Author's response to reviews

Title: Genetic variations on chromosome 15q14-22.1 for type 2 diabetes candidate region in the Japanese population

Authors:

Yuka Yamaguchi (yama@genome.tokushima-u.ac.jp)
Maki Moritani (moritani@genome.tokushima-u.ac.jp)
Dai Osabe (d_osabe@ip.fujitsu.com)
Toshihito Tanahashi (tana@genome.tokushima-u.ac.jp)
Kyoko Nomura (nomura.kyoko@ip.fujitsu.com)
Yuka Fujita (fujita@genome.tokushima-u.ac.jp)
Parvaneh Keshavarz (parvaneh@genome.tokushima-u.ac.jp)
Kiyoshi Kunika (kunika@genome.tokushima-u.ac.jp)
Naoto Nakamura (naoto@koto.kpu-m.ac.jp)
Toshikazu Yoshikawa (toshi@koto.kpu-m.ac.jp)
Eiichiro Ichiishi (ichiishi@XG7.SO-NET.ne.jp)
Hiroshi Shiota (shiota@clin.med.tokushima-u.ac.jp)
Natsuo Yasui (nyasui@clin.med.tokushima-u.ac.jp)
Hiroshi Inoue (hinoue@genome.tokushima-u.ac.jp)
Mitsuo Itakura (itakura@genome.tokushima-u.ac.jp)

Version: 3 Date: 27 October 2007

Author's response to reviews: see over
October 27, 2007

Prof. Melissa Norton, MD
Editor-in Chief

*BMC Medical Genetics*

Dear Prof. Melissa Norton, MD:

We sincerely appreciate your kind e-mail letter of September 17, 2007 concerning our original manuscript (MS: 1340804552153919) with two Referees’ constructive comments and instruction by Dr. Chrissie Kouremenou, Senior Assistant Editor. These comments have been very helpful to revise our manuscript. We thoroughly revised our manuscript in response to all the comments raised by Referee 1, Referee 2, and the Editorial office, particularly those of Referee 1 one by one.

Please find enclosed our revised manuscript entitled “Genetic variations on chromosome 15q14-22.1 for type 2 diabetes candidate region in the Japanese population” and a complete list of our responses to two Referees’ comments.

We sincerely hope that our revised manuscript now meets the high scientific standard of your journal, and you would now consider our revised manuscript suitable for publication in the *BMC Medical Genetics*. This manuscript has not been previously published elsewhere, nor is under consideration for publication elsewhere. All authors approved of the revised manuscript submitted for publication.

Respectfully yours,

Mitsuo Itakura, M.D., Ph.D.
Professor and Director
Division of Genetic Information
Institute for Genome Research
The University of Tokushima
Response to Referee 1’s comments (1598237750157653_comment.pdf)

We sincerely appreciate your thoughtful comments. Your comments have been very helpful to revise our manuscript. First of all, we apologize for insufficient description in our original manuscript and for not following the guidelines for Abstract. We have changed our original manuscript in response to your comments. The revised manuscript, thus, contains substantial changes in the main text, Tables, Figures, and Additional files. In our revised manuscript, the revised portions are in red font for your attention. We believe that original manuscript has been completely changed in response to your comments. Please find below our response to each of your comments in detail.

General comment

This paper examines a region of replicated linkage on chromosome 15q14-22.1 (32.6-51.2 cM). The reason for choosing this region is not entirely clear, and one wonders if the real region might have been missed. Much of the data is relegated to 10 "additional files" which are not easily read. This makes the data difficult to interpret. Much data are not very helpful, whereas essential data (how much genetic variation was captured ?) are not included. This is a tremendous amount of work, but in the end the results are essentially negative. The authors need to decide whether this is really the correct time to publish what is essentially a progress report.

Response-General comment

Our replies to three points in General comment are as follows;

First, you raised one question, which is the same as Comment-4 in Major Compulsory Revisions, if our choosing of the target region is appropriate. Our response to this question is summarized in Response-4b to Major Compulsory Revisions.

Second, our response to your comment that much data are relegated to 10 Additional files, is explained in Response-1, 2, and 6a to Major Compulsory Revisions and in Response-1 and 7 to Minor Essential Revisions.

Third, you raised question that the authors need to decide to publish or not what essentially is a progress report. As you mentioned, genome-wide association study
(GWAS) is currently applied to common disease including T2D using new technologies such as genome-wide array or beads can potentially find many novel regions including disease susceptibility genes in a large scale experiment. In contrast to the recent status, GWAS technology was not available for the selected region at the onset of our study. We believe that we should publish our data because our data in this study could help uncover the insights on the susceptibility variant(s) on the candidate region on chromosome 15q in the Japanese, which could be used as a guide for further association tests including replication and/or meta-analysis.

**Major Compulsory Revisions**

[Comment-1]

The amount of genetic variation captured based on either JSNP or HapMap Asian samples should be provided. Were there gaps that were not covered? The essential information is based on r^2 rather than D'.

**Response-1**

The point about how much genetic variation in the target region was captured based on JSNP is the same as raised in General comment. In response to this comment, we replied the followings; 1) we compared the amount of genetic variation captured by our SNPs with JSNP data, and presented the results in a newly created Table 2 in our revised text, 2) we inserted our original Additional file 4 presenting the comparison of our SNPs distribution with JSNP in the target region as Figure 1 in our revised text, because one can visually detect gaps that were not covered, and 3) we cited your comment in our revised Results.

JSNP database is aimed to focus on gene-centric SNPs, even though the gene coverage is not complete due to the lack of well-annotated human genome information at the time of its construction. In addition, no sufficient SNPs database such as HapMap was available at the onset of this study. We, thus, originally constructed the Japanese common SNPs (MAF > 0.15) database based on the genome-wide genotyping data from 45 unrelated Japanese control subjects (23 males and 22 females) [De La Vega. 2005, Genome Res. 15: 454-462]. These data were different from either JSNP or HapMap. Our original database is aimed to focus on gene-centric and intergenic SNPs on chromosome 15q14-22.1. Because our coverage of genetic variation is apparently more limited than that with HapMap in the
target region, we only summarized the comparison of our coverage of genetic variation with JSNP. We created our new Table 2, in which we combined genetic variation with our original Additional file 2 in our revised text. We cited this comment in our revised Results as follows;

[revised Results; line 22 on page 11 - line 6 on page 12]

The distribution of the common 1,317 SNPs was compared with the distribution of 1,578 SNPs from JSNP (Figure 1) and genetic variations based on JSNP in the target region are presented in Table 2. The distribution of SNPs in a bin of 300 kb of gene-centric region showed that the total number of 1,382 SNPs (87.6% in total) from JSNP is larger than our selected 974 SNPs (74.0% in total). However, there are more SNPs in the intergenic region (343, 26.0%) in our study than in the corresponding JSNP (196, 12.4%). Figure 1 showed only one gap around 43.3 Mb in our study, but 3 big gaps of 34.0-37.6, 43.9-45.4, and 51.1-53.2 Mb in JSNP database. Thus, the distribution map of SNPs in our study both in gene-centric and intergenic regions is more uniform than that found in the JSNP database.

[Comment-2]

The two stage strategy would appear to lack power for effects of the size reported. Power in the Discussion is estimated on the full sample size, not on the two stage design, and even then the discussion is uninterpretable as no effect size is provided. This must be incorporated in the manuscript, based on the actual study design (2 stage), and presented with the effect sizes.

Response-2

The two stage approach is highly recommended to minimize the chance of false positive association results [Hirschhorn. 2005, Nat Rev Genet. 6: 95-108], and results in the better outcome using the similar sample size as in our previous reports [Moritani. 2007, Diabetologia 229: 175-184, Kato. 2006, Genomics 87: 446-458]. However, the power for effects of sample size is essential in an association test, especially when a nominal result was obtained as in our study. We agree that our sample size in the two-stage design exhibited rather the poor analytical power.

As you mentioned, we corrected insufficient description about the power for effects of sample size in the two-stage design, and cited this comment in our revised Discussion, as shown below. In addition, we inserted a newly created Figure 3 in our
revised text instead of our original Additional file 10 presenting power calculation in this study.

**[revised Discussion; lines 16 - 24 on page 17]**

Accordingly statistical power, with a MAF of 0.30, our sample size in the first or second stage showed 22-60 or 29-76% to detect the OR of 1.2-1.4 at a significance level of 0.05, respectively (Figure 3A, B). Although the power using the whole sample size was estimated as 45-93%, power is insufficient to detect an association of T2D (Figure 3C). Thus, the limited number of samples in this study could cause the true variants to be missed. To detect with 60-80% power with a MAF of 0.30 and OR of 1.2, 1,400-2,200 each for case and control samples (almost twice as many as those in each stage of this study) are needed (Figure 3D).

[Comment-3]

The introduction and Discussion are quite out of date with regard to known genes; this should be corrected.

**Response-3**

Although we addressed the result of recent GWAS data in our original Background with references [original references 7-12], it was insufficient. With your suggestion, which is the same as in Discretionary Revisions, we incorporated the results of recent GWAS into Background and Discussion in our revised text, and cited a recent reference [Frayling. 2007, Nat Rev Genet. 8: 657-662] as follows;

**[revised Background; lines 4 - 13 on page 4]**

Together with a large amount of data using a candidate gene approach in the past 10 years, and several reports conducting a recent genome-wide association study (GWAS) with comprehensive 300 to 500 k SNPs [2-7] have replicated confirmed 11 genes to T2D with the statistically confident levels in Caucasians [8]. However, the assessment in the Japanese population, in which patients with T2D characteristically have a lower body mass index (BMI) and a lower fasting insulin level than Caucasians [9] remains largely unknown. Thus, Japanese individuals with T2D might have a different genetic background from other populations, and susceptibility variant(s) or gene(s) for the development of T2D in the Japanese can only be identified with genetic assessments.

**[revised Discussion; lines 2 - 5 on page 17]**
We applied a region-wide association test as a different approach from GWAS with new technologies, which has been shown effective for the identification of candidate susceptibility genes in the target region for T2D [17, 18] and rheumatoid arthritis [40, 41] in the Japanese.

[Comment-4]

The authors must show IN THE MANUSCRIPT the exact region of linkage with the 1 LOD confidence interval. Why this region was chosen is not at all clear, but is essential to the interpretation of the paper.

Response-4a

Although we presented the exact region of linkage with the 1-LOD drop interval from STS markers showing positive linkage region in our original Figure 3A and the legend for Figure 3A, the presentation was not clear and illegible in our original format. We collapsed them into a newly created Figure 2 including the exact region of linkage of the 1-LOD drop interval from STS markers, and names of STS markers in our revised Figure 2.

Response-4b

As you pointed out, the explanation why we chose this region was insufficient in our original Background. In this study, we selected the region for the region-wide association test based on following criteria; 1) a region that showed replicated positive linkage to T2D and its related traits with significant LOD levels in at least one of reports in the Japanese, 2) a region reported in the Japanese that replicated those found in different populations. Until now, only three independent genome-wide linkage scans for T2D and its related traits have been performed in the Japanese (revised references 11-13), and identified several putative susceptibility regions. We selected the target region of 15q14-22.1 based on overlapping of the linkage region in the Japanese and other multiple populations (as shown in Additional file 1, this file is the same as our original Additional file 1).

We think that the difference of the genome-wide linkage scans results from the low sensitivity with a limited number of microsatellite markers or subjects. In contrast to the recent status, GWAS to show the appropriate narrow region was not performed at the onset of this study. Thus, in the indirect association test in the
Japanese, we hypothesized that we would be able to efficiently detect the disease susceptibility gene(s) by using evenly-spaced, common SNPs in the target region defined by replicated linkage evidence.

We explained the reasons for the choice of the target region in this study in our revised Background as follows;

[revised Background; line 18 on page 4 - line 5 on page 5]

Among a number of proposed candidate regions for type 2 diabetes and its related traits [10], we selected chromosome 15q14-22.1 based on the replicated linkage signals in the Japanese. In previous three studies of genome-wide linkage scan in the Japanese [11-13], this region showed replicated linkage significance (LOD = 2.41) for T2D [12] and significant evidence of linkage, with a maximum LOD score (MLS) of 3.91, for early onset T2D in the Japanese [11]. This region reported in the Japanese overlaps those found in different populations including Mexican Americans [14] and Pima Indians [15]. These results were replicated in a subsequent study of Mexican Americans [16] (Additional file 1). This candidate region supported by replicated linkage signals is expected to contain susceptibility gene(s), but there has not been a detailed association test on chromosome 15q. The challenge remains to identify the disease susceptibility variant(s) or gene(s) that definitively contribute to T2D. We focused on chromosome 15q14-22.1 at 32.6-51.2 cM as the susceptibility region in the Japanese.

[Comment-5]

Much of the manuscript presents LD data that are available from the HapMap project. Please remove these data - this would be appropriate for supplemental files.

Response-5

As you pointed, we eliminated much of LD data, especially our original Figure 4 presenting LD block around the SNP2140 using HapMap data from our revised text. We changed our original Figure 4 to a newly created Additional file 7 in our revised manuscript.

[Comment-6]
Supplemental files need to be moved into the text. Too much essential data are relegated to "additional files" that in turn are not easy to download and read. Most of these files should be included in the text. A figure showing the linkage with 1 Lod confidence intervals is essential. File 3 is helpful, but the authors never comment on the large number of SNPs tested in Stage 2 with only a single positive result. The number expected by chance in Stage 2 is 5% of the Stage 1 findings. Clearly they are below this threshold. The study is essentially negative, and likely underpowered. These issues must be addressed.

**Response-6a**

We presented many Additional files in our original manuscript. According to your suggestion, which is the same as in General comment, we moved 4 Additional files out of 10 into our revised text, and also revised the names of our original Additional files as follows:

<table>
<thead>
<tr>
<th>original manuscript</th>
<th>revised manuscript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional file 2</td>
<td>Table 2 (newly created) in text</td>
</tr>
<tr>
<td>Additional file 4</td>
<td>Figure 1 in text</td>
</tr>
<tr>
<td>Additional file 5</td>
<td>Figure 2 (newly created) in text</td>
</tr>
<tr>
<td>Additional file 10</td>
<td>Figure 3 (newly created) in text</td>
</tr>
<tr>
<td>Additional file 1</td>
<td>Additional file 1</td>
</tr>
<tr>
<td>Additional file 3</td>
<td>Additional file 2</td>
</tr>
<tr>
<td>Additional file 6</td>
<td>eliminate</td>
</tr>
<tr>
<td>Additional file 7</td>
<td>Additional file 3</td>
</tr>
<tr>
<td>Additional file 8</td>
<td>Additional file 5</td>
</tr>
<tr>
<td>Additional file 9</td>
<td>Additional file 8</td>
</tr>
</tbody>
</table>

**Response-6b**

In response to your comment that a figure showing the linkage with 1 Lod confidence intervals is essential, which is the same as Comment-4 in Major Compulsory Revisions, we prepared our Response-4a to Major Compulsory Revisions. Please refer to that response.

**Response-6c**
As you mentioned, our explanation of the number of positive SNPs was insufficient.

Out of tested 1,317 SNPs in the first stage, 112 (8.5%) showed significant associations with T2D ($p<0.05$) for the allelic and genotypic frequency. Out of these 112 candidate SNPs, only a single positive SNP (SNP2140) replicated association for allele frequency with $p<0.05$ in the second stage. As you mentioned, the number expected by chance in the second stage is about 5% of the SNPs that proceeded from the first stage tests, and the detection of only one SNP may suggest an under power or false positive result. Following your suggestion, we discussed insufficient power in our revised Results. In addition, to improve the presentation about the result of association test, we newly created a revised Figure 2 by collapsing our original Figure 3A into revised Figure 2 in our revised text.

[revised Results; lines 9 - 16 on page 12]

In the first stage discovery panel, we assayed 1,317 SNPs, and 112 (8.5%) showed significant associations with T2D ($p<0.05$ in a $\chi^2$ test) for the allelic or genotypic frequency (Figure 2A). These 112 SNPs were analyzed in the second stage independent replication panel. In the second stage, 11 SNPs yielded the significant associations ($p<0.05$) for the allelic or genotypic frequency (Figure 2B). However, only one SNP, namely SNP2140 (rs2412747), showed a replicated association for allele frequency with $p<0.05$ in both stages (Figure 2C, Additional file 2). These results suggest that our association test is likely underpowered.

[Comment-7]

Correction for multiple testing is not addressed. Genotypic and allelic tests were performed. Was a correction imposed for obesity, or only age?

Response-7

Although we mentioned the result of correction for multiple testing in our original Results, our explanation did not sufficiently address your concern. Following your comment, we added the explanation of correction for multiple testing for genotypic and allelic tests, and a logistic regression analysis for age, sex, and BMI in our revised Methods. We provided the result of FDR analysis in this response letter (Fig. 01 for the reviewing purpose only) and cited this comment in our revised Methods and Results in detail.
Fig. 01

[revised Methods; lines 18 - 22 on page 8]
To evaluate the false-positive, standard Bonferroni’s correction was used [24]. In addition, we used the false discovery rate (FDR) approach [25] as implemented with the FDR control program in the R language [26]. A logistic regression analysis to adjust for age, sex, and BMI was carried out using the SPSS program (ver.12, SPSS Japan Inc., Tokyo, Japan).

[revised Results; line 22 on page 12 - line 4 on page 13]
However, a logistic regression analysis revealed that the association between SNP2140 and T2D was not statistically significant after adjustment for age, sex, and BMI. When the standard Bonferroni’s correction for multiple testing was applied (corrected for 1,317 SNPs for genotypic and allelic tests), the association with SNP2140 was no longer significant. In addition, FDR with 112 SNPs in the second association test was calculated as a threshold value of 0.1 [25]. Unfortunately, none of the 112 SNPs reached the statistical confidence level.

[Comment-8]
The authors should explain on Page 10 why they performed the studies of the LD block around SNP 2140 given the lack of significant association. Is there any reason to pursue URB1? Is it a candidate gene?

Response-8
In all association tests, SNP2140 was weakly associated with T2D exhibiting the nominal $p$ values. These results are statistically non-significant after strictly applying Bonferroni’s correction for multiple testing. The correction for multiple testing is an important issue in genetic association test. However, we also recognized that the Bonferroni’s correction is too conservative. After analyzing a haplotype-based test by a multi-locus approach, which was a more sensitive method for detecting
associations than the assessment of individual SNPs, the nominal \( p \) values were observed. We were unable to completely rule out the possibility of genetic contributions of SNP2140 to T2D in this study. Therefore, we searched around SNP 2140 for the additional putative common variants.

As you suggested, SNP2140 showed the nominal \( p \) value in this study. We only analyzed 904 case versus 890 control samples. We know that our samples are insufficient in number to strictly confirm the susceptibility variant. We would like to leave this issue to be answered in a future study with a large number of samples. We cited this comment in our revised Results as follows;

[revised Results; lines 23 - 26 on page 13]

To find the additional putative common variants, and investigate the fine LD structure around a landmark SNP2140, we searched the public dbSNP databases [39] and genotyped seven additional SNPs (rs16957168, rs10467975, rs11070380, rs10518779, rs6493068, rs7166467, and rs11070392).

[Comment-9]

Please explain the haplotype tests. Was this an evaluation of one haplotype against all others? Did you perform an omnibus test of all haplotypes?

Response-9

As you pointed, we added the explanation in our revised Methods as follows;

[revised Methods; lines 12 - 16 on page 9]

Haplotype frequencies for multiple loci were estimated with the expectation-maximization (EM) algorithm in SNPAlyze ver 5.1 Pro software (DYNACOM, Japan). This procedure was an interactive method based on maximum-likelihood. In addition, the permutation method was used to test the deviation of allelic frequencies of haplotypes on the basis of 10,000 replications.

[Comment-10]

The discussion on pages 11-12 are not clear, and are not adequately explained for the reader to understand them.

Response-10
We apologized for our insufficient description about comparison with LD block with HapMap data and especially EHH and REHH in our original Results on page 11-12. In response to the former description, as shown in our previous Response-5, we added the explanation in our revised Results. Regarding the description about EHH and REHH, we made our explanation simpler in our revised Results. This response relates to Response-3 in Minor Essential Revisions later.

[revised Results; line 18 on page 15 - line 3 on page 16]

It is important to explore the haplotype-specific decay of LD pattern such as disease associated haplotype in the association test. To understand how LD breaks down with increasing distance from a specified core region, we calculated haplotype homozygosity (HH) in a stepwise manner as EHH [33] using the Sweep program. EHH estimates the level of haplotype splitting due to recombination at extended region on both sides of a specified core region. EHH profile also shows an indicator of positive selection. Frequent haplotype having high long-range LD is, thus, supposed to be positively selected.

We applied this method to the landmark LD block including the UBR1 gene, because the size of this LD block was unusually long-range and it was longer than in other populations. A region of approximately 355 kb of the LD block did not show any sign of selection (additional file 8).

Minor Essential Revisions

[Comment-1]

The manuscript is redundant in several places, and needs to be shortened.

Response-1

According to your suggestion, we shortened the length of our revised manuscript as follows;
1) We eliminated the repeated description throughout our revised manuscript.
2) We eliminated both our original Table 2 presenting the summery of LD in the entire target region and Table 3 presenting the result of haplotype analysis.
3) We eliminated both our original Figure 1 presenting study design and Figure 4 presenting comparison of LD structure with HapMap data. In addition, we collapsed our original Figure 3A into a newly created Figure 2 in our revised text.
4) The length of our revised main text was shortened from 4,917 words to 4,456 words.

Our revised manuscript, thus, contains 2 Tables, 3 Figures, and 8 Additional files.

We changed Tables and Figures in our revised text as follows:

<table>
<thead>
<tr>
<th>original manuscript</th>
<th>revised manuscript</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Table 1</td>
</tr>
<tr>
<td>Table 2</td>
<td>eliminate</td>
</tr>
<tr>
<td>Table 3</td>
<td>Additional file 6</td>
</tr>
<tr>
<td></td>
<td>Table 2 (newly created)</td>
</tr>
<tr>
<td>Figure 1</td>
<td>eliminate</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Additional file 4 (newly created)</td>
</tr>
<tr>
<td>Figure 3</td>
<td>eliminate</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Additional file 7 (newly created)</td>
</tr>
<tr>
<td></td>
<td>Figure 1 (newly created)</td>
</tr>
<tr>
<td></td>
<td>Figure 2 (newly created)</td>
</tr>
<tr>
<td></td>
<td>Figure 3 (newly created)</td>
</tr>
</tbody>
</table>

[Comment-2]

Given evidence for age dependency on linkage in this region, was any attempt made to examine the association by age of onset?

Response-2

As you mentioned, the association test for the age of diabetic onset is important. Following your suggestion, we examined the association test between age of diabetic onset. After analysis by age of diabetic onset of >50 years (482) and <50 years (419), none of 1,317 SNPs except for SNP2140 ($p = 0.0004$) was significantly associated with the onset >50 years of age with T2D. In approximately 904 case/890 control samples in this study, when case samples are divided by age of diabetic onset, sample size becomes very small. It is apparently difficult to conclude the significant association in this sample size. Further studies with a larger sample size, involving at least several thousand cases and controls, is required to completely rule out the possible involvement of age of onset for association to T2D.

[Comment-3]
On page 15, the text in the last paragraph is difficult to understand.

**Response-3**

As you pointed out, we rewrote the last paragraph in our revised Discussion as follows;

[revised Discussion; lines 13 - 20 on page 19]

Generally, a difference in allele frequencies, LD block size, and haplotype specific decay of LD pattern between populations may cause the population-specific attributable risks to T2D. Our observations based on HapMap data for a candidate LD block showed differences between the Japanese and other populations, especially one from YRI. Frequent haplotype having high long-range LD suggests recent positive selection, and such signal played a role at loci influencing susceptibility to common complex diseases such as T2D. We could not find any evidence for EHH of the core haplotype, suggesting no sign of selection.

[Comment-4]

Why publish this paper now, if in the paper the authors state that remaining genes should be tested?

**Response-4**

As you commented, our description that SNPs in the remaining genes should be tested for association in our original Discussion was rather confusing and insufficient.

As described in Response-1 in Major Compulsory Revisions, we originally constructed the Japanese common SNPs database. Thus, there were uncovered gene-centric regions in our target region based on current information of genome sequence. We would like to leave the association to be analyzed in a future study with a large number of samples. We eliminated the description from our revised text.

[Comment-5]

The authors should provide some information on the UBR1 gene if they are to make the case that this is a potential diabetes gene.

**Response-5**
Because certain SNPs did not exhibit a strong effect on the risk of T2D within the target region, replication studies in other independent populations are a priority if the consistency of our observations is to be determined. By taking a possibility of observing a positive association for UBR1 with T2D in the future study, we provided the minimized information on the UBR1 gene in our revised legend of Additional file 5 instead of text as follows;

**[revised legend of Additional file 5]**

A landmark UBR1 gene consists of 47 exons spanning approximately 163.2 kb of genomic DNA and the 7.8 kb UBR1 transcript encodes a protein of 1,750 amino acids.

**[Comment-6]**

Table 2 is not particularly useful. Tables in the additional files are more helpful. The simulation and raw p values on Table 3 are extremely similar, and see surprising for 10,000 simulations.

**Response-6**

According to your suggestion, we eliminated our original Table 2 and 3 completely from our revised text.

**[Comment-7]**

I could not successfully open File 5, but the legend suggests that this should be in the manuscript, not supplemental data. The power calculation absolutely must be in the manuscript. Figure 1 does not add that much and could be removed, but the power analysis of this strategy must be shown. Figure 2 B shows the same data twice. Overall, I wish to see the p values across the region, not the block structure. Please add that figure and take these out.

**Response-7**

We apologize for not successfully constructing our original Additional file 5 presenting results on the association test in each stage. According to your constructive suggestion, we added our original Additional file 5 to a newly created Figure 2 and Additional file 10 presenting power calculation to a newly created Figure 3 in our revised text. We also eliminated our original Figure 1 presenting scheme of
two-stage association test and original Figure 2 presenting the pattern of LD from our revised text.

[Comment-8]

Figure 8 B is illegible. Overall this is not helpful. We need to see the original linkage signal. Figure 4 is totally illegible and should not be shown in this format.

Response-8

Although your comment is “Figure 8B is illegible…”, we think that it may be Figure 3B in our original manuscript. We changed our original Figure 3 by totally collapsing it into our newly created Figure 2 in our revised text. In addition, we changed our original Figure 4 to a newly created Additional file 7 in our revised manuscript.

Discretionary Revisions

Please make the abstract match the text. The two stage design must be explained. Insulin sensitivity is not necessarily the first defect in Caucasians; considerable data support an early role for the beta cell. This should be corrected on Page 5. Results from recent GWAS should be incorporated into the text in the introduction and discussion.

Response-1

We apologize for not following the guidelines for the Abstract strictly in our original manuscript. We structured our revised Abstract into Background, Methods, Results, and Conclusions (on page 3).

Response-2

In our original Methods of “Study design of the association test”, we described about the two stage design. However, the explanation was not clear. This comment was cited in our revised Methods as follow; a) we changed the name of section as “Two stage study design of the association test”, b) we explained this strategy in detail.

[revised Method; lines 8 - 16 on page 8]
Two-stage study design of the association test
We performed a first association test in the first stage discovery samples (372 versus 360) for 1,317 SNPs. SNPs exhibiting significant allelic or genotypic association ($p < 0.05$) were further examined in the second stage independent replication samples (532 versus 530). Each stage sample was selected by randomly assigning all samples to two independent panels. For each stage, the association was evaluated by a $\chi^2$ test with a standard $2 \times 2$ allelic frequency and $2 \times 3$ genotypic frequency contingency tables for the status of cases or controls. Finally, the association was evaluated in the combined samples of the first and second stage.

**Response-3**

According to your suggestion, we eliminated the description about insulin sensitivity in Caucasians in our revised Background.

**Response-4**

Although we addressed the result of recent GWAS data in our original Background and references, description was insufficient. According to your suggestion, which is the same as Comment-3 in Major Compulsory Revisions, we corrected our revised original Background and Discussion, and prepared our response-3 in Major Compulsory Revisions.
Response to Referee 2’s comments (1848788880158202_comment. pdf.)

We sincerely appreciate your kind comments. We are very encouraged by your assessment of our manuscript. In our revised manuscript, the revised portions are in red font for your attention. We believe that the original manuscript has been completely changed according to your suggestions. Please find below our response to each of your comments in detail.

General comment

In this manuscript, Yamaguchi Y. et al. have performed an extensive search for genes conferring susceptibility to type 2 diabetes within the candidate locus at chromosome 15q14-22.1 using evenly-spaced common SNP markers in the Japanese. They found modest association of 1 SNP locus with T2D, but the association was not strong enough to overcome the threshold of the significance for the present large-scale analysis. This paper has considerable interest, and provides important information for scientists in the field of genetics for type 2 diabetes.

Major Compulsory Revisions

[Comment]

The authors are recommended to state how completely their evenly-spaced SNPs could cover the present locus. Also please explain the reason why the authors selected those SNPs rather than tag SNPs.

Response-a

We described the target region with our evenly-spaced common SNPs markers in Figure 1 presenting the distribution of evenly-spaced common SNPs markers and Table 2 in our revised text. As we showed in these files, we spaced SNPs with the average of 9,169 bp in the gene-centric region occupying 45% of LD block. In addition, 974 SNPs were mapped within 152 genes (65.8% of the 231 genes), and 343 SNPs were located in intergenic regions. We focused on the gene-centric region, and chose SNPs yielding priorities in their locations and allele frequencies. There are more SNPs in the intergenic region (343) than in the corresponding JSNP region.
However, SNPs in gene-centric region covers approximately 65.8%, and the gene coverage is not complete due to the lack of well-annotated human genome information at the onset of this study. We cited this comment in our revised Results.

(on page 11 - 12)

**Response-b**

Tagging-SNP (tag-SNP) analysis is expected to infer the allelic state of all the common polymorphisms and cover maximum genetic variability, and the utilization of tag-SNP has been proposed as an efficient method to localize susceptibility variants. In contrast to the current situation, there was insufficient information of public SNPs database such as HapMap. Unfortunately, tag-SNPs resource from the HapMap did not supply a sufficient number of tag-SNPs at the onset of this study. We cited this comment in our revised Discussion as follows;

[revised Discussion; lines 2 - 5 on page 19]

Even though tag-SNPs was not available at the onset of this study, tagSNPs with LD information and a high-density SNP map from the HapMap database could afford the useful resources.
Response to Editorial office comments

Please also ensure that your revised manuscript conforms to the journal style (http://www.biomedcentral.com/info/ifora/medicine_journals). It is important that your files are correctly formatted:

- your abstract must be structured into Background, Methods, Results, and Conclusions.
- you must include a ‘Competing interests’ section between the Conclusions and Authors’ contributions. If there are none to declare, please write ‘The authors declare that they have no competing interests’.

Responses-1

According to the comments by the Editorial office, we structured our revised Abstract into Background, Methods, Results and Conclusions. (on page 3)

Responses-2

According to the comments by the Editorial office, we added a ‘Competing interests’ after the Conclusions. (lines 11 - 12 on page 20)