Author's response to reviews

Title: Replication of the Association of Chromosomal Region 9p21.3 with Generalized Aggressive Periodontitis (gAgP) Using an Independent Case-Control Cohort

Authors:

Florian D Ernst (ernst@uni-greifswald.de)
Katharina Uhr (katharina.uhr@gmx.de)
Alexander Teumer (ateumer@uni-greifswald.de)
Jutta Fanghänel (fanghj@uni-greifswald.de)
Susanne Schulz (susanne.schulz@medizin.uni-halle.de)
Barbara Noack (Barbara.Noack@uniklinikum-dresden.de)
Jose Gonzales (Jose.Gonzales@dentist.med.uni-giessen.de)
Stefan Reichert (stefan.reichert@medizin.uni-halle.de)
Peter Eickholz (eickholz@med.uni-frankfurt.de)
Birte Holtfreter (holtfret@uni-greifswald.de)
Peter Meisel (meiselp@uni-greifswald.de)
Gerard J Linden (G.Linden@qub.ac.uk)
Georg Homuth (georg.homuth@uni-greifswald.de)
Thomas Kocher (kocher@uni-greifswald.de)

Version: 3 Date: 24 June 2010

Author's response to reviews: see over
Concerning: Our revised manuscript

Dear Jenny,

We would like to submit our revised manuscript: „Replication of the Association of Chromosomal Region 9p21.3 with Generalized Aggressive Periodontitis (gAgP) Using an Independent Case-Control Cohort“ for publication in „BMC Medical Genetics“.

All changes in the revised manuscript are highlighted in yellow. We paid particular attention in modifying our abstract to making it easier to read and understand. We shortened it, and, as requested, removed the reference. The revised manuscript has been carefully read by our co-author Gerard Linden from Belfast, who is a native English speaker. He focused in particular on the changes that have been made to the manuscript.

We hope that you can now consider the revised paper for publication.

Sincerely yours,

Georg
Our answers to the points raised by the reviewers:

We thank the reviewers for their helpful advice and critiques. The questions and points were addressed in detail as outlined below.

Reviewer 1: Jamal M Stein

Comment 1: All study participants have been examined at different centres (different locations in Germany and Northern Ireland). How many examiners were involved? Was there an inter-examiner calibration performed as to periodontal measurements (CAL values) and/or estimation of radiographic bone loss?

Answer: Since Aggressive Periodontitis (AgP) is a rare disease, it is challenging to identify sufficient affected individuals to obtain a total sample size that allows the generation of statistically significant results. Therefore, we had to employ a multi-centre study design, using samples collected from different specialist dental centres where AgP patients were being treated. The subjects had already been examined and diagnosed with AgP at the respective treatment centres by experienced specialist dentists according to the generally accepted criteria described in the “Methods” section of our manuscript. To adjust for any putative bias of different examiners, T. K. performed a final review of all the available phenotype data and decided which samples met stringent criteria for inclusion in the case and the control cohorts. Furthermore, it has to be emphasized that the methods used for the dental examination that resulted in the diagnosis of AgP are, when carefully applied, robust. Therefore, the bias produced by the involvement of several different examiners can be assumed to be small, and no inter-examiner calibration was performed. A related study, Schaefer et al. (2009), that produced the major association finding which was replicated using an independent case-control cohort in our own study, used a comparable multi-centric approach.

Comment 2: As explained in the discussion, the predicted risk haplotypes may account for gAgP susceptibility to a certain degree; however, lifestyle factors and other genetic polymorphisms may influence disease manifestation as well. Thereby, smoking is one of the most important dose-dependent confounders for periodontitis. Please provide more detailed data on the definition of smoking status and its incorporation in the multivariable analyses and, if possible, on the amount of smoking.
Answer: Detailed data about smoking status as defined by the number of smoked cigarettes per day (cpd) were not available for all patients and all control individuals, only for the probands from the epidemiological cohorts from Greifswald and Belfast (SHIP and PRIME) which served as controls without gAgP. Therefore, we decided to restrict the classification of the individuals from all sub-cohorts to the categories “non-smoker”, “former smoker” and “current smoker”. The respective smoking status of each individual was used as additional covariate in the adjusted model, where “former smoker” and “current smoker” were added as dummy variables with the “non-smoker” status as reference into the regression model. The “Methods” section of our manuscript has been accordingly extended to address this issue.

Comment 3: The discussion would benefit from a few sentences regarding the limitations and clinical implications of the findings of this study.

Answer: The limitations of our findings in explaining the complete genetic background of AgP are discussed in the last paragraph of the “Discussion” section. In the context of putative clinical implications, we have added the following sentence:

“As a clinical implication of the verified association between gAgP and the chromosomal region 9p21.3, that has been clearly associated with CHD in earlier studies, it may be worth informing AgP patients routinely of a possible increased risk of CHD.”

Reviewer 2: Alexandre Rezende Vieira

Comment 1: Introduction and Discussion sections are too long. This work is not about the correlation between cardiovascular diseases and periodontitis and most of the text elaborating on this is superfluous and must be deleted.

Answer: We do not agree with the referee. In several different Genome-wide analyses, the chromosomal region 9p21.3 was first associated with CHD. Based on the well-known observation that, on the other hand, CHD is frequently associated with AgP, Schaefer and co-workers designed their study to test for a putative association between the chromosomal region 9p21.3 and AgP. Their finding of a detectable significant association between AgP and the chromosomal region 9p21.3 was successfully replicated in our study, using an independent case-control cohort. Therefore, it would be very difficult to explain our motivation to test for an association between the chromosomal region 9p21.3 and AgP in our case-control-cohort in the
“Background” section without explaining the links between 9p21.3, AgP, and CHD that were already known before we performed our study.

The same considerations apply in the “Discussion” section: A considerable proportion of this section deals with the putative pathological molecular mechanism that underlies the development of CHD in the carriers of the 9p21.3 risk allele, namely the down-regulation of the cell-cycle inhibitor encoding genes CDKN2A, ARF and CDKN2B which promotes a pro-proliferative and pro-inflammatory phenotype. However, exactly the same mechanism that causes an excessive, self-destructive immune response may be responsible for the development of AgP. Therefore, it seems reasonable to relate the genetic susceptibility locus that pre-disposes for the development of AgP, which was successfully replicated in our study, and the proposed patho-physiological mechanism modulating the development of AgP based on this inheritable genetic risk variant in the “Discussion” section.

Comment 2: In the Introduction section (page 5) the authors mention the limited sample size of the initial study, and a few lines after, that the current work exhibits similar sample size. Since there is at least one individual co-authoring both works, and there is a short time span between the two pieces, it is not clear why these data was not presented in the original manuscript.

Answer: There is indeed one single co-author (B.N.) who contributed to both studies. However, the participation of Dr Noack in the study of Schaefer and co-workers was already completed when we (G.H. and T.K.) asked her to join our small consortium. Therefore, the participation of Dr Noack in our study occurred after her involvement in the study of Schaefer and co-workers. Importantly, the patient samples she contributed to the Schaefer et al. study were, of course, different from those she contributed to our study. We would emphasize that the two mentioned consortia formed autonomously and were completely independent from each other, with the exception of this one common member.

Comment 3: Methods section, cases and controls have a significant difference regarding gender distribution. This may indicate a good chance of undetected population substructure bias. More consideration to this should have been given in the Methods section and some minimal description on how the authors addressed this should be provided. Genomic control approaches, conditional logistic methods?

Answer: In our study, conditional logistic regression analysis to avoid population substructure bias caused by differences in gender distribution was carried out by including sex as an
additional covariate in the adjusted regression model. A more detailed description of the adjustments we performed has now been added to the methods section.

Genomic control approaches represent methods to control inflation of potential false positive associations due to undetected population stratification or undetected relationship among individuals in Genome-Wide Association Studies (GWAS). Since we genotyped only four single tagging SNPs in our case-control cohort, in order to validate the associations between the chromosomal region 9p21.3 and AgP described by Schaefer and co-workers, we were not able to apply any genomic control.

**Comment 4:** Discussion section. The first two paragraphs, as I mentioned above are about cardiovascular diseases, something not tested in this study. However, the authors mention the four genes in this critical region, CDKN2A, ARF, ANRIL, and CDKN2B. There is mention to some previous expression work of these genes, as well as the possibility of linkage disequilibrium to causal genetic variants. Why these four genes were not sequenced in the studied population?

**Answer:** As explained in detail in the „Discussion“ section, the three genes CDKN2A, ARF, and CDKN2B were demonstrated to be regulated by interaction with the regulatory non-coding ANRIL RNA. The ANRIL RNA overlaps with the upstream promoter region of the genes CDKN2A and ARF, whereby ARF shares two exons with CDKN2A, and the complete sequence of the CDKN2B gene. While CDKN2A, ARF, and CDKN2B are transcribed in the same direction, ANRIL is transcribed in the opposite direction. The regulation of the expression of CDKN2A, ARF, and CDKN2B is modulated via interaction with different allele-specific variants of ANRIL:

One SNP (rs1333045) seems to be responsible for an increased expression of short splicing variants of the ANRIL regulatory RNA in carriers of the CHD and AgP risk allele as compared to carriers of the non-risk allele, whereas the long ANRIL variant is present in decreased amounts. There is a positive correlation between the long splicing variant of ANRIL and the amount of CDKN2A and CDKN2B specific mRNA. On the other hand, the amount of CDKN2A and CDKN2B specific mRNA is lower in carriers of the CHD and AgP risk allele who exhibit decreased amounts of the long ANRIL variant and increased amounts of the short splicing variants. According to this model, the risk allele causes increased amounts of short ANRIL variants, which in turn results in reduced expression of CDKN2A and CDKN2B (and ARF).

As an important consequence of this regulatory model, there will be no differences in the coding sequences of the three genes CDKN2A, ARF, and CDKN2B between risk allele carriers and the carriers of the non-risk allele. Only their expression level is expected to be different, and therefore, it would not make sense to sequence these genes in our case-control cohort.
Concerning ANRIL, the crucial sequence variation that causes increased expression of either long or short splicing variants and therefore serves as a "molecular switch" might indeed be rs1333045, as suggested in one of the cited studies (Jarinova et al., 2009). On the other hand, since no detailed in-depth-sequencing analysis of the corresponding chromosomal region in carriers of the risk-allele and the non-risk-allele has been performed until now, it might be that the real causative sequence variation has not yet been identified. However, it is clear that this hypothetical not-yet identified sequence variation has to be in strong linkage disequilibrium with rs1333045, and 1333045 is in turn in strong linkage disequilibrium with the rs1333048 within the LD region 1 (R²-value: 0.96) that has been genotyped in our study. Therefore, rs1333048 definitely represents a reliable proxy for the putative not-yet identified causative sequence variation.

As the aim of our work was the replication of the association finding between AgP and the chromosomal region 9p21.3 first described by Schaefer and co-workers, it was sufficient to genotype four tagging SNPs covering the locus in the complete case-control cohort. To achieve this aim, in-depth-sequencing of the ANRIL gene which encompasses 163.3 kb (!) was not necessary. This approach, although it would certainly produce highly interesting results, would be far beyond the scope of our study.

Reviewer 3: Paula Trevillato

Comment 1: Obesity is not considered a known risk factor for gAgP as it is for heart diseases. In this context, shouldn’t the authors consider removing this condition from examples of common risk factors between both gAgP and coronary heart disease in the Introduction section on the 4th page, second paragraph, line 7?

Answer: We changed the corresponding text from “Both diseases have several environmental and behavioural risk factors in common such as education, smoking, diabetes mellitus and obesity [9, 10].” to “Both diseases have several environmental and behavioural risk factors in common such as education, smoking, and diabetes mellitus [9, 10].”

Comment 2: The authors ascertain that the four analyzed SNPs were found to be in Hardy-Weinberg equilibrium (HWE) (Results, page 8, first paragraph), but in table 2 p values of SNP rs496892 for all and unaffected subjects are smaller than 0.05. Are the p values correct? If yes, this marker is not in HWE. I suggest you include a brief comment on that in the Discussion section and also this information in the appropriate part of the Results. Thus, this alteration
should be simple once this SNP was not associated with gAgP and does not belong to the LD block associated with it.

**Answer:** The testing of SNPs for HWE is commonly used as a quality control step to exclude SNPs from the analysis that exhibit genotype frequencies in the analysed sample which are biased due to genotyping errors. Using a p-value cut-off of 0.05 implies that 5% of all SNPs are expected to be removed via HWE test by chance even if there are no genotyping errors. In most cases, this cut-off is too conservative. Moreover, there are known violations of the common assumptions for HWE and other limitations of the analysed sample set that can result in deviations from HWE, like migration, limited sample size, and the combination of slightly different populations. Therefore, current association studies of SNPs are using p-value cut-offs for the HWE testing of 0.001 (for instance, Newton-Cheh, C. et al.: Genome-wide association study identifies eight loci associated with blood pressure. Nature Genetics 2009, Supplemental). Since the HWE is reached under theoretical conditions only, we do not think that the p-values we obtained from the HWE test suggest substantial deviation from HWE, especially because all p-values were greater than 0.01. However, for clarification, we modified the “Results” section accordingly and noted the p-value cut-off we used.

**Comment 3:** Are the authors sure that the SNP ascertained rs496892 for the studies of Greifswald and Schaefer et al. cohorts (Table 1), whose alleles were termed A/G and C/T, is the same one? If yes, I suggest you consider the A/G alleles, with their frequencies: 49 and 51% frequencies are very close and, in my opinion, are not “opposite” as the authors mention in the Results, first paragraph?

**Answer:** We changed the corresponding text from “In contrast, the minor allele for rs496892 being the effect allele was the opposite compared to Schaefer and co-workers (A to G) [1].” to “The effect allele of rs496892 was around 50% in our study as well as in the study of Schaefer and co-workers (2009) (A to G) [1].” Table 1 was also modified accordingly.

**Comment 4:** Using adjusted SNP data, shouldn’t analyses be adjusted by the covariate sex as well, once the frequencies of males and females invert in the case and control subjects (Results, page 8, last paragraph)?

**Answer:** The adjusted analyses indeed included a covariate for sex. A more detailed description of the adjustments that were used has now been added to the methods section.
Comment 5: There is a parenthesis which should be removed after the last word of the first paragraph on page 10 (after respectively).

Answer: The complete sentence reads as follows: "Using a dominant model in the meta-analysis, only rs496892 representing LD region 2 exhibited significant association after Bonferroni correction with $p = 1.24 \times 10^{-02}$, with this value being lower than those obtained using the two single cohorts ($p = 4.94 \times 10^{-02}$ and $p = 2.60 \times 10^{-02}$ for the Greifswald and the Schaefer et al. cohort [1], respectively). Therefore, the mentioned parenthesis is at the correct position: It represents the second of the two parentheses bordering the sentence "($p = 4.94 \times 10^{-02}$ and $p = 2.60 \times 10^{-02}$ for the Greifswald and the Schaefer et al. cohort [1], respectively)."

Comment 6: In the Discussion section where is figure 1 referenced? Shouldn’t it be placed in the first paragraph, on page 11?

Answer: We now added a reference to Figure 1 to the first paragraph of the discussion section.

Comment 7: The authors could include some information about ANRIL gene in the Discussion section, once most of the SNPs, including the putative functional ones, are inside and nearby this gene. Why didn’t the authors perform the genotyping of rs10757278 and rs1333045 SNPs? Haplotype analysis could be done with their own results instead of using data from the Hapmap.

Answer: To the best of our knowledge, we included all relevant information about ANRIL in our manuscript, namely in the “Discussion” section. As ANRIL obviously does not encode a protein, its main function seems to be the regulation of the expression strength of the three mentioned cell-cycle inhibitor encoding genes CDKN2A, ARF, and CDKN2B. How this regulation is modulated in detail is extensively described in the “Discussion” section of our manuscript. The SNPs rs10757278 and rs1333045 were not directly genotyped in our study, because we wanted to perform a direct replication of SNPs that have been originally genotyped in the study of Schaefer and co-workers. This was not the case for rs10757278 and rs1333045. However, as both SNPs are located within the LD region 1 and exhibit strong linkage disequilibrium with the directly genotyped SNPs rs1333048, rs1333042, and rs2891168, the latter could be used as reliable proxies of rs10757278 and rs1333045.

Comment 8: In the Conclusions section: …an association between the chromosomal region 9p21.3 AND gAgP…
Answer: We changed the corresponding text from: “In conclusion, using an independent case-control cohort, we positively replicated the finding of an association between the chromosomal region 9p21.3 with gAgP which was first described by Schaefer and co-workers [1].” to “In conclusion, using an independent case-control cohort, we positively replicated the finding of an association between the chromosomal region 9p21.3 and gAgP which was first described by Schaefer and co-workers [1].”