Author’s response to reviews

Title: Schizophrenia susceptibility and NMDA-receptor mediated signalling: an association study involving 47 tagSNPs of DAO, DAOA, PPP3CC, and DTNBP1 genes

Authors:

Emilio Sacchetti (genetica@fatebenefratelli.it)
Catia Scassellati (cscassellati@fatebenefratelli.it)
Alessandra Minelli (alessandra.minelli@med.unibs.it)
Paolo Valsecchi (labgen@fatebenefratelli.it)
Cristian Bonvicini (cbonvicini@fatebenefratelli.it)
Patrizio Pasqualetti (patrizio.pasqualetti@afar.it)
Alessandro Galluzzo (scat_08@hotmail.com)
Rosaria Pioli (rpioli@fatebenefratelli.it)
Massimo Gennarelli (gennarelli@fatebenefratelli.it)

Version: 3 Date: 8 October 2012

Author’s response to reviews: see over
Author's covering letter for initial submission

Title: Schizophrenia susceptibility and NMDA-receptor mediated signalling: an association study involving 32 tagSNPs of DAO, DAOA, PPP3CC, and DTNBP1 genes

Authors:

Version: 1 Date: 24 September 2012

Comments: see over
Dear Dan Rujescu,

We thank the reviewers for their constructive comments and suggestions. In the following, we first summarize each reviewer comment in underlined text and then give our response. The modifications in the manuscript have been indicated in *italics*, whereas the sentences eliminated have been indicated with the strikethrough. In general the manuscript has been revised at the light to emphasize that it is the first replication study performed in an Italian population and we have better underlined the use of the phenotype dissection strategy to strengthen the informative power of genetic association studies.

**Reviewer #1: Erik Jönsson**

1. **RE:** The major limitation of the study is the relatively low number of subjects. Therefore, the authors should emphasize - both in the abstract and in the conclusions - that the study results are preliminary and that there is need for replication.
   We agree with the reviewer and we have emphasized this limitation in the abstract (pages 4-5) and in the discussion (page 18).

2. **RE:** In the first paragraph of page 17, the authors state that the possibility of type I errors is low because there are still significant associations after correction for multiple testing. However, there are plenty of genetic association reports which never have been replicated irrespective of solid results in single studies. Also, using a strict Bonferroni correction (32 SNPs analysed for total sample, gender, clinical subtype and age of onset + unknown number of haplotypes analysed for the same parameters) would give a substantial number of analyses performed, probably making none of the analyses remaining significant. Therefore, the authors should be more careful in this respect.
   We agree with the referee. Thus we have better clarified how we have performed the different corrections (see the answer of the next comment) and because, mainly in relation to gender, the results remain still significant also after the different corrections (including corrections according to the different sliding windows approach), we have left this sentence, even though it has been simplified (page 18).

3. **RE:** Please describe more in detail how the correction for multiple testing was performed.
   The corrections for multiple testing were conducted according to the PLINK software consisting of Bonferroni correction for allele frequencies analyses and 10000 permutation for genotype frequencies analyses (page 10). In details the corrections were performed considering each gene separately (6 SNPs for DAO, 6 SNPs for DAOA, 8 SNPs for PPP3CC, 12 SNPs for DTNBP1). In the haplotype analysis by using the Haploview program the corrections were performed according to 10000 permutation (page 10) and, according to another referee, we have also added the corrections in relation to the sliding window approach (page 10). Also for these analyses, we have considered each gene separately. In the results section we have added how these corrections were performed. In details, there were 9 sliding windows for DAO gene. According to the LD panel (see supplementary table 2), we conducted the analyses considering the rs3741775 along with the other single SNPs, and rs3918347 along with the other single SNPs and between them, setting a p value significant of 0.006 (0.05/9). The patient-control difference in the allele estimated GT diplotype (rs2070586 A/G-rs3741775 G/T) in females sample remained still significant (p=0.003). It is also
important underlying that the significance in females sample is found also in the genotype frequencies and in the carriers of GT diplotype and it is supported by a significant interaction between GT diplotype carriers and gender. Moreover the p values were corrected also according to permutation analyses (see table 1).

For PPP3CC, there were 11 sliding windows. According to LD panel (see supplementary table 2), we conducted the analyses considering rs4872499 along with two SNPs in the Block 1 and along with all SNPs of the Block 2, setting the p value to 0.005 (0.05/11). The results related to CAG triplotype (rs4872499 T/C-rs11780915 A/G-rs13271367 G/A) remain still significant in the total sample (p=7x10^{-4}), whereas in the females sample the difference was lost (p=0.006). However the significance in females sample remains in the genotype frequencies and in the carriers of CAG triplotype and it is supported by a significant interaction between CAG triplotype carriers and gender. Moreover the p value remains significant according to permutation analyses (see table 1).

On the basis that rs6459409 presented significant results as single SNP and along with all other SNPs (see supplementary table 2), we calculated 11 sliding windows for DTNBP1, setting the p value to 0.005 (0.05/11). The patient-control difference in the allele estimated CT (rs6459409 C/T-rs9476886 T/C) in males sample remained still significant (p=0.003). It is also important underlying that the significance in males sample is found also in the genotype frequencies and in the carriers of CT diplotype and it is supported by a significant interaction between CT diplotype carriers and gender. Moreover the p values were corrected also according to permutation analyses (see table 1).

(4) RE: The authors should give the power of their sample for an OR of 1.1-1.2, which is more in line with the literature of recent schizophrenia association results of common polymorphisms. The low power to detect differences of this effect size should also be acknowledged among the limitations of the study.

As the Reviewer indicated, our study does not have an adequate power to detect OR as low as 1.2. However, because of we have based our study on the phenotype dissection strategy in order to strengthen the informative power of genetic association studies, we have computed the power for higher OR. Moreover, when a significant effect is reported, the power is less relevant, whereas the effect size dimension is more important.

(5) RE: Among the limitations of the study the authors should also include the limited covering of some parts of the investigated genes giving that one third of the investigated SNPs was not available for analysis.

We had tested the quality control by using another technique (Affymetrix Human Mapping GeneChip 6.0 arrays) on a subsample of approximately half of the whole sample with a genotyping completion rate of approximately >95%.

We have clarified better this issue (page 9), reporting that the SNPs analyzed were 32. In details for the referee three (rs9558574 in DAOA, rs17733133 in PPP3CC, rs10456775 in DTNBP1) were excluded as it showed a significant deviation from HWE; whereas twelve (rs7967441 in DAO; rs2153674 in DAOA and; rs2469758, rs2461482 and rs17671456 in PPP3CC, and rs9370823, rs2619539, rs7768128, rs760761, rs2619521, rs13192791, rs7752070 in DTNBP1) were excluded for a completion rate <90%.

(6) RE: I suggest that the authors add supplementary figures with the linkage disequilibrium structure including LD blocks for each of the genes. The authors should also state how many haplotype analyses that were performed for each gene.

We have added as supplementary figures the linkage disequilibrium structure including LD blocks for each of the genes. Moreover we have better explained how many haplotypes sliding window analyses were conducted (see the answer at (3) RE).
The authors write: “We considered significant the allele haplotype frequencies <5%”. Do the authors mean that they considered a p-value equal or less than 5% significant in the haplotype analyses? Or do they mean anything else? Please explain. 

We mean that only the allele frequencies of the haplotypes > 5% were included in the analyses, those <5% the haplotypes were excluded. We have explained better this sentence (page 10).

In the Results section (page 13, 3rd paragraph) the authors refers to the analysis of variance (ANOVA). However, I cannot find this analysis mentioned in the Methods section. 

In the Methods section ANOVA is mentioned as Factorial analyses of variance and we have added in the page 10 the acronyms.

Gene abbreviations should be in italic font, e.g. DAO, DAOA, PPP3CC and DTNBP1. 

We have explained better this sentence (page 10).

Abbreviations should be explained at their first appearance. This refers to e.g. NMDA, DAO, DAOA, PPP3CC, DTNBP1, SNP, RRR, OR, and CI in the abstract and DSM-IV-TR, OR, and CI in the main text.

We have explained better this sentence (page 10).

The authors should give references for DSM-IV-TR, SCID-CV, MINI and MMSE.

We have added these references.

Concerning the haplotype analysis, a significant effect of the estimated CAG...” I suggest: “Concerning the haplotype analysis, there was a significant effect of the estimated CAG ...”

We have explained better this sentence.

Table 1. The legend could be simplified, e.g.: “Estimated haplotype distributions for selected D-amino acid oxidase (DAO), protein phosphatase 3 catalytic subunit gamma isoform (PPP3CC) and dystrobrevin-binding protein 1 (DTNBP1) single nucleotide polymorphisms in patients with schizophrenia and control subjects.” It is not necessary to state the specific polymorphisms in the legend, because they are given in the table.

We have explained better this sentence.

Reviewer #2: Daimei Sasayama

Major comment: The sample size was not large enough to achieve sufficient statistical power to detect small effect size, especially when stratified by gender and diagnostic subtypes. Ideally, the positive findings presented in this study should be confirmed using an adequately powered replication sample.

We agree with the referee, however we have emphasized that this work represents a replica of the well known candidate genes for schizophrenia in the Italian population and among the limitations we have commented the low statistical power of the our sample, especially after stratification by gender and diagnostic subtypes.

However, because of we have based our study on the phenotype dissection strategy in order to strengthen the informative power of genetic association studies, we have computed the power for higher OR. Moreover, a significant effect is reported, the power is less relevant, whereas the effect size dimension is more important.
(1) RE: In Genotyping of the Methods section, I do not understand what the authors mean by the fourth criterion for the inclusion of the SNPs in the analyses (i.e. including SNPs with “successful genotyping of at least 50% of the SNPs”). Is this a criterion for inclusion of the sample (i.e. including samples with successful genotyping of at least 50% of the SNPs)?

We agree with the referee and because a typing error, we have deleted the fourth criteria of inclusion of SNPs in the analyses (page 9) and we have simplified the sentence. According to the other referees, we had tested the quality control by using another technique (Affymetrix Human Mapping GeneChip 6.0 arrays) on a subsample of approximately half of the whole sample with a genotyping completion rate of approximately >95%.

We have simplified the work showing that the analyses were conducted on 32 SNPs.

(2) RE: In Genotyping of the Methods section, the authors listed SNPs excluded from the analyses. How many of the SNPs were excluded due to low call rates? How many due to deviation from HWE? How many due to low MAF? If sample replicates have not been included in the assay to ensure consistent genotypes, the quality of genotyping can only be assessed by call rates and deviation from HWE. Thus, it is important to state whether these SNPs were excluded due to low quality typing or simply due to low MAF.

We have clarified better this issue (page 9), reporting that the SNPs analyzed were 32.

In details for the referee three (rs9558574 in DAOA, rs17733133 in PPP3CC, rs10456775 in DTNBP1) were excluded as it showed a significant deviation from HWE; whereas twelve (rs7967441 in DAO; rs2153674 in DAOA and; rs2469758, rs2619521 and rs13192791, rs7752070 in DTNBP1) were excluded for a completion rate <90%.

(3) RE: What algorithm did the authors use to select the tagging SNPs?

For tagging selection we used SNPbrowser version 3.5 (de La vega et al 2006) (page 9)

(4) RE: In Table 1, show the number of male and female samples included in the haplotype analyses.

We have added these numbers in Table 1

(5) RE: Line 11 of Page 6: Correct this sentence: “A recent convergent functional genomics in schizophrenia has identify top genes among which DTNBP1.”; Line 10 of Page 14: non-GAC should be corrected to non-CAG.

We have corrected as suggested

Discretionary Revisions

(6) RE: The Supplementary Table showed that none of the examined SNPs were significantly associated with schizophrenia susceptibility. It would be more informative if odds ratio and 95% CI are shown in the table.

We have added OR and CI as requested.

(7) RE: Is it possible to show the results of the secondary analyses also in a supplementary table?

The major significant results from the secondary analyses come from the gender stratification that are reported in the table 1. Thus we have reported the significant results related to interactions evaluation and to the stratification according to paranoid subtype

Reviewer #3: Jubao Duan
(1) RE: Without discussing the most recent genome-wide association studies of schizophrenia, in particular the Psychiatric Genetic Consortium SZ GWAS, the background introduction seems very weak to make the case of studying association of these 4 genes in a small Italian sample. The rationale of the study should not be simply based on meta-analyses of classical candidate gene studies. A post-GWAS genetic association study of schizophrenia in relatively small sample should be better justified.
We have inserted in the introduction section a discussion about the most recent genome-wide association studies of schizophrenia (page 6).

(2) RE: QC issues are major concerns: (1) “successful genotyping of at least 50% of the SNPs” seems to be a too relaxed cut-off for sample inclusion, which may create sample bias or false positives. This could have substantially affected the haplotype analysis. The cut-off should be at least 90%.
We agree with the referee and because of a typing error, we have deleted the fourth criteria of inclusion of SNPs in the analyses (page 9) and we have simplified the sentence.

(3) RE: (2) 15 out of 47 SNPs were excluded from the analyses due to failed genotyping, then why the title of the paper is “…47 tag SNPs.”. More importantly, if 15 out of 47 tag SNPs were failed (which should not be the quality standard of a typical SNPlex genotyping experiment), it could indicate some potential systematic technical problems.
We had tested the quality control by using another technique (Affymetrix Human Mapping GeneChip 6.0 arrays) on a subsample of approximately half of the whole sample with a genotyping completion rate of approximately >95%.
We have clarified better this issue (page 9), reporting that the SNPs analyzed were 32. In details for the referee three (rs9558574 in DAOA, rs17733133 in PPP3CC, rs10456775 in DTNBP1) were excluded as it showed a significant deviation from HWE; whereas twelve (rs7967441 in DAO; rs2153674 in DAOA and; rs2469758, rs2461482 and rs17671456 in PPP3CC, and rs9370823, rs2619539, rs7768128, rs760761, rs2619521, rs13192791, rs7752070 in DTNBP1) were excluded for a completion rate <90%.

(4) RE: (3) any SNPs showed HWE departure? (4) is there potentially population substructures in the studied sample? If not technically feasible to control for this potential confounder, it should be at least discussed as limitation of the conclusion.
Because we have a very few SNPs, one for each gene (rs9558574 in DAOA, rs17733133 in PPP3CC, rs10456775 in DTNBP1) showing deviation from HWE, we exclude the potential population substructures

(5) RE: As genotyping QC is not up to current standard, it is necessary to confirm some of the positive findings (e.g., association with PPP3CC Haplotype (rs4872499|rs11780915|rs13271367) by re-genotype those SNPs by other genotyping techniques (e.g., TaqMan) and compare the concordance of genotyping results. Alternatively but less convincingly, the authors could address this concern by only analyzing a smaller subset of samples with genotyping completion rate >90%.
We had tested the quality control by using another technique (Affymetrix Human Mapping GeneChip 6.0 arrays) on a subsample of approximately half of the whole sample with a genotyping completion rate of approximately >95%.

(6) RE: “For the PLINK software, a nominal level of significance of p=0.05 was accepted and corrected according to the permutation procedure and Bonferroni’s correction.” It is not clear whether this statement refers to HWE test or allele r association test.
We have specified that this sentence is referred to all genotype/allele association tests.
The haplotypic analysis: It is not clear why two-marker haplotypes were reported for some genes but 3-marker haplotypes for others? Assuming that they were tested systematically by different window size and with different sample stratification, the multiple testing correction should be better addressed. For example, is a simple permutation correction for that particular reported haplotype sufficient? Should the authors consider correcting for the total # of tests for other tested haplotypes?

We have clarified this issue specifying better that we have performed both the estimate haplotype frequencies for each block of linkage and sliding window haplotypes. In the haplotype analysis by using the Haploview program the corrections were performed according to 10000 permutation (page 10) and also we have added the corrections in relation to the sliding window approach (page 10). For these analyses, we have considered each gene separately. In the results section we have added how these corrections were performed. In details, there were 9 sliding windows for DAO gene. According to the LD panel (see supplementary table 2), we conducted the analyses considering the rs3741775 along with the other single SNPs, and rs3918347 along with the other single SNPs and between them, setting a p value significant of 0.006 (0.05/9). The patient-control difference in the allele estimated GT diplotype (rs2070586 A/G-rs3741775 G/T) in females sample remained still significant (p=0.003). It is also important underlying that the significance in females sample is found also in the genotype frequencies and in the carriers of GT diplotype and it is supported by a significant interaction between GT diplotype carriers and gender.

For PPP3CC, there were 11 sliding windows. According to LD panel (see supplementary table 2), we conducted the analyses considering rs4872499 along with two SNPs in the Block 1 and along with all SNPs of the Block 2, setting the p value to 0.005 (0.05/11). The results related to CAG triplotype (rs4872499 T/C-rs11780915 A/G-rs13271367 G/A) remain still significant in the total sample (p=7x10^{-4}), whereas in the females sample the difference was lost (p=0.006). However the significance in females sample remains in the genotype frequencies and in the carriers of CAG triplotype and it is supported by a significant interaction between CAG triplotype carriers and gender. Moreover the p value remains significant according to permutation analyses (see table 1).

On the basis that rs6459409 presented significant results as single SNP and along with all other SNPs (see supplementary table 2), we calculated 11 sliding windows for DTNBP1, setting a p value to 0.005 (0.05/11). The patient-control difference in the allele estimated CT (rs6459409 C/T-rs9476886 T/C) in males sample remained still significant (p=0.003). It is also important underlying that the significance in males sample is found also in the genotype frequencies and in the carriers of CT diplotype and it is supported by a significant interaction between CT diplotype carriers and gender.

(8) RE: Minor commen: please be more careful in identifying spelling errors and formatting problems of the manuscript, e.g., Page 4, in “(p=3x10^{-4} OR=0.46 95% CI:0.30-0.70)”, there should be a comma between p-value and OR, CI.

We have corrected as suggested