Author's response to reviews

Title: High frequency of known copy number abnormalities and maternal duplication 15q11-q13 in patients with combined schizophrenia and epilepsy

Authors:

Larissa Stewart (ls12@bcm.edu)
April L. Hall (burant@bcm.edu)
Sung-Hae L Kang (skang@bcm.edu)
Chad A Shaw (cashaw@bcm.edu)
Arthur L Beaudet (abeaudet@bcm.edu)

Version: 3 Date: 3 August 2011

Author's response to reviews: see over
Guian Paolo Declaro  
on behalf of Prof David Collier  
Journal Editorial Office  
BioMed Central

To the editor,

I apologize for the delay in revising this manuscript. In this case, I think that the manuscript is substantially improved in response to the reviewers’ comments. We are still trying to obtain specific information about other studies which might have included these samples, but we do not have responses to some of our inquiries. I think that we have some differences of opinion as to whether chromosomal microarray analysis is suitable for use in the clinic at this time.

Reviewer 1 Kirov  
Reviewer. 1) Have patients from this NIMH sample been included in other publications on CNVs? If not, they should attempt to contact the authors of any papers reporting on these collections and clarify this.

Response. We have contacted various authors but we have not received answers to date. We will continue to try to clarify this. We believe that some of these samples likely were studied in the paper published by Levinson et al. (PMID: 21285140), and we have sent an inquiry in this regard to Dr. Levinson, but he has not responded yet. Otherwise we are unaware of any publications that would include this sample. We have added a statement to the methods in this regard as follows. “4+.

Reviewer. 2) The authors are unsure which CNVs are accepted as increasing risk for schizophrenia, e.g. NRXN1 is in or out for some of their analyses. In addition they seem unaware of several recent papers of CNVs and schizophrenia, which implicate new loci, and clarify the rate of CNVs in this disorders. I would suggest to change the list of accepted findings to: 1q21.1, NRXN1, 3q29 (Moreno de Luca 2010), 15q11-q13 (maternal duplications, Ingason et al 2011), 15q13.3, 16p11.2(dupl, McCarthy) and 22q11.2. The paper by Levinson et al (2011), which the authors don't cite, should also give them established frequencies of most of these CNVs in cases and controls. They can then re-calculate their findings, which are likely to get even stronger.

Response. The comments of the reviewers and recent references are helpful. We have now stated explicitly which CNVs we consider to be well established as assocaited with schizophrenia as follows. “For this assessment, we accepted the following CNVs as being known to be associated with schizophrenia based on published data [12,19]: deletions of 1q21.1, 3q29, 15q13.3, 22q11.2, and NRXN1 and duplications of 15q11-q13 (maternal), 16p11, and 16p13.3. We included maternal duplications of 15q11-q13 in this group because of the extensive literature linking it to autism and to epilepsy and based on recent reports of its occurrence in schizophrenia [14, 52].” We have recalculated the data and the significane is indeed stronger.
Reviewer. 3) The use of bipolars with epilepsy in this paper is a bit confusing. If the authors wanted to include them with the schziophrenics from the start, then they should be reported together, and the calculations done for the two sets. This is probably not a good idea, but if they are reported separately, the sample size is too small. It looks best not report these samples at all.

Response. We have not tried to make any calculations for the bipolar data, because the numbers are small, and we have moved most of the information to Supplementary on line materials.

Reviewer. Minor suggestions:
The "artifact" CNV in Table 5 appears a genuine artifact and should be excluded.
Give again the numbers of cases/controls in the Methods. By the way, are the controls useful in the analysis?

Response. We have left these findings in the Table since they are real and would be detected by other groups using the same samples. We do not want to introduce discrepancies as to what various labs detect when they analyze a sample.

Reviewer. The 1st paragraph in the Background section refers to frequencies of CNVs of 0.0021% and others, they are actually 0.21% (correct the decimal points). The authors should be aware that the Stefansson and ISC papers had an overlap, so the true frequencies are a bit lower.

Response. The error has been corrected.

Still in the Background, the authors have to cite the papers by Ingason et al 2011 about the role of maternal duplications at 15q11-q13, accept the established role of 16p11.2 and 3q29.

Response. This citation is valuable and has indeed been added.

Table 2 lists "Cus2" and "Cus3", there is a discrepancy with the legend to the table.

Response. This problem is corrected.

Reviewer 2 Gershon

Reviewer. Some of the clinical and analytical methodological choices made by the authors detract from the acceptability of their results. The most serious problems are with diagnosis. Epilepsy in patients is based on a self-report questionnaire. The NIMH controls were collected through an anonymous web-based questionnaire and not by direct interview and medical records exam. It is well-known that many psychiatric problems were not well-screened for, and the data available have been rescreened in previous publications. In this paper, rescreening of controls for psychiatric problems was based on ad hoc criteria. Other studies have screened out smaller numbers of the controls. Nonetheless, since the great majority of NIMH controls was screened...
out by these authors, a comparison of current results on patients with results on all the NIMH controls would have been appropriate.

**Response.** We are aware of these limitations and have stated them in the text as follows. “There are weaknesses in the current study including the use of cell line DNA rather than blood derived DNA, self-reporting of epilepsy findings, less than optimal phenotypic information, lack of availability of parental DNA, and lack of complete matching of ethnicity of cases and controls.” We have defended our rationale for rescreening and stated it in the text as follows. “We chose to exclude controls with even modest or questionable behavioral findings in an attempt to exclude individuals who might have mild manifestations of a deleterious CNV. The rationale was to obtain a control group with under-representation of any behavioral abnormalities.”

**Reviewer.** One might also ask about continental origins of the samples. The included controls are noted to be Caucasian; is this true of the patients as well? The use of lymphoblastoid cell lines is explained as necessary because whole blood was not available for the vast majority of specimens in the repository. The reason an explanation is needed is that lymphoblastoid cell lines are well known to create CNV artifacts, as noted recently in the WTCCC 2010 CNV article (doi:10.1038/nature08979) with illustrative comparisons of blood and lymphoblast CNV detection on a custom CNV array. Since the data are meager in any case, this reviewer is concerned about possible artifactual results, even though both patients and controls had the same type of samples available. The major conclusion of the paper is, “We had a 4.5 fold (3% vs. .66%) increase detection rate of known causative CNVs in combined schizophrenia and epilepsy patients,” [vs. patients in a previous report by Vassos et al. 2010]. Even though the Vassos paper is a meta-analysis with larger samples, the use of historical controls is always suspect. One worries whether the methods are comparable.

**Response.** Ethnicity data has been added for all samples.

**Reviewer 3 Rucker**

**Reviewer.** The manuscript is generally skewed in favour of the pathogenicity of the CNVs presented, which I do not think is an accurate reflection of all the evidence. There are a significant number of typographical errors. Some sentences are needlessly wordy, and need re-writing. The manuscript could also be better organized, with less repetition. It requires more detail and references in the methods section, especially surrounding the re-use of arrays, and the use of cell-line derived DNA. The discussion section contains too much biological ‘padding’ text about putative biological effects of the genes within the CNVs in question, rather than presenting a balanced view of the findings, limitations and advantages of the study.

**Response.** Although we believe that many of the CNVs “cause” schizophrenia with high or moderate penetrance in genetic terminology, we have removed the use of “cause” throughout the manuscript and generally replaced it with language such as “associated with” as an accommodation to this reviewer. We have made every effort to correct any typos, reduce wordiness, and eliminate needless repetition. Citation of references regarding individual genes is described as “padding” but we feel that this is justified in the interest of thoroughness. Some of this has been moved to Supplemental on line materials especially for bipolar findings. We have
attempted to now provide a balanced summation in the last paragraph of the discussion immediately before conclusions. We have addressed the reuse of arrays in the text as follows. “The reuse of clinical arrays provided a major cost reduction for these studies and allowed for comparison to a large body of clinical data using the same array design. We have experience that the used arrays detect known variants reliably if the Agilent DLR score is less than 0.30. These data indicate that false negative results would be rare. There is no significant concern regarding false positive results, because all CNVs were validated using new arrays with coverage suitable for the putative CNV.”

**Reviewer.** 2 Apropos the above, it is also of major concern that there is no direct comparison of the frequencies of CNVs seen in cases with the author’s screened controls. Why is this not presented, when you have gone to all the trouble of collecting and analysing them? I imagine that the statistical comparison between the frequencies of CNVs between the cases and screened control groups is in fact non-significant. But if this is the case, then why not report it?

**Response.** This is an excellent suggestion, and the comparison is presented in the abstract as follows. “We detected 10 schizophrenia plus epilepsy cases in 235 (4.3%) with the above mentioned CNVs compared to 0 in 191 controls (p = 0.003). Other likely pathological findings in schizophrenia plus epilepsy cases included 1 deletion 16p13 and 1 duplication 7q11.23 for a total of 12/235 (5.1%) while a possibly pathogenic duplication of 22q11.2 was found in one control for a total of 1 in 191 (0.5%) controls (p = 0.008).”

**Reviewer.** 3 On the assumption that (2) is done, and more broadly, the methods section requires much more detail, especially with regard to the nature of the control samples (there are no numbers), on what basis the items used for exclusion were used, how many DNA samples were derived from cell-lines and how many from lymphocytes, and the evidence base for repeatedly using arrays. The authors say in the results that they used 191 controls (but this isn’t stated in the methods), but then also that they excluded 1,284 out of 1,920 control samples. So what happened to the rest? 1,284+191 does not equal 1,920. With regard to the aCGH, information on QC needs to be covered, especially in light of the repeated use of arrays, which is a methodology I haven’t come across before, and is also not referenced.

**Response.** We excluded 1284 samples out of 1920 leaving 636 control samples for study. Then 191 of the 636 were selected at random for study.

**Reviewer.** I don’t understand the rationale for the follow-up. You have used array-CGH. Why the need to validate on other, generally inferior array platforms? No results for the validation appear to be presented in any event. Perhaps this is done to mitigate against problems caused by the re-use of the Agilent arrays, but this isn’t stated.

**Response.** This is validation was done in part to mitigate against false positives related to reuse of arrays as now stated in the text. The validation arrays were higher resolution, often custom designed, not reused, and superior to the first array analyses.

**Reviewer.** The discussion presents lots of biological data about the putative plausibility of these variants in causing the phenotypes being investigated. But I think the more interesting question is whether or not microarray based genetic screening is appropriate in these groups of individuals at
all. With the low pick up rate, low specificity/sensitivity, uncertain causative mechanisms and scarcity of any data relating these variants to treatment response or prognosis (let alone diagnosis), it is unlikely that they will be in the near future. The conclusions mention this issue, but it is not discussed in any length in the discussion, which is incongruous.

**Response.** We strongly disagree with the reviewer regarding clinical utility of array testing. For decades, cytogenetic testing has been performed on children with mental retardation with similarly low pick up rates. Although the pick up rate is low, it is steadily increasing as higher resolution arrays are put into clinical use. In cases where a well characterized CNV is detected, this can clarify the diagnosis, inform genetic counseling, and potentially lead to CNV-specific interventions. This is now discussed in the conclusions paragraph.

**Specific comments.**

**Reviewer.** 1 Page 4, first para.

**Response.** We have eliminated the “cause” terminiology throught the manuscript.

**Reviewer.** 2 Page 4, second para.

**Response.** Again we have eliminated the “cause” terminiology throught the manuscript.

**Reviewer.** 3 Page 5, para 2.

**Response.** We have eliminated repetition.

**Reviewer.** 4 Page 5, para 2 and 3. I think these two paragraphs could be combined and re-written for greater clarity. I do not understand the basis for the exclusion of control subjects. This needs more detail. I do not think that the commentary about the high exclusion rate should be in the methods section, but left for the discussion (where it is, in fact, repeated).

**Response.** The rescreening of controls is discussed and defended above.

**Reviewer.** 5 Page 6, para 1. “All cases and controls were hybridized with the same male reference DNA isolated from fresh blood”. OK, but this is only part of the protocol. Needs more detail.

**Response.** This is routine methodology in array CGH. We have expanded the description and referenced methods appropriately. We have added the fact that the control DNA is from a healthy Caucasian male.

**Reviewer.** 6 Page 6. The re-use of microarray slides is not something I have come across before.

**Response.** This is addressed above.

**Reviewer.** 7 Page 6-7. “Scanned images were quantified using Agilent Feature Extraction software (v10.7.3.), then analyzed for copy-number change using our in-house
analysis package, as described previously[18-20]” I think some brief detail of the in-house package used is required.

**Response.** We believe that this is relatively complex and not easily described briefly and is best dealt with by reference to published papers.

**Reviewer.** 8 Page 7, para 1. “Common calls, as defined by being seen more than 20 times in 20,000 cases analyzed by the medical genetics laboratories at Baylor College of Medicine, were also not included”. But this is not the definition of a common CNV (or CNP), which is usually defined as a variant occurring in more than 1 in 100 members of the population. Why the need for a new definition?

**Response.** We are not proposing any new definition. We are simply stating that certain CNVs are seen very frequently in the clinical laboratory where they are usually inherited from a healthy parent, and we did not report CNVs of this type here.

**Reviewer.** 9 Page 7, para 1. “All coordinates are in hg19”. A non-geneticist would not understand what this means.

**Response.** A sentence of explanation has been added.

**Reviewer.** 10 Page 9, para 1. “Based on these results, clinicians can reasonably expect a pathological CNV detection rate of about 5% if they study patients with both schizophrenia and idiopathic epilepsy”. This suddenly introduces a new agenda. Are you saying that your results are translatable into the clinic? If so, I think this is going too far. Why would clinicians (and patients) currently want to know about these variants (over and above the fact that it is interesting) given that A) we do not know if they are associated with treatment and outcome (let alone diagnosis) and B) There is nothing that can be done to rectify the abnormality. I do not think it is reasonable to say that the pick-up rate of 5% is likely to be a reasonable expectation for a clinician in any event. The authors comment that “This is the highest detection rate for pathological CNVs in any schizophrenia population studied to date”. So it's pretty likely then that the real rate lies somewhere between their own estimates and others.

**Response.** We are saying that these results do translate into the clinic. We simply do not agree with this reviewer on this topic. No one else has reported a study of patients with both schizophrenia and epilepsy, and we believe that our number are significant and are the best available.

**Page 9, para 1.** “Although deletion 15q11.2 is considered to be associated with schizophrenia by some authors, based on our clinical laboratory experience, we still question this association, and if it is correct, the penetrance is relatively low”.

**Response.** We have provided a reference for the low penetrance.
Page 14, para 1. “The findings in the cases discussed above were compared to findings in the NIMH controls screened to reduce the frequency of psychiatric symptoms”. Where is this comparison? I see no comparison table in the text. Where is the p value?

Response. These data are now included as presented above.

Page 14, para 3. “Most clinicians accept that this duplication is capable of causing abnormal neurodevelopmental phenotypes and heart defects” [54]. I don’t think this is an accurate précis of the reference, which says “Whether duplication 22q11.2 could be a non-pathogenic polymorphism or a real syndrome with a great clinical variability and reduced penetrance is uncertain at this time”

Response. We have revised the text to better address the information regarding duplication 22q11.2 as follows. “Duplications of 22q11.2 show incomplete penetrance and cause a highly variable phenotype of neurocognitive dysfunction [50,51], although there is at least one report suggesting an association with schizophrenia [10].”

Page 15. The conclusion paragraph does not bear much relationship to what has been presented or discussed. I think the words ‘strongly’ and ‘disease-causing’ should be removed from the first sentence. The second sentence mentions that the authors ‘argue’ for the use of CMA in the clinic, but the arguments aren’t stated in the text.

Responses. The conclusions paragraph is substantially rewritten. We have used the terminology disease-associated rather than disease-causing. We still do use the phrasing “strongly suggest” and state the following “The data presented here strongly suggest that disease-associated CNVs were present at a significantly higher frequency in cases with both schizophrenia and epilepsy compared to control samples from the same repository.” With the statistical analyses now presented, we believe that this statement is justified.