Author's response to reviews

Title: Admixture Mapping of End Stage Kidney Disease Genetic Susceptibility Using Estimated Mutual Information Ancestry Informative Markers.

Authors:

Liran I Shlush (lshlush@tx.technion.ac.il)
Sivan Bercovici (sberco@cs.technion.ac.il)
Walter G Wasser (lcdialysis@aol.com)
Guennady Yudkovsky (yudgena@yahoo.com)
Alan Templeton (temple_a@wustl.edu)
Dan Geiger (dang@cs.technion.ac.il)
Karl Skorecki (skorecki@tx.technion.ac.il)

Version: 2 Date: 24 July 2010

Author's response to reviews: see over
Dear Professor McCarthy

re: Manuscript 7311936332793889 "Admixture Mapping of End Stage Kidney Disease Genetic Susceptibility Using Estimated Mutual Information Ancestry Informative Markers."

Thank you for the careful editorial consideration and excellent reviews of the above noted manuscript. We were very pleased that the reviewers found the manuscript to merit publication after revisions, and for the opportunity to re-submit a revised version. We have addressed the reviewers' thoughtful concerns and comments, and now re-submit a revised version, which we believe addresses all of the reviewers' comments. We are grateful for the comments and suggestions, and feel that the manuscript now incorporating the revisions in response to the reviewers' comments is substantially improved, and hopefully worthy of publication in BMC Medical Genomics.

As requested, in order to facilitate the further decision making process, we provide below a detailed item-by-item list of responses and revisions.

Thank you and best regards,

Karl Skorecki

on behalf of the authors
Detailed Responses to Reviewers Comments

Reviewer: 1 (Peter Maxwell)

-Major Compulsory Revisions:

1. The source of reference data should be stated for SNP and gene positions. For example the manuscript (page 9, line 18) includes the following text, "33-35Mbs (rs3752596 to rs2076084)". According to reference Human Genome Build 37.1 (accessed through mapview) rs3752596 maps to 35.6 Mbs while rs2076084 is positioned ~ 36.9 Mbs. MHY9 is currently located at 36.7 Mb which is within the distance, but it may be difficult for readers to follow the data without specific details for all reference maps.

We thank the reviewer for seeking this clarification, and now state explicitly in the "Methods" section the number of the Human Genome Build and dbSNP Build, for both the screening panel and the enrichment panel.

This clarification now appears on page 20, lines:1-2, of the revised manuscript, as follows:

"All physical positions and SNP details reported in the current study were based on NCBI Genome Build 36, and dbSNP Build 127."

2. On page 10, lines 3 and 4 it is stated "SNP rs7286127, whose physical location is 34.8Mbs also located in the MYH9 gene". Mapview (http://www.ncbi.nlm.nih.gov/projects/mapview) displays this SNP at ~ 36.5 Mbs in genome build 37.1 with no genes in the immediate vicinity based on sequence map and 11 genes (none named MYH9) on cytogenetic map. To clarify this relationship, please cite the evidence for this SNP located in MYH9 as the current build of dbSNP (Build 131) does not show this SNP mapped to a particular gene.
We agree with the reviewer that rs7286127 in the current dbSNP Build and Human Genome Build 37.1 is not located in the region of MYH9, and thank him for pointing this out. This is related to the introduction of new Builds since the time the manuscript was submitted. We prefer to omit the sentence specifying the location of the SNPs, since the validation of MYH9 as the causative gene has not been proven, and since submission of the manuscript, new candidate genes in the admixture mapping peak region, such as APO-L1 have been suggested to carry the causative mutation, as pointed out in the revised manuscript, together with the new citations. Accordingly, we have omitted the designation of MYH9, and only state the physical position of the SNP according to Human Genome Build 36 as above.

This clarification now appears on page 10, lines: 5-9, of the revised manuscript, as follows:

" spacing of > 50Kb. Analysis using ADMIXMAP yielded a - LOG (P) score of ~15-16 spanning 7 consecutive SNPs in the region of rs4821667 to rs739016 (Supp. Table 2) (Figure 4). Notably, this region contains the MYH9 gene, as well as the neighboring APOL1 gene, more recently shown to contain coding region variants which are candidates for ESKD risk causation [23, 24]."

We have further changed the wording regarding MYH9 in other parts of the manuscript, due to the recent data pointing to variants in alternative plausible candidate genes for harboring the disease risk phenotype causative mutation.

(Abstract page 2, lines 17-18, and Discussion page 10, line 27, and page 11, lines: 1-3).

3. For Table 2: It would be helpful to include sequence position for each SNP in terms of bases in addition to cM map position. Please specify which map was used as the reference.

We thank the reviewer for this comment, and accordingly have now added the physical position for each SNP, and the Human Genome Build Number 36.

The revised table now appears on page 30, and includes a further modification in response to the request by Reviewer 2 to exclude X chromosome analysis.
Table 2 - LOG (P) score of candidate loci by ADMIXMAP screening panel.

<table>
<thead>
<tr>
<th>Locus Name</th>
<th>Position (Cm) Build 36</th>
<th>Position (Bp) Build 36</th>
<th>Chromosome</th>
<th>- LOG (P) All Cases</th>
<th>- LOG (P) NonDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6811604</td>
<td>50.57</td>
<td>29549192</td>
<td>4</td>
<td>3.32</td>
<td>0.37</td>
</tr>
<tr>
<td>rs3923114</td>
<td>76.42</td>
<td>68362054</td>
<td>5</td>
<td>3.05</td>
<td>0.34</td>
</tr>
<tr>
<td>rs7700504</td>
<td>81.30</td>
<td>73347096</td>
<td>5</td>
<td>3.15</td>
<td>0.46</td>
</tr>
<tr>
<td>rs16872999</td>
<td>83.74</td>
<td>75235507</td>
<td>5</td>
<td>3.38</td>
<td>0.81</td>
</tr>
<tr>
<td>rs7712675</td>
<td>85.80</td>
<td>76174323</td>
<td>5</td>
<td>3.06</td>
<td>0.72</td>
</tr>
<tr>
<td>rs12938039</td>
<td>94.49</td>
<td>64732231</td>
<td>17</td>
<td>3.05</td>
<td>0.01</td>
</tr>
<tr>
<td>rs7286127</td>
<td>37.43</td>
<td>34828811</td>
<td>22</td>
<td>3.77</td>
<td>3.95</td>
</tr>
<tr>
<td>rs5756133</td>
<td>38.16</td>
<td>35023926</td>
<td>22</td>
<td>3.46</td>
<td>3.95</td>
</tr>
</tbody>
</table>

All AIMs with a - LOG (P) greater than 3 are presented, for the all ESKD patients (576 samples) and for the nondiabetic only patients (299 samples) Cm - centimorgans, Bp – base pairs.

- Minor Essential Revisions

1. There is inconsistent style in the way journal names are cited and how references are capitalised in the reference list provided e.g. References 5 and 7 Jama should be JAMA Reference 15 American journal of human genetics should Am J Hum Genet n addition please fix reference style of journal citation in References 19, 20, 21, 22, 23, 24, 25, 29, 31, 32, 38, 39, 40, 41, 42 and 43.

We thank the reviewer for this important and helpful attention and for pointing out the need for uniform style and format. We have now changed the format of all the references so that they are according to a uniform style and conform to the BioMed Central format. The revised reference list is located on pages: 24-27 of the revised
manuscript. As noted above, we have also included additional references which have become relevant since the initial submission date.

2. Minor typographical errors
   page 9, line 19. Please correct spelling of ANCESTRYMAP instead of ANCESSTRYMAP page 14, lines 16 and 22 Please correct spelling of ANCESTRYMAP instead of ANCESTYMAP page 19, line 18 Please correct spelling of ANCESTRYMAP instead of ANCESTYMAP.

   We thank the reviewer for pointing out these typographical and spelling errors, and we have made these corrections accordingly.

3. The readability of the manuscript would be further improved by formatting to introduce multiple paragraphs that would then break up long sections of text on pages 4, 5, 11, 13 and 14.

   We thank the reviewer for this important comment. Indeed, we have gone ahead and introduced new paragraph breaks, which do indeed render the manuscript more readable.
Reviewer: 2 (Elad Ziv)

-Major Compulsory Revisions:

1. The neg log p values reported from admixmap seem too high compared with the results from ancestrymap, esp in figure 4. Assuming the authors use base 10 negative log p values then values of 15-20 suggest a very strong association result that is more significant than a genome wide association. I am wondering whether the authors are reporting negative LN(p) or negative Log base10(P). They do not clarify this in their methods section. Based on the fact that a neg log p of ~20 vs. an LGS of ~4-5 is seen in figure 4 it seems more likely that this is a natural log. If so, please convert to log base 10. If not, please clarify the discordance in statistical significance.

We thank the reviewer for this very insightful observation and for seeking this important clarification. Indeed the base of the negative log was not defined in the manuscript. The results were presented in minus log base10 of the P value. Indeed, as the reviewer correctly points out discordance in statistical significance can be observed when comparing the p values of ANCESTRYMAP and ADMIXMAP. We now address this point in the revised manuscript page 16 lines 21-25, and page 17, lines 1-6:

"In the current study it may be observed that the p values obtained from ADMIXMAP, were more significant in comparison to those obtained using the ANCESTRYMAP LOD scores (Figure 4 reduced panel). One possible explanation is that ADMIXMAP does not apply any correction for multiple hypothesis testing [41]. Furthermore as can be observed in Fig. 4 ADMIXMAP results are greatly influenced by the omission of dense SNPs (with high LD in ancestral populations), while the corresponding changes for ANCESTRYMAP are only moderate. However, we caution that while highly tenable, we cannot conclude definitively that this is the source of discordance between ANCESTRY MAP and ADMIXMAP, since each of these analytical programs calculates a different statistic for LD"
We further describe the log base in the figure legends 3 and 4, as well as in Methods: page 21 lines 22-23.

2. The fact that the authors are running a case only analysis (as opposed to case-control) analysis should be noted in the introduction or methods and the ramifications should be addressed in the discussion. In particular, the fact that false positive results may occur due to mis-estimation of locus specific ancestry should be explained. Since there are already previous publications on this locus, this is unlikely to be the case, but it should still be addressed as a limitation.

We thank the reviewer for this thoughtful and instructive comment. In the previous version of the manuscript we had only partially addressed the issue of case only design with relation to false positive results due to the use of dense markers and LD in the ancestral population. We have added this important cautionary reservation as requested by the reviewer, in the revised Methods section, page 18 line3:

"Study Design

A multi-step case only MALD study was designed, suitable to the available sample set and genomic association question. Study design consisted of the following steps:"

We also discuss the implications of case only statistics, as suggested by the reviewer, on page 12, lines: 24,25, and page 13, lines 1-9 of the revised manuscript, as follows:

"Furthermore, we have used a case only study design, which for a given sample size and marker set, has been suggested by others to be more powerful than a case control study design (reviewed in Montana et.al [38]). However, it should be noted that such a case only study design, might be more sensitive to inaccuracies in estimation of specified allele frequencies. This is less problematic with current builds of dbSNP, which are based on larger sample sets in comparison to the SNPs datasets that were available in earlier admixture mapping studies. Notwithstanding these
reservations, the analysis based on real samples in the current study, in which the admixture peak and its genetic loci have been validated in independent studies, lends further confidence to the predicted power and accuracy of an appropriately formulated case only study design."

3. The inclusion of the X chromosome is confusing. In fact, as the authors note, the X is often skewed in terms of ancestry due to differential proportions of sexes in the founding of the admixed population. Therefore, it seems simpler to exclude the X-chromosome, or at least account for it differently (use the X mean ancestry to adjust for the locus specific ancestry at each locus on the X.) Currently, table 2 makes it seem like there is a signal on the X which is not fully addressed by the authors.

We thank the reviewer for this astute and helpful comment. Indeed the results of chromosome X might be confounding for the reasons noted by the reviewer. Therefore, in the revised manuscript, we have omitted X chromosome markers from the analysis, and specifically explain the rationale of the possible confounding effect of differential gender proportion in the founding and the admixed populations. The X chromosome results were omitted from: Figure 2 and Table 2, and also better explained in the text of the revised manuscript, page 7, lines: 4-10:

"X chromosome variation marker analysis, generally shows greater differentiation among human populations than autosomes, related to the difference in effective population size, and as reflected by gene flow and or drift patterns [1]. Inclusion of X chromosome markers, can possibly introduce a confounding effect of differential gender proportion in the founding and the admixed populations. Therefore, the results of X chromosome admixture mapping are not included in the current study."
-Minor revisions:

1. The use of the term "MALD" should probably be avoided. The term "admixture mapping" is more appropriate for this paper. While the 2 terms are often used interchangeably, they historically refer to different approaches. MALD, as originally proposed in AJHG 1994, used long range LD in admixed populations to do association mapping (ie marker and trait direct association). But there was no estimation of ancestral states of the chromosomes (locus specific ancestry.) In contrast, admixture mapping, as proposed by McKeigue in AJHG 1997 explicitly infers ancestral states of chromosomes and associates these with the trait. Thus, MALD and admixture mapping are qualitatively different if these original definitions are used. This paper uses admixture mapping and should refer to the procedure by that term.

We thank the reviewer for the informative clarification and accordingly have changed MALD to admixture mapping throughout the revised manuscript.

2. For the presentation of the results, it may be instructive to present in either figure or table form the excess ancestry or the risk associated with locus specific ancestry at chromosome 22. That is essentially the key parameter. The discussion should also comment about whether this excess ancestry explains only part or most/all of the excess non-diabetes ESKD seen in the population.

We thank the reviewer for the helpful comment. Indeed previous publications have provided the analyses as suggested by the reviewer and these are indeed highly informative regarding the key parameter of risk associated with locus specific ancestry. Accordingly we now provide in the revised manuscript a reference in which the analysis requested by the reviewer has been cited (reference 45 in the revised manuscript), and also provide the requested results, page 10, lines 13-16, and add a new table (Table. 3), as suggested:

"The excess of African ancestry associated with the genomic region on chromosome 22 noted above, is also presented in Table 3. Furthermore the genome wide score for
disease risk attributed to having one African ancestry allele is 1.6 and for African ancestry alleles the risk is 2.6.

We further added a paragraph in the Discussion, addressing the contribution of the genetic risk as estimated in the current study analysis to the epidemiological risk (page 11 lines 22-25 of revised manuscript):

"In the sub-group of non-diabetic ESKD AA, the increased genetic risk conferred by the African ancestry risk allele explains the majority of the increased risk (1.6 for heterozygote and 1.87 increased risk in the epidemiological studies[8])."

Table 3 Excess of African ancestry for markers in risk region compared to genome average in African American ESKD subjects

<table>
<thead>
<tr>
<th>Locus Name</th>
<th>Expected African Ancestry Among AA*</th>
<th>Estimated African Ancestry Among AA ESKD patients**</th>
<th>% Change in African Ancestry</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7286127</td>
<td>0.833</td>
<td>0.923</td>
<td>10.8%</td>
</tr>
<tr>
<td>rs5756133</td>
<td>0.865</td>
<td>0.954</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

*Expected locus specific African ancestry was calculated from the African allele frequency as was reported by dbSNP build 127, multiplied by genome average African ancestry as was reported by ANCESTRYMAP.

**Estimated African ancestry was reported for each locus using ANCESTRYMAP.