authors describe a study investigating the association of HLA-B and HLA-C alleles and KIR genotypes with outcomes of TB, HIV-1 infection and immune reconstitution inflammatory syndrome (IRIS). Patients were from Rio de Janeiro, Brazil - followed from 2006 to 2016 - and constituted 4 groups of patients: Grp1 - TB+/HIV+ (n=88; 11 with IRIS), Grp2 - HIV+ (n=24), Grp3 - TB+ (n=24) and Grp4 - healthy individuals (n=26).

The following findings were reported: With TB as an outcome, KIR2DS2 associated with risk, while HLA-B*08 and female gender. The absence of KIR2DL3 (KIR2DL2 homozygosity) and HLA-C*07 associated with protection against TB in HIV-infected patients.

With IRIS as an outcome - HLA-B*41, KIR2DS2, KIR2DS1 + HLA-C2, KIR2DL3 + HLA-C1/C2 absence, KIR2DL1 + HLA-C1/C2 absence, and CD8 count ≤ 500 associated with an increased risk of IRIS in TB/HIV coinfected patients. Associations of HLA-B*41, KIR2DS2 and KIR-HLA-C pairs with IRIS have previously not been reported.

It needs to be very clear what is meant by KIR2DL3 + HLA-C1/C2 absence, KIR2DL1 + HLA-C1/C2 absence, as this can be very confusing to the reader. See comment for line 307 below.

Specific comments and suggested edits:

Line 36: KIR “alleles” should be “genes” to not confuse with allelic variants at KIR loci. (now line 37).

Line 46: 500 "cells/ul" missing units. (now line 45).

Line 57: repetition of "worldwide".

Line 58,60,62,64: inconsistent use of "HIV" and "HIV-1" - should use "HIV-1" throughout the document.

Line 60,63: inconsistent use of "."or "," in numbers "69.500" should be "69,000" - check consistency throughout document.


Line 75: repeated definition for CTL. (now lines 75/76)

Line 76: add "serving as" before "ligands".

Line 112: "specific single nucleotide polymorphisms (SNPs)" - add definition in full.

Line 127: spelling "enrolment".

Line 129: use "and", not "&".

Line 132: submitted "for" publication. (now line 131)

Line 137: n = 26, inconsistent with spacing "n=26". (now line 162)

Line 143,146: superscript mm3 = mm3.

Line 154: singular not plural: "one non-nucleoside reverse transcriptase inhibitor".

Line 162: repetition with respect to previous sentence.

Answer: We apologize for these misspellings, which were adequately corrected in the text. The manuscript cited in Line 132, ("submitted for publication") was published and was replaced by reference 47. Please note that, due to the new editions in the text, some number of lines may have changed.

Line 170: It is unusual to use "skin colour" rather than "Caucasian, Black or Mixed ancestry" to define individuals racial background?

Answer: We categorized the individuals according to skin color following the classificatory system
employed by the Brazilian Institute of Geography and Statistics (IBGE) (which is an entity linked to the Brazilian Federal Government that is responsible for the official collection of statistical, geographic, cartographic, geodetic and environmental information in Brazil). This system uses the following definitions for skin color or race: "White," "Black," "Yellow," "Brown" and "Indigenous" (Características Étnico-Raciais da População: Classificações e Identidades. Rio de Janeiro: IBGE, 2013. Available on: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9372-caracteristicas-etnico-raciais-da-populacao.html?=&t=publicacoes> Accessed October 30, 2019). For the demographic census, IBGE collects this information based on the self-declaration of the individuals. In order to clarify this point, we added this criterion of classification in the footnote of Table 1.

Line 184: Authors used 4-digit level for HLA typing in methods, but only 2-digit resolution is reported in results. It is more informative to use 4-digit.
Answer: Due to the extent of HLA-B and –C allelic variability in our sampling, which would limit the statistical associations, requesting a higher sample size, we decided to work with the 2-digit resolution that would be informative enough for our proposed analyses, as explained in lines 261-263 of the text describing HLA-B alleles results. In order to complement the information for clarifying the HLA-C alleles analyses, we included a phrase in lines 277-279. It is essential to point out that this decision could only be made after the generation of the results, which revealed this high variability in the allelic composition of HLA-B and –C loci.

Line 190: A mention should be made that Bw4 80I and 80T epitopes were combined and HLA-A Bw4 was not included.
Answer: The reviewer is right about this question. We included more precise details about the HLA epitopes in the Methods section, page 8 (lines 192-195), as follows: "All HLA-B alleles with the Bw4 epitope were grouped, regardless of the amino acid composition in the position 80 (80I or 80T). In this study, we did not evaluate HLA-Bw4 epitope-associated specificity found in some HLA-A alleles”.

Line 197: "were verified" change to "was determined".
Line 216: "KIR alleles" should be KIR genes"; "estimated" change to "determined".
Line 223: "Allelic frequencies" do you mean "HLA allele frequencies".
Answer: All the required changes were corrected in the manuscript (lines 200, 218, and 231 in the new version of the manuscript, respectively).

Line 232,233: One may be introducing a bias by including HIV+ individuals in the TB+ and TB-group. Were the groups combined to increase sample size? The higher numbers of HIV positive patients because of the TB+HIV+ TB group in the "TB combined group" vs the smaller numbers in the HIV group in the "non-TB group" could affect the outcomes. There should be some explanation for combining of different phenotypes within these groups, as one ideally would not do this. If the statistics used considers this adequately (you have adjusted for HIV as a possible confounder etc) then this should be stated in the text as well. It is appreciated that sample numbers are what they are and cannot be added to at this stage. The grouping of different phenotype groups should be more clearly justified and how you have accounted for the heterogeneity between groups and interpretation of the findings when you combine different groups. In Table 2 you show various comparisons of combined and specific groups (HIV patients with and without HIV) - does this help with teasing out the relationships found in the combined groups.
Answer: We would like to thank the reviewer for the opportunity to clarify some aspects of the analysis. We would like to emphasize that the "group" definition given in the methodology section, organized as Group 1 - TB+/HIV+ (n=88; 11 with IRIS), Group 2 - HIV+ (n=24), Group 3 - TB+ (n=24) and Group 4 - healthy individuals (n=26), served only as a reference for the
inclusion/enrollment of patients in the study. We intentionally combined Group 1 and Group 3 and compared them with the union of groups Group 2 and Group 4 for two main reasons: (1) The validity of the analysis. The sample was collected in the city of Rio de Janeiro (Brazil), an endemic area for TB and the Brazilian capital with the second-highest incidence coefficient (89.9 cases/100,000 inhabitants; Brazilian Ministry of Health, 2019). All patients, whether they were living or not with HIV, were at risk for TB infection. So both the Odds-ratio (OR) nominator (Group 1 + Group 3) and the denominator (Group 2 + Group 4) were at risk for TB, which is a requirement for conducting a logistic analysis and the validity of the interpretation of an OR (or aOR). If, otherwise, we had not a binomial outcome, in our case, having or not TB given that they were living in an endemic area for TB, but a time to event outcome, for instance, time until developing TB (again assuming that they were all at risk) and we decided to use a Proportional hazard Cox regression to model our data, then we would probably be violating the proportional risk assumption once the HIV+ individuals would have a distinct hazard function then the HIV- individuals; which, we repeat, is not an assumption of the logistic regression; and, (2) The opportunity to estimate the risk increase of individuals living with HIV. Without the inclusion of HIV+ individuals in these analyses, it would be impossible to estimate the associated risk of developing TB in an endemic area with the co-infection with HIV. Indeed, when the reviewer correctly suggests that we would "maybe introducing a bias by including HIV+ individuals in the TB+ and TB- group" is because someone before us had included HIV+ individuals in a similar analysis and concluded that they had an increased risk when compared to HIV- individuals. We do not regret our analysis decision of including HIV+ individuals. It was worth noting that although in the non-adjusted OR HIV infection appeared associated with an increased risk to the TB co-infection (OR = 3.97 (1.94-8.12); p = 2e-04), in the adjusted OR analysis the increased risk was no longer significative (aOR = 3.85 (0.3-50.09); p = 0.3033). It may suggest that in a high endemic area, living or not with HIV is not as important as we may believe for the TB co-infection. As you wisely anticipated, we introduced HIV-infection factor in the modeling of all other genetic or phenotypic analysis specifically to eliminate any possible bias introduced for having more or fewer individuals living with HIV in the aOR numerator or denominator of any analysis; for instance, between carriers and not carriers of the HLA-B*08 gene. Indeed, we introduced as confounders, any clinical phenotypic marker associated with the different outcomes for this same reason. For example, we introduced gender, skin color, education, HIV status, CD4 count, and CD4/CD8 ratio as confounders to adjust the ORs in the TB analyses among individuals living in an endemic area for TB; all confounders were explicit in table’s footnotes. Following your recommendations, we have now included the phrase “We introduced as confounders, any clinical phenotypic marker associated with the different outcomes in the modeling of all other genetic or phenotypic analysis to eliminate any possible bias introduced for having more or fewer individuals living with HIV in the aOR numerator or denominator of any analysis” in the Methods section, page 9 and 10 (lines 222-225), to make this point clearer to the reader.

Line 237,238: Unusual way to define individuals ethnic background. Can the individuals be assigned Caucasian, Black and Mixed Ancestry based on the skin colour?

Line 253: Using 2-digit resolution may not be the best since some functional differences exist at the 4-digit level. Eg. B*58:01 and B*58:02 have opposite effects of disease progression in HIV-1 infected
patients.
Answer: We agree with the reviewer about the importance of the 4-digit level analysis. As answered above (Line 184 question), due to the extent of HLA-B and –C allelic variability found in our studied population, which restricted the statistical analyses, we decided to work with the 2-digit resolution that would be enough information for our proposed analyses. We would like to note that none of the HLA-B or –C associations found in this study (HLA-B*08; HLA-B*41 and HLA-C*07) involved alleles with functional differences at the 4-digit level, so our decision have no impact in the interpretation of the results.

Line 264: HLA-C and REDOME sentence should be moved to the next paragraph.
Answer: This change was inserted in the Results section, page 12 (lines 284-285), as follows: “Unfortunately, HLA-C frequencies are not available at REDOME dataset for comparisons of our data with the general Brazilian population”.

Line 270: "HLA-B*03" should be "HLA-C*03".
Answer: We are very sorry for this mistake already changed in the manuscript (line 280 in the new version of the manuscript).

Line 279: Most publications that mention KIR genotypes use the nomenclature e.g. "Bx6" meaning all 16 KIR genes were present. It may be better to put Bx in front of the number instead of GID #6, to avoid confusion that a different genotyping nomenclature system was used.
Answer: We agree with the reviewer about this point. Therefore, we changed the nomenclature of the KIR genotypes in the manuscript (lines 288 and 291) and Table S5, as suggested.

Line 291: inconsistency - in other places the 95% CI ranges are omitted, but here they are included.
Answer: We reviewed the entire manuscript, and all 95% CI ranges are included in the new version.

Line 294: Can absence of KIR2DL3 be stated as KIR2DL2 homozygous. Gene association studies make sense with respect to the presence of a gene, not the absence.
Answer: We understand that KIR2DL2 and KIR2DL3 segregate as alleles of the same locus (Jiang et al., 2012; Martin and Carrington, 2013), but we are not secure to say that the absence of KIR2DL3 directly means KIR2DL2 homozygosity. The reason for that is the existence of 54 KIR genotypes (4 AA and 50 Bx) in which KIR2DL2 and KIR2DL3 are simultaneously absent in the "KIR Database – Genotype Reference List" resource included in the "The Allele Frequency Net Database (http://www.allelefrequencies.net)." In the same way, in the study performed by Gentle et al., 2017, a new KIR genotype (named BxN2) was characterized - among other features - by the absence of KIR2DL2 and KIR2DL3, which is described in the article as an unusual (but possible) phenomenon. Additionally, Zwolińska et al., 2016 also describe individuals (n=14 from 577) without these genes.
In our study, besides all KR2DL3, not carriers (n=13) are KIR2DL2 positive, we think that it is not appropriate to consider the term "KIR2DL2 homozygous" in the text, as the used methodology (PCR-SSP) does not give such discrimination, together with the reasons explained above. To make the text more appropriated, we changed the "KIR2DL3 absence" for "not carrying KIR2DL3" and "the presence of HLA-C*07" for "carrying HLA-C*07" (lines 306 and 307 in the new version of the manuscript).

With respect to the second comment, in all phenotypic and genotypic analyses, the reference, or the denominator in the OR or aOR, is represented by the more representative level in our sample, which, in the case of a genotypic analysis, is also representative of the populational genetic background. Indeed, for the KIR2DL3, carriers represent 93.75% (n = 105) and 88% (n = 44) in our sample. Therefore, using KIR2DL3 carriers as a reference, the unconditional logistic multiple regression model performed...
in our study showed an association between the protection for TB onset and not carrying the KIR2DL3 gene (Table 2).

Line 307: It is confusing to write "KIR2DL3 + HLA-C1/C2 absence". Does this mean both KIR and ligands must be absent or can it be absence of KIR, but presence of ligands or vice versa? This could be seen as presence of KIR2DL2 homozygous and either C1 or C2 homozygous was significant. KIR2DL1 + HLA-C1/C2 absence - unlikely this can be absence of KIR2DL1 as a gene as have very high percentages across the groups (most individuals will have at least one copy). Do you mean absence of HLA-C1/C2 only?

What is meant needs to be explicitly stated so the reader understands.

Answer: "C1/C2" means the presence of one HLA-C allele from the C1 epitope group and the other HLA-C allele from the C2 epitope group. When we use "KIR2DL3 + HLA-C1/C2 absence" or "KIR2DL1 + HLA-C1/C2 absence", we refer to those individuals who are not carriers of the respective pair (please see Table 3), due to fact that they present C1/C1 or C2/C2 alleles, instead of C1/C2 alleles. Both KIR2DL1 and KIR2DL3 are present in all IRIS individuals (please see Table S2). In order to avoid misinterpretation, the text in lines 318-324 was fully reviewed ("...an increased risk for IRIS onset was associated with having a CD8 count ≤ 500 cells/mm³ [aOR=18.23 (95% CI, 1.71-193.79), P=0.016]; carrying the KIR2DS2 gene [aOR=27.22 (95% CI, 1.33-558.6), P=0.032], the HLA-B*41 allele [aOR=68.84 (95% CI, 1.41-3369.9) P=0.033], the KIR2DS1 + HLA-C2 pair [aOR=28.58 (95% CI, 1.54-530.65) P=0.024]; and not carrying the KIR2DL3 + HLA-C1/C2 pair [aOR=43.04 (95% CI, 1.32-1404.01) P=0.034], and the KIR2DL1 + HLA-C1/C2 pair [aOR=43.04 (95% CI, 1.32-1404.01) P=0.034] (Table 3). All the discussion about these results was fully reviewed as well (Discussion section, lines 395-398, and 408-410).

Line 309: superscript mm3 = mm3.

Answer: The correct unit is 500 cells/mm³, which was inserted in the revised version of the manuscript (now line 319).

Line 325: "accelerating" change to "improving".

Line 334: Conesa-Botella et al - take out "and collaborators".

Line 339: Pean et al - take out "and colleagues".

Line 343: submitted "for" publication.

Answer: All suggested changes were made (lines 338, 346, 350/351, 355 in the new version of the manuscript, respectively). The manuscript cited in Line 355 ("submitted for publication") was published and was replaced by reference 47.

Line 355: "To the best of our knowledge, HLA-B*08.." See Indian study Shankarkumar et al 2009, JAIDS, vol 51, number 5, August 15, page 640. There is mention of HLA-B*08 in this study that should be discussed.

Answer: The article by Shankarkumar et al., 2009 (78) was inserted in the Discussion session, page 15, and the text (lines 367-372) was modified, as follows: "To the best of our knowledge, HLA-B*08 has not yet been associated with TB protection. However, an increased frequency of this allele was described among TB-HIV coinfected individuals and HIV-1 infected patients with rapid disease progression, reflecting different roles for this allele in the context of TB and HIV-1 infection".

Line 367: KIR2DL2 homozygosity instead of KIR2DL3 absence. KIR2DL2 recognises C1 ligands with a higher affinity than KIR2DL3 - this might be a contributing factor.

Answer: Please see the answer above (line 294 question) regarding the relation between KIR2DL2 and KIR2DL3. We thank the reviewer for raising this point regarding the possible contribution of the
KIR2DL2 in these mechanisms. Unfortunately, in our study, we did not find an association with KIR2DL2 and TB outcomes. To agree with the text in the results section, we also modified this point in the Discussion section (lines 381-382): "Not carrying KIR2DL3 and the carriage of HLA-C*07 were protective factors for TB onset among HIV-1-infected individuals studied here (Table 2)".

Line 383: mm3 = mm3; interesting that CD8 cell count was significant and not CD4; no discussion on why CD4 count below 50 cells/mm3 was not significant (was it also non-significant at 100 cells/mm3?).

Answer: Thank you for pointing out an interesting, non-expected result found in the risk analyses for IRIS onset in individuals co-infected with HIV-TB. First, we would like to emphasize that our cut, either for CD8 or CD4+ T-cell counts, were based exclusively on achieving a good balance among individuals with or without the outcome (in this particular case, IRIS). Indeed, we often choose the round integer number closest to the medians of the outcome's sets defined by the continuous numeric variable. Clarified that, and following your suggestion, we have just now run the analysis with the CD4 count cut-off of 100 cells/mm3, and found a similar result either before or after the corrections by the confounders (the same ones used for the CD4 count cut-off of 50 cells/mm3), OR = 0.29 (0.06-1.48), p = 0.1376, or aOR = 0.34 (0.03-3.58), p = 0.3654. One possible reason for these results comes from the inclusion criteria of HIV+/TB+ individuals with CD4 counts below 350 cells/mm3, as explained in the text (lines 142-144) and, possibly, due to the low frequency of individuals who evolved to IRIS in our study.

Line 384: HLA-B*41 molecules or allotypes - not "antigens".
Line 426: remove "responses " not part of the definition.
Line 434: "8" no"18".

Answer: All suggested changes were made (now lines 398, 440 and 448, respectively).

Table 1: Was there a rationale for choosing 500 as the cut-off for CD8 analysis? Would a different cut-off have yielded different statistical results? It would be good to include a rationale for this choice.

Answer: Thank you for the question. Our rationale for selecting cut-off values for categorizing any numerical continuous variable was based exclusively on achieving a good balance among individuals with or without the outcome. Indeed, we often choose the round integer number closest to the medians of the outcome's sets defined by the continuous numeric variable. This rationale allowed us to incorporate these categorized variables, whenever appropriated or needed, as confounders in the adjusted analyses, even with small contingency sets defined by the outcome, the variable of interest, and the categorized confounder variable; as was the case for the IRIS onset among individuals co-infected with HIV and TB. Following your suggestion, we included the following phrase: “Whenever needed, we categorized continuous numerical variables using as cut-offs the round integer number closest to the medians of the outcome's sets defined by the continuous numeric variable”, just after the modeling approach described in the Methods section, page 10 (lines 225-228).

Table 3: It is surprising not to see any associations of CD4 count < 50 cells/mm3 in the TB-IRIS cases. Low CD4 count (<100 cells/mm3 has been seen as a hallmark of developing IRIS. Would a cut-off of < 100 cells/mm3 have shown an association?

Answer: We would like to thank the referee for raising this question. The rationale for the cut-off value for categorizing the continuous numerical variable CD4 was the same used for all continuous numerical variables described above. However, following your suggestion, we have now run the analysis with the CD4 count cut-off of 100 cells/mm3, and found a similar result either before or after the corrections by the confounders (the same ones used for the CD4 count cut-off of 50 cells/mm3), OR = 0.29 (0.06-1.48), p = 0.1376, or aOR = 0.34 (0.03-3.58), p = 0.3654. Therefore, we decided to keep the cut-off in
50 cells/mm³, based on the same rationale described in the question above (line 383) and because it is a hallmark of immunosuppression in HIV-1 positive individuals.

Supplementary tables:
Viral load cut-off set at 20,000 copies/ml? Rationale for this choice? What happens if 10,000 copies is used as cut-off?
Answer: Thank you for the question. The rationale for the cut-off value for categorizing the continuous numerical variable Viral Load was the same used for all continuous numerical variables described above. Answering what would happen if a new cut-off of 10,000 copies were used instead, we would have very unbalanced contingency sets formed by the cross among non-TB and TB individuals living with HIV, 0 (0%) and 6 (5.4%) for the lower viral load range (< 10,000 copies/mm³), and 24 (21.4%) and 74 (66.1%) for the higher range (≥ 10,000 copies/mm³), for the non-TB and TB groups respectively (p = 0.3321).
Correct : "All de groups" heading of Supplementary Table S4 - to "All the groups".
Answer: We are sorry for this misspelling, that was corrected in the revised manuscript.