A genome-wide association study of *de novo* deletions identifies a locus on chromsome 7p14.1 associated with non-syndromic isolated cleft lip/palate

Samuel G. Younkin [1], Robert B. Scharpf [2], Ingo Ruczinski [1], Mary L. Marazita [3], Alan F. Scott [4], Terri H. Beaty [5]

[1]Johns Hopkins University, School of Public Health, Dept. of Biostatistics, Baltimore, MD, USA
[2]Johns Hopkins University, School of Medicine, Dept. of Oncology, Baltimore, MD, USA
[3]University of Pittsburgh, School of Dental Medicine, Dept. of Oral Biology, Pittsburgh, PA, USA
[4]Johns Hopkins University, School of Medicine, Inst. of Genetic Medicine, Baltimore, MD, USA
[5]Johns Hopkins University, School of Public Health, Dept. of Epidemiology, Baltimore, MD, USA

Abstract

Copy number variants (CNVs) may play an important part in the development of common birth defects such as oral clefts, and individual patients with multiple birth defects (including clefts) have been shown to carry chromosomal deletions. We are interested in identifying regions of the genome in which a deletion is present in a child with cleft lip/palate, and absent in both parents. Such de novo regions were compared in children with cleft lip/palate and unaffected children. We used probe intensity data from the Illumina 610K quad array to identify CNVs in two independent sets of child-parent trios. The control group was drawn from a family based study of dental caries, and the cleft group was composed of trios ascertained through a child with an isolated oral cleft (either cleft lip, cleft palate or clefts lip and palate). All subjects are of European ancestry, and the control families are from rural Appalachia. We performed CNV discovery among these trios using two approaches: a joint hidden Markov model implemented in PennCNV and an algorithm specific for *de novo* CNV detection in case-parent trios referred to as *MinimumDistance*. We then conducted a one-sided Fisher's exact test for increased frequency of de novo deletions among offspring with an oral cleft. After adjusting for correlation due to overlapping CNVs and multiple testing, we identified a significant region on 7p14.1 (38.7 kB) and a suggestive region on 14q11.2 spanning 26.8 kB that was marginally significant.

Methods

Copy number variants (CNVs) may play an important role in the development of common birth defects such as oral clefts, and individual patients with multiple birth defects (including clefts) have been shown to carry chromosomal deletions of varying size. The aim of our study is to identify deletions among subjects with an isolated, nonsyndromic oral cleft that have not been inherited from either of parent, each of which does not have cleft lip/palate. We hypothesize development of a *de novo* deletion in a critical region of the genome increases risk for oral clefts, and perform an association study to identify such regions.

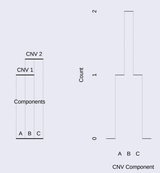

Figure 1: CNV components are constructed by decomposing CNVs into segments that either are unique or overlap entirely with another component. Above we see an example created by two partially overlapping CNVs. In this case three components (A, B & C) are created from two partially overlapping CNVs, and have counts of 1, 2 & 1, respectively.

We used two independent methods for identifying *de novo* deletions — the joint hidden Markov model (HMM) implemented in PennCNV, and a novel approach named *MinimumDistance* that segments the parent to offspring difference in marker probe intensity [1, 2]. We collected case-parent trios as part of the Oral Cleft Project in the GENEVA Consortium, and were generously provided with data from small pedigrees collected from rural Appalachia as part of a study of dental caries performed by Marazita *et al.* to serve as controls.

Methods (cont.)

Both methods rely on measures of the marker probe intensity and B allele frequency generated by hybridization to the *Illumina* 610 quad array as performed by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. We then restricted our analysis to *de novo* deletions spanning at least ten markers.
To perform the association analysis, we first decomposed the set of all *de novo* deletions into a partition containing no partially overlapping CNVs, referred to as CNV components (see Figure 1). With these we simply count the frequency of CNV components in the cleft group and control group and perfrom Fisher's one-sided test for CNV components with frequency of at least five. To correct for the correlation among these adjacent CNV components, as well as to consider multiple tests over all CNV components, we compute the rejection region for genome-wide significance at the $\alpha = 0.05$ level through permutation.

Results

Figure 2 displays the $-\log_{10} p$ values for each of the 470 CNV components, along with a dashed horizontal line indicating the value needed for genome-wide significance, as well as a dashed line indicating the level for genome-wide significance using the conservative Bonferonni correction. We see two peaks, one on chromosome 7p14.1 which is highly significant, and one on 14q11.2 which achieves a marginal, but suggestive, level of significance in association. Figure 3 displays the frequency of *de novo* events among the oral cleft and control offspring for an 80 kb region on chromosome 7, and we see that in places 22 subjects with an oral cleft carry a *de novo* deletion, while no more than two individuals in the controls carry such a deletion.

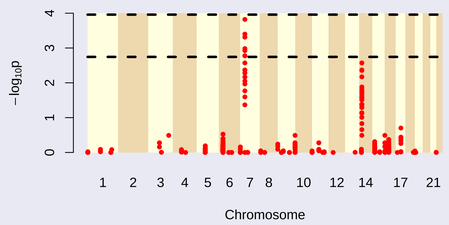

Figure 2: The $-\log_{10} p$ values are plotted by genomic position for each of the CNV components created from *de novo* deletions with coverage greater than 10. The two dashed lines indicate the necessary level for genome-wide signifcance using the overly-conservative Bonferroni correction, and a permutation based correction.

Of these 22 cleft lip/palate subjects with a *de novo* deletion in this region nine had cleft lip, seven had cleft palate and six had cleft lip and palate. The gene nearest to this region on 7p14.1 is the *TARP* gene. The *TARP* gene is not known to be related to TARP syndrome. Other genes in the vicinity include *AMPH*, *FAM183B*, *STARD3NL* and *TXNDC3*. The CNV component giving the most significant $-\log_{10} p = 3.82$ corresponds to a natural log relative frequency of 2.67 (21/1,384:1/953).

Results (cont.)

The Database for Genomic Variation identifies 16 known copy number variants, 14 of which may be deletions, in this region of 7p14.1, yet it is not clear if these known CNVs are frequent enough in the population to represent copy number polymorphisms (CNPs), and here we conclude only that small deletions near *TARP* are significantly more common among children born with an oral cleft compared to unaffected children of European origin.

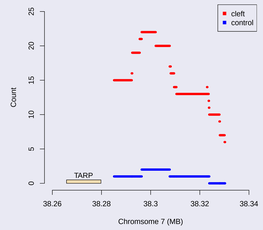

Figure 3: The counts of CNV components created from *de novo* deletions in the significant region on 7p14.1 are displayed by genomic position in red for subjects with cleft lip/palate, and blue for control subjects. The closest gene, *TARP*, is only a few kB away.

To visualize the degree to which these *de novo* deletions found in the cleft group display a decrease in probe intensity, as measured by the well-known log(R ratio), we display raw values in Figure 4. The clear depression in the log(R ratio) levels among the 22 cleft cases indicates a decreased amount of genetic material present in this region near *TARP*, as expected in a deletion. The log(R ratio) among the 22 cleft cases identified as carrying a *de novo* deletion is displayed in Figure 4 in red, along with their parents in blue.

Discussion

This analysis is the first comprehensive analysis of CNVs based on probe intensities generated with high-throughput genome-wide marker panels in case-parent trios. Here we compared *de novo* deletions in children born with an isolated, non-syndromic oral cleft (cleft lip, cleft palate or cleft lip & palate) to a sample of unaffected trios. CNV discovery was carried out using PennCNV and *MinimumDistance* in both groups, and only apparent *de novo* deletions spanning > 10 adjacent SNPs were considered to minimize erroneous calls of CNVs. Trios where the child had fewer copies than either parent were the focus of this analysis, and a one-sided test was used to compare cleft case children to control children.
The distribution of estimated CNVs is compared in Figure 5, and while the vast majority of subjects had zero *de novo* deleted CNVs, there were some differences between cleft cases and controls. Oddly enough, more control children carried several *de novo* deletion CNVs over the entire genome. Examining the distribution of *de novo* CNVs across the genome, however, revealed one chromosomal region on 7p14.1 (near the *TARP* gene) where cleft cases showed significantly more deletion CNVs than did control children. This difference in the counts of *de novo* deletions achieved genome-wide significance when adjusted for the correlations in counts of CNVs and the multiple testing done here.

Discussion (cont.)

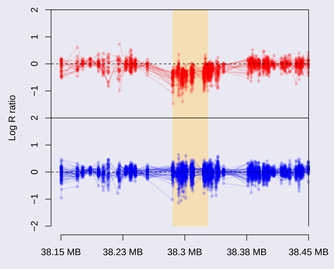

Figure 4: log(R ratio) values in the significant region on 7p14.1 are displayed in red for the 22 cleft cases with a *de novo* deletion, along with their parents in blue. The significant region is shaded in yellow. A clear depression is present among the offspring, but not their parents, as is expected from a *de novo* deletion.

TARP codes for a TCR gamma alternate reading frame protein and is embedded within an intron of the T-cell receptor-gamma locus. This gene has never been suggested as being related to oral clefts, but this analysis showed an odds ratio of 14.7 of being a cleft case compared to a control if the child carried a *de novo* deletion in this region. Further studies will be required to fully understand the role of this gene in the etiology of oral clefts.

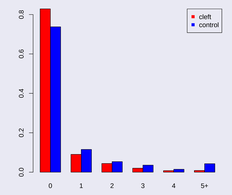

Figure 5: Bars represent the proportion of subjects with a given number of *de novo* deletions. Subjects with cleft lip/palate are represented in red and control subjects in blue. We see that approximately 80% of subjects contain no *de novo* deletion and there are significantly more control subjects with five or more *de novo* deletions.

References

[1] K. Wang, Z. Chen, M. G. Tadesse, J. Glessner, S. F. A. Grant, H. Hakonarson, M. Bucan, and M. Li, "Modeling genetic inheritance of copy number variations.," *Nucleic Acids Res*, vol. 36, p. e138, Dec 2008.

[2] R. B. Scharpf, S. Younkin, T. H. Beaty, H. Schwender, A. F. Scott, and I. Ruczinski, "Fast detection of de novo copy number variants from SNP arrays of case-parent trios," *(in revision)*, 2012.

Send all correspondence to Samuel G. Younkin at syounkin@jhsph.edu