PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	PARK2 and pro-/anti- inflammatory cytokine gene interactions contribute to the susceptibility to Leprosy – A Case/Control study of North Indian population
AUTHORS	Chopra, Rupali; Kalaiarasan, Ponnusamy; Ali, Shafat; Srivastava, Amit; Aggarwal, Shweta; Garg, Vijay; Bhattacharya, S.; Bamezai, Ramesh

VERSION 1 - REVIEW

REVIEWER	Aurélie Cobat McGill University Human Genetics Department Montreal, Canada
REVIEW RETURNED	26-Nov-2013

GENERAL COMMENTS	In this paper, the objective of Chopra et al. was to assess the combined role of significantly associated SNPs in PARK2 and cytokines genes in Leprosy susceptibility. Leprosy is a chronic infectious disease caused by Mycobacterium Leprae. There are accumulating evidence that host genetic factors play an important role in the outcome of the disease. To date several loci (including PARK2 and cytokines genes) have been identified. However, effect of interactions between those loci on leprosy susceptibility have never been studied. To answer this question, the authors performed an interaction analysis in a North Indian population including 829 Leprosy patients and 1476 Healthy controls. - I'm wondering to what extent the proposed analyses really assess the combined role of the selected SNPs and interaction between those SNPs. Individual risk for each SNPs should be presented. The authors should discuss the possibility that only one variant captures the observed effect or they should explain why it is not possible if they think so. Why multivariate analysis (i.e. all variants at a time) was not conducted? Overall, the analysis strategy used to answer the question should be better justified and explained. - Genes and SNPs selection is not described enough. Why other loci more recently identified by GWAS (Zhang et al 2009; Zhang 2011; Liu et al 2012) are not included? Quality control (SNPs / samples) is missing. I guess the authors are studying in this paper the SNPs previously identified as significantly associated with Leprosy in the same population. This should be stated more clearly and discussed. Knowing that each SNPs was individually associated with Leprosy in the same population, the observed results seem expected. Do the authors have access to a replication cohort? An interesting analysis could be to show that combining all the risk genotypes provides a better case-control classification tool. - To what analysis does the forward selection method described in the method section refer? Only univariate interaction term analyses are presented. In the interaction analysis, what is the reference

	genotype for computation of ORs? All the samples not carrying the presented genotype combination? Or samples not carrying the presented genotype combination but carrying at least one risk genotype? The latter should be considered if it is not already the case. - Why and how were designed the groups of protective and risk SNPs? A SNP is not by himself protective or at risk but has a risk and a protective allele. Separate analysis for "SNPs providing protection" and "SNPs providing risk" seems somehow artificial. Why not test all possibilities (e.g. risk genotype at PARK2 with risk genotype at LTA, IL-10RB2, BAT1 ...?) There is also a lack of justification for the genetic model chosen for each SNPs. - Discussion (of the results, strengths and weaknesses of the study, meaning of the study, implications, unanswered questions, future research) is almost non-existent. - How was made the diagnosis of Leprosy? - In table 2, results for PB or MB only are presented but they are not discussed in the manuscript. - Could population stratification be an explanation of the present findings?

REVIEWER	Monot, Marc Institut Pasteur Paris France
REVIEW RETURNED	29-Nov-2013

GENERAL COMMENTS	The statistical analysis is well described in the method part of the paper. The authors used published in-silico data to determine a target, PARK2, that contribute to the susceptibility to Leprosy. The methods described in the paper is not enough to allow the study to be repeated, so a supplementary table is needed which describes carefully information about the 2305 samples used in the study : patient, disease forms (pauci, multi or healthy), accession number of the sequence and the genes belonging to pro-, anti-inflammatory cytokines and if available SNP-type of the Mycobaterium leprae strains. Could you performed a genotype interaction analysis between genotype and the SNP-type of the Mycobaterium leprae strains (if available) using the same statistical method ? Could you define the "OR" values ?

VERSION 1 – AUTHOR RESPONSE

Query 1:
I'm wondering to what extent the proposed analyses really assess the combined role of the selected SNPs and interaction between those SNPs. Individual risk for each SNPs should be presented.
The authors should discuss the possibility that only one variant captures the observed effect or they should explain why it is not possible if they think so.
Why multivariate analysis (i.e. all variants at a time) was not conducted? Overall, the analysis strategy

used to answer the question should be better justified and explained.
Reply 1:
As the Study involved the SNPs previously found by us to be significantly associated with Leprosy in the Indian population, providing complete individual information would have duplicated the already published information (Chopra et al, 2013; Ali et al, 2012; Aggarwal et al, 2011). The objectives of the study were to assess the interaction among the SNPs independently associated with leprosy and then look at the pathways or networks among genes which emerged uniquely. This paves the way for precise and efficient diagnostics and therapeutics.
We appreciate the Reviewer's query about the possibility for only one variant to capture the whole effect, usually which would be a possibility when there is haplo-insufficiency or complete dominance in simple diseases. Since Leprosy is a complex disease, the contribution of more than one gene and its variants is reflected and emphasized upon, as shown through overall results for all the significant SNPs towards disease susceptibility, where SNPs with only one allele were non-informative and couldn't be analyzed further.
We have used multivariate analysis in our study, considering all the SNPs together for the interaction analysis. Knowing the fact that the study includes only two categories; Cases and Controls, we chose binary logistic regression analysis for examining interaction among all SNPs together; the results have been shown in Table 1 & 2 of the Ms and adequately explained in the result section.
Query 2:
Genes and SNPs selection is not described enough. Why other loci more recently identified by GWAS (Zhang et al 2009; Zhang 2011; Liu et al 2012) are not included? Quality control (SNPs / samples) is missing. I guess the authors are studying in this paper the SNPs previously identified as significantly associated with Leprosy in the same population. This should be stated more clearly and discussed. Knowing that each SNPs was individually associated with Leprosy in the same population, the observed results seem expected.
Do the authors have access to a replication cohort?
An interesting analysis could be to show that combining all the risk genotypes provides a better case-control classification tool.
Reply 2:
Yes, as pointed out by the reviewer, the analysis is confined to the North Indian population studied by our Laboratory in which these SNPs have shown significant association independently. We have emphasized upon this point in the revised Ms and marked the point in Red Color.
Yes we have access to the replication cohort data for individual SNPs, however, when used for SNP-SNP interaction analysis, the considerable reduction in the sample size has prevented us to carry out combined interaction analysis.
As pointed out by the Reviewer regarding combined analysis, we did the combined analysis of all possible SNP genotypes and selected only those combinations which were providing significant Risk or Protection; others i.e. non-significant combinations are not shown for optimizing space and better understanding. Details are provided in Table 2 of the Ms.
Query 3:
To what analysis does the forward selection method described in the method section refer?
Only univariate interaction term analyses are presented. In the interaction analysis, what is the reference genotype for computation of ORs?
All the samples not carrying the presented genotype combination? Or samples not carrying the presented genotype combination but carrying at least one risk genotype? The latter should be considered if it is not already the case.
Reply 3:
In this manuscript we performed multiple genotype interaction analysis of previously studied SNPs between leprosy cases and controls recruited from North Indian population. For interaction analysis (both pair-wise and multiple), all possible genotype combinations of selected SNPs in different genes were ascertained. However, only the combinations of significantly associated SNP genotypes were presented in the Ms for convenience. These interactions were tested using binary logistic regression

with the forward likelihood ratio based selection method considering all variables independently. In this selection method, entry testing based on the significance of the score statistics and removal testing based on the probability of a likelihood ratio statistics were applied. Furthermore, in multiple gene interaction analysis, all interactions with either risk or protection were combined against other interactions to observe the overall effect of all risk versus protective interactions.

For all our analysis, we considered major allele as reference for computation of ORs.

In this analysis, only those genotype combinations entered into a model, which showed at least minimum threshold of significance (p value ≤ 0.05), while non-significant interactions were omitted due to constraint of size (Table 1 & 2 shown in the Ms).

Query 4:
Why and how were designed the groups of protective and risk SNPs? A SNP is not by himself protective or at risk but has a risk and a protective allele. Separate analysis for "SNPs providing protection" and "SNPs providing risk" seems somehow artificial.
Why not test all possibilities (e.g. risk genotype at PARK2 with risk genotype at LTA, IL-10RB2, BAT1 ...?) There is also a lack of justification for the genetic model chosen for each SNPs.
Reply 4:
In our study, we used binary logistic regression for combined interaction analysis of all the studied SNPs, indicating only few samples (10 cases and 59 controls) showing significant association of SNP genotypes providing protection towards the disease. Similarly, 57 cases and 44 controls showing significant association providing risk towards the disease. In rest of the cases and controls either only one risk or protective allele was present or combination was not significantly associated with the disease.
The analysis for all the different combinations including SNPs for the genes found significantly associated with the disease has already been done in the Ms. Firstly, we performed the analysis by combining significant SNPs in PARK2 gene region with the SNPs of pro-/ anti- inflammatory cytokine gene regions and Secondly, we performed the analysis by combining all the SNPs providing either Risk or the Protection towards the disease to see their overall combined effect towards the disease susceptibility. Results are shown in Table 1 & 2 of the Ms and have been discussed in the discussion section of the Ms.
Query 5:
Discussion (of the results, strengths and weaknesses of the study, meaning of the study, implications, unanswered questions, future research) is almost non-existent.
Reply 5:
As suggested by the Reviewer, we have now included the strengths and weaknesses, implications and future prospects in the discussion section.
Query 6:
How was made the diagnosis of Leprosy?
Reply 6:
According to the WHO guidelines, Diagnosis of leprosy is most commonly based on the clinical signs and symptoms. In an endemic country or area, an individual should be regarded as having leprosy if he or she shows ONE of the following cardinal signs: skin lesion consistent with leprosy and with definite sensory loss, with or without thickened nerves and positive skin smears test. The details has been included in the Ms and marked in Red color.

Query 7:
In table 2, results for PB or MB only are presented but they are not discussed in the manuscript.
Reply 7:

As pointed out by the reviewer we have made the changes in the discussion part of the Ms and marked the changes in Red Color.

Query 8:
Could population stratification be an explanation of the present findings?
Reply 8:
In our earlier published work on the North Indian population group, we had already carried out the experiments to rule out population stratification in our population group by typing 61 individual identifying autosomal SNPs (II SNPs) (Chopra et al, 2013) in the complete set of samples. The results showed a compact cluster of all studied samples from North India ruling out any possibility of population stratification and justifying homogeneity among all samples.
Reviewer
Name Monot
Institution and Country Institut Pasteur Paris
France
Please state any competing interests or state 'None declared': None declared

Query 1:
The statistical analysis is well described in the method part of the paper.
Reply 1:
Authors thank the reviewer for the appreciation.

Query 2:
The authors used published in-silico data to determine a target, PARK2, that contribute to the susceptibility to Leprosy.

Reply 2:
Yes

Query 3:
The methods described in the paper is not enough to allow the study to be repeated, so a supplementary table is needed which describes carefully information about the 2305 samples used in the study: patient, disease forms (pauci, multi or healthy), accession number of the sequence and the genes belonging to pro-, anti-inflammatory cytokines and if available SNP-type of the Mycobaterium leprae strains.
Reply 3:
The information has been added to the respective table and within the section of Materials and Methods of the revised Ms. The study reflects on the human host SNPs in their genes and not the the pathogen (M. leprae). Since, the data for individual SNPs were adopted from our already published work, the detailed information regarding patients/healthy controls, disease forms, accession number of related genes and SNPs is already available online in the respective articles (Chopra et al, 2013; Ali et al, 2012; Aggarwal et al, 2011).
Query 4:
Could you performed a genotype interaction analysis between genotype and the SNP-type of the Mycobaterium leprae strains (if available) using the same statistical method?
Reply 4:
The aim of our study was to see the combined effect of the different genetic loci of host found significantly associated independently with the disease. We concluded that the outcome of this complex disease, Leprosy, is dependent on the host factors controlling the immune response, especially when M. leprae possesses lowest level of genetic diversity (Monot et al, 2013).

Query 5:

Could you define the "OR" values ?
Reply 5:
OR stands for odds ratio. For example, if we have a set of samples (Cases and Controls) with a SNP having alleles X and Y and distributed as Xa, Ya in cases and Xb, Yb in controls. Therefore, odds ratio for the above example will be ratio of Xa/Ya and Xb/Yb.

VERSION 2 – REVIEW

REVIEWER	Aurélie Cobat McGill University Human Genetics Department Montreal, Quebec, Canada
REVIEW RETURNED	10-Jan-2014

- The reviewer completed the checklist but made no further comments.

REVIEWER	Monot Marc Institut Pasteur
REVIEW RETURNED	09-Jan-2014

- The reviewer completed the checklist but made no further comments.