Reviewer's report

Title: A Method for Determining Haploid and Triploid Genotypes and their Association with Vascular Phenotypes in Williams Syndrome and 7q11.23 Duplication Syndrome

Version: 0 Date: 24 Jan 2018

Reviewer: Bernt Popp

Reviewer's report:

In their manuscript "A Method for Determining Haploid and Triploid Genotypes and their Association with Vascular Phenotypes in Williams Syndrome and 7q11.23 Duplication Syndrome" the authors describe their approach of performing a case-only region-based association study with microdeletions/duplications in the 7q11.23 region with aortic stenosis/dilatation stenosis in 25 and 13 individuals, respectively.

I will first go through and answer the points recommended for review [PMID: 10391655] and then list points which could be addressed in a possible revision.

(1) Importance: Both Williams Syndrome and 7q11.23 Duplication Syndrome are rare diseases with considerable variability in the affected individuals phenotype. This is both true for the frequent and rare symptoms. Despite advances in genomic analyses this variability can in most cases not be explained today. Knowing the risk of an affected person to develop certain life-threatening symptoms (e.g. cardiac) is important for individual management. Thus approaches addressing the phenotype-genotype gap in rare entities, like the present manuscript, are of scientific interest.

(2) Originality: While several genomic regions show recurrent reciprocal deletions and duplications with partially mirrored phenotypes (PMID: 23747035) I could not find a similar study attempting to associate SNPs within CNV regions with phenotype variability.

(3) Strengths and weaknesses: The strengths of this study the relative high number of samples for both conditions and that all individuals in the study are genotyped on the same platform and that the phenotype information is compiled similarly. Obvious weaknesses are that the authors do not provide detailed phenotypic descriptions of the individuals enrolled. While they state that "Individual-level data used during this study are not publicly-available due to restrictions", not even summary statistics for the analyzed vascular phenotype are given. Same hold true for genetic data, neither the exact CNV locations nor the genotypes for the two SNPs discussed are reported in an accessible tabular way. While the manuscripts title begins with "A Method[…]" and also the article type is "Technical advance" the authors choose to not make any computational
procedures directly accessible and also not describe these in detail. The authors also make no attempt to further explain how the associated SNPs could influence the phenotype (e.g.: gene dosage effect, tagging a functionally relevant variant, etc.) and provide a clinical utility for this approach (e.g.: Can these SNPs be used to make predictions about phenotype severity pre-/postnatally and would the authors recommend this?). In its current state the lack of openness regarding both genotype/phenotype data and analysis pipeline greatly reduces reproducibility and this manuscripts value to the scientific community.

(4) Presentation: Generally the manuscript is well written and the authors approach is viable to address the studies objective (with the limitations detailed in point 3). Several sentences are unnecessarily long and could be split (e.g. 3:30-43) and there is quite some redundancy (introduction, between results and discussion). Other minor problems are the use of use of unusual terms like "reciprocal genetic disorder", "stereotyped hemideletions" and "syndromic symptoms"). The figures accompanying the manuscript are at printable resolution (300dpi) but relatively uncommon dimensions (width 127mm). Figures 1A-D and 2A are produced with an unspecified plotting library and then manually labeled using different fonts, font sizes and colors. Figures 1A,C,D are extremely over-plotted, the colors used are neither print friendly nor colorblind-save and it is impossible to differentiate the "Varying shades of blue and orange" used to "represent individual participants". Figure 1B does not hold meaningful information and could easily be presented in a more intuitive way (dot-plot with quantiles). Figure 2A seems to have been vertically resized as it is somewhat blurry; the gene names are very small and genomic coordinates are not plotted, also one could add significance thresholds as a horizontal line. Figure 2B has been produced using a different software and thus has a different style; the presentation of the data in confusing (bar plot of percentages not scaled to 100%, 0% data still has a bar height, genotype for the duplication is uninformative) and it does not hold any information not present in the main text. Generally, some figure panels are not referenced in the main text while they are sometimes referenced without context (e.g. 5:30)

(5) Interpretation of results: The authors' interpretation of the results and the conclusions drawn are principally valid. However there is not too much to interpret, as the main results are the higher variability of duplication calls and the association of two SNPs with the cardiac phenotype. Some points are overemphasized or over-interpreted (they repeat 12 times that the developed a "method[...]""). The final conclusion that this approach can be used to "uncover novel genetic mechanisms" is not supported in any way by the data presented as no mechanistic insight was reported.
Suggestions for improvement:

Mayor:

- Decide about the purpose of this manuscript:

  o If it is a methodological manuscript provide detailed descriptions of the method which enable other researchers to replicate and conduct similar studies (e.g. flow-chart, open-source useable code deposited to a repository). Also add a flow-diagram for the proposed workflow as a figure to help the readers understand the steps needed to conduct a similar study. In this context try replacing the term "method" with something like workflow or pipeline as no new method has been developed but instead available methods have been used together for a new purpose.

  o If it is clinical provide details about the individuals involved in the study and their phenotype.

  o If it is functional, try to elucidate how the associated SNPs are influencing the phenotype (e.g. splicing or dosage effect, tagging of another disease relevant SNP, etc.).

- The statistical model and software used for association analysis is not described sufficiently. Please add this to the methods section. Please also discuss the suitability of the method used for non-diploid association studies.

- The cardiac phenotype in the cohort is missing. Not even summary statistics are given (e.g. XX/25 individuals with deletion had severe SVAS). If possible, provide a table with individual data as this phenotype is essential for the manuscript.

- The evaluation of SVAS phenotype is described vaguely. Please describe in more detail the categories for aortic phenotype and what the exact criteria used for inclusion in in each group were.

- Also the description of haploid and triploid calls and how these were achieved is incomplete. It seems like the authors used fixed cutoffs, determined by visual inspection of the aggregated B allele fraction plots, for calling non-diploid genotypes. Please describe the approach used and give the cutoffs for this project. In this context please also describe the call rate for the SNPs in the CNV region with absolute numbers (e.g. "haploid call were made for 99,0% (xxx/yyy) of SNPs") and discuss reasons for the miscalled SNPs.

- The authors mention the variability observed in microdeletion syndromes like WS and discuss and analyze variants inside the affected genomic region as a cause. In fact, there could be other explanations (common variants outside of the CNV region with low effect
size, additional rare variants with high effect size like other CNVs and presence of other confounding genetic disorders). Please discuss these possibilities properly. As there is dense genome wide data SNP available for all cases one could check for variants already associated with aortic diameter (e.g. GWAS Central).

Minor concerns:
- Use only HGNC approved gene symbols and do not emphasize common names or protein names in italic.
- Give OMIM numbers for ELN associated disorders (185500, 123700).
- Give the version number of all software used.
- For commercial components and software used (HumanOmni5-4v1.1 SNP, GenomeStudio) give complete manufacturer information.
- The first sentence of the results section is actually methods (6:51 "CNVs were identified by PennCNV in the 7q11.23 locus for all individuals (Figure 1).")
- 9:20-23 Neither does the cited publication "detail[s] the effects of ELN variation on cardiovascular phenotypes in WS" nor was it really "recently published". Update or remove this statement.
- Citation 1 and 16 are referencing articles GeneReviews intended for genetic counselors. Also at least references 17 and 18 are general reviews. Please cite primary literature if possible.
- 2:13-16 "Though software exists to identify areas of copy number variation (CNV) using commonly-available SNP-chip data, this has not been translated into non-diploid genotypes of the CNV regions." This sentence is not readily understandable. What exactly has not been translated? I believe the authors want to say that the commonly used software used for CNV calling does not emit genotype calls for SNVs in this region. Please clarify.

Suggestions (encouraged but not needed):
- Have the SNPS rs2528795 and rs37609 been previously associated with a phenotype?
- A supplementary table with detailed phenotypic information for each individual included.
- A supplementary table with CNV sizes for each individual in HGNC approved nomenclature.

- A supplementary table with genetic analyses performed in each individual prior to diagnosis and additional genetic findings (other CNVs, variants from next generation sequencing etc.)

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Unable to assess

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Unable to assess

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests' below. If your reply is yes to any, please give details below

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal