Copy number variation analysis of twin pairs discordant for cleft lip with or without cleft palate

Luca Scapoli¹, Francesco Carinci², Annalisa Palmieri¹, Francesca Cura³, Alessandro Baj⁴,⁵, Giada Beltramini⁵, Raffaella Docimo⁶ and Marcella Martinelli¹

Abstract
Non-syndromic cleft lip with or without cleft palate (nsCL/P) is a frequent orofacial malformation. The comparison of concordance rate observed in monozygotic and dizygotic twins supports high level of heritability and a strong genetic component. However, phenotype concordance for orofacial cleft in monozygotic twins is about 50%. The aim of the present investigation was to detect postzygotic events that may account for discordance in monozygotic twins. High-density SNP microarrays hybridization was used to genotype two pairs of monozygotic twins discordant for nsCL/P. Discordant SNP genotypes and copy number variants were analyzed to identify genetic differences responsible of phenotype discrepancy. A number of differences were observed, none involving known nsCL/P candidate genes or genomic regions. Considering the limitation of the study, related to the small sample size and to the large-scale investigation method, the results suggest that the detection of discordant events in other monozygotic twin pairs would be remarkable and warrant further investigations.

Keywords
Cleft lip, cleft palate, monozygotic twins

Date received: 26 September 2018; accepted: 17 May 2019

Introduction
Cleft of the lip with or without cleft palate (CL/P) is the most common orofacial malformation, with a prevalence close to 1/1000 at birth.¹ However, the prevalence varies depending on ethnic origin.²,³ The non-syndromic cleft lip with or without cleft palate (nsCL/P) is a heterogeneous disorder with multiple phenotypic presentations and is considered a typical example of trait with complex inheritance, where a combination of multiple genetic and environmental factors contributes to phenotype expression. Twin studies are commonly used to investigate etiology of common diseases with complex inheritance. Monozygotic (MZ) or identical twins result from a single ovum, fertilized by one sperm, while dizygotic (DZ) twins result from two different ova, fertilized by two different sperm. Otherwise from DZ twins, which originate from two zygotes and
share on average half of the genome, MZ twins are long thought to share 100% of their genomic information, because they originate from the same zygote. However, additional genetic components, such as epigenetic factors and postzygotic somatic mutation events, may explain different traits of expression in MZ twins.\(^5\) Increasing evidences of genetic differences have been reported both in typically developing and in clinically discordant MZ pairs.\(^6\)

Twin studies demonstrated a consistent genetic component in nsCL/P etiology, indeed a higher concordance rate in MZ (25%–50%) was often observed compared to DZ (3%–6%) twins.\(^7\) Molecular analysis of discordant MZ twins has been attempted to identify nsCL/P genetic factors. A de novo nonsense mutation in \textit{IRF6} was detected in the affected twin of a twin pair discordant for the presentation of Van der Woude clefting syndrome.\(^8\) However, other investigations, using different technical approaches, were unsuccessful to identify genetic differences in discordant nsCL/P twin pairs.\(^9\)–\(^12\)

Discordant MZ twin pairs, that are informative in respect to variability of phenotypic expression, epigenetics, and postzygotic mutagenesis, may represent an alternative approach to identify genes in inherited disorders. We hypothesized that postzygotic de novo mutations could cause discordant MZ twin pairs for nsCL/P, that are otherwise genetically identical. To test this hypothesis we have investigated two MZ twin pairs by means of high-density SNP genotyping arrays that consent the analysis of postzygotic de novo copy number variation (CNV) events.

Materials and methods

Discordant twin pair collection was part of a broader investigation aimed to identify inherited susceptibility factors of nsCL/P.\(^13\) A team of clinicians performed the diagnosis and excluded additional birth malformations or metabolic diseases. A detailed interview excluded families that may be subjected to known or suspected clefting agents, such as phenytoin, warfarin, ethanol, and smoking. The study was approved by the local ethics committees and it complied with the Helsinki Declaration’s Ethical Principles for Medical Research Involving Human Subjects. Written informed consent was obtained from all patients and parents.

Five twin pairs discordant for nsCL/P were identified. Genomic DNA was extracted and purified from whole blood using standard techniques. Twin pairs were analyzed for zygosity by direct genotype comparison of a panel of highly polymorphic microsatellite DNA loci. Three twin pairs were excluded from the investigation because the originated by different zygotes. Two molecularly ascertained MZ twin pairs that were discordant for nsCL/P were analyzed by high-density SNP microarray. Genotyping was performed using the Illumina HumanOmni1-Quad array, which contains nearly 1.14 million markers including SNP and CNV probes.

BeadChip data were processed using GenomeStudioV2011.1 (Illumina Inc.) and PennCNV.\(^14\) Primary data analyses, including raw data normalization, clustering, and genotype calling were performed using algorithms in the genotyping module. The software derives, for each sample, log R ratios (LRRs) and B allele frequencies for each probe on the Quad array; the LRR reflects relative probe fluorescence intensity, which varies with the discrete number of copies of probe-specific DNA present within an individual’s genome. A copy number state of 2 per individual is considered normal (one copy per chromosome); lower value reflects copy number loss and higher values, a copy number gain. Each sample CNV pool was subjected to filtering steps in order to remove alteration smaller than 10 kb in size and containing lower than 5 probes. CNV that passed these filtering steps were retained for downstream analysis. Chromosome regions annotations were obtained from UCSC Refseq track Human genome build 19. All analyses were conducted with R version 3.4.3, Platform: x86 64-pc-linux-gnu (64-bit) running under Ubuntu 16.04.3 LTS.

Results

Genotyping of SNPs of the four DNA samples by microarray hybridization produced high quality results; indeed, for each sample, the genotype call rate was >99.7%. As expected, the comparison of genotypes between the affected and the unaffected twin revealed a high level of concordance in each twin pair (Table 1). Indeed, only 25 (0.002% of total genotypes) discordant calls were observed in each pair. The high level of concordance confirmed that twin pairs were actually MZ, while discordant SNP genotypes could be explained as either genotyping errors or de novo mutations. Discordant polymorphisms did not alter gene coding sequences, and they were not classified as pathogenic in the ClinVar database.
The intensities of allele probe hybridization in the SNP array platforms were analyzed to evaluate the ploidy of each tested locus. Indeed, CNVs such as duplication and deletion increase or decrease the total measured intensities; moreover, for large CNVs that span multiple SNPs, intensity ratios have patterns distinct from normal disomic genomic regions. In this investigation, we considered CNV regions spanning more than 10 kbp. In the four samples, the number of detected CNVs varied between 51 and 70 with a median length of 23 kbp. In order to identify inherited CNVs that could act as nsCL/P susceptibility loci, we first looked for CNVs detected in all the investigated samples (Table 2). Two CNVs of the list consisted of deletions that did not include any transcripted sequence. The remaining CNVs spanned 12 genes, including JAG2 a possible genetic factor of nsCL/P.

Then we focused on genetic differences in each twin pair, particularly to CNVs that may account for phenotype discordance. The CNVs detected exclusively in the affected individual of each pair are shown in Table 3. Such CNVs include 34 out of 66 variations detected in patient ID = 100101, and 13 out of 50 variations detected in patient ID = NBF3.

No overlap between the two CNV lists, specific for each twin pair, was found.

**Discussion**

Several factors could contribute to discordance of diseases between MZ twins, including postzygotic somatic mutations, X chromosome inactivation, differential methylation, stochastic factors, and non-genetic intrauterine environmental factors such as unequal cell allocation at twinning and disproportionate placental blood supply.\(^5\)\(^15\) Discordant MZ twins can be a valuable resource for complex diseases, indeed genetic comparison of discordant twins could potentially help to increase reliability of candidate genes in complex diseases or to find novel disease susceptibility genes that could partly explain missing heritability.

The current study reports genome-wide SNP and CNV results on two MZ twin pairs discordant for nsCL/P. A small number of in-pair discordant SNP genotypes were found; none of them appeared as a probable causative mutation. The genotype discrepancy may be related to genotyping inaccuracy of large-scale microarray typing, although at a level similar to those previously reported.\(^11\) We searched for postzygotic CNVs that may account for the discordant phenotype. In addition, we analyzed the shared CNVs among twin pairs looking for variants of face development genes. Lists of selected CNVs were reported along with annotations including involved genes and previous contribution to clinical relevant data. The reported genetic regions and genes did not overlap with any of the candidate regions by previous genome wide allelic association analyses. These data partially agree with a previous report by Shi et al. who investigated 333 nsCL/P candidate genes for CNVs; they found that CNVs could have a role in nsCL/P etiology but with relatively rare occurrence. Indeed, analyzing 725 nsCL/P Scandinavian families, they identified only seven deletions.\(^16\)

Previous investigations attempted the identification of nsCL/P genetic factors by comparison of discordant MZ twins. Mansilla et al.,\(^9\) by comparing sequences of 18 candidate genes, did not find etiologic somatic

---

**Table 1.** Comparison of SNP genotypes between the discordant twins.

| Twin pair | Sample ID | # concordant SNPs | # discordant SNPs |
|-----------|-----------|--------------------|-------------------|
| 1         | NBF3-NBF4 | 1,011,267          | 25                |
| 2         | 100101-100104 | 1,011,764      | 25                |

**Table 2.** List of CNVs detected in all analyzed samples.

| Chr. | Start | End | Width | # of SNPs | CNV_TYPE | Genes |
|------|-------|-----|-------|-----------|----------|-------|
| 2    | 41,092,961 | 41,103,770 | 10,810 | 13 | Loss | – |
| 2    | 88,932,848 | 89,090,893 | 15,8046 | 59 | Gain | RPIA, ANKRD36BP2 |
| 6    | 103,850,891 | 103,868,723 | 17,833 | 9 | Deletion | – |
| 8    | 32,799,628 | 32,810,651 | 11,024 | 14 | Deletion | – |
| 11   | 55,122,337 | 55,175,539 | 53,203 | 35 | Loss | OR4A15 |
| 14   | 105,275,606 | 105,697,201 | 421,596 | 244 | Gain | JAG2, CEP170B, PLD4, AHNAK2, CDC4, GPR132, NUPT14, BRF1, BTBD6 |

CNV: copy number variation.
Table 3. List of CNVs that were detected only in the CL/P affected twin.

| Patient ID | Chr. | Start     | End       | Width    | # of SNPs | CNV type | Genes involved                                      |
|------------|------|-----------|-----------|----------|-----------|----------|-----------------------------------------------------|
| 100101     | 2    | 14,109,052| 14,119,079| 10,028   | 10        | Loss     | −                                                   |
| 100101     | 2    | 52,607,219| 52,621,681| 14,463   | 5         | Loss     | −                                                   |
| 100101     | 2    | 89,904,056| 89,920,851| 16,796   | 10        | Gain     | −                                                   |
| 100101     | 2    | 97,150,351| 97,165,854| 15,504   | 6         | Loss     | NEURL3                                             |
| 100101     | 2    | 153,489,894| 153,508,850| 18,957   | 22        | Loss     | FMNL2, PRPF40A                                      |
| 100101     | 2    | 238,262,529| 238,275,105| 12,577   | 15        | Gain     | COL6A3                                             |
| 100101     | 3    | 149,649,355| 149,660,146| 10,792   | 5         | Loss     | RNF13                                              |
| 100101     | 4    | 14,529,946 | 14,543,205| 13,260   | 12        | Loss     | −                                                   |
| 100101     | 4    | 100,728,344| 100,744,538| 16,195   | 11        | Loss     | DAPPI                                              |
| 100101     | 4    | 144,879,245| 144,889,446| 10,202   | 9         | Loss     | −                                                   |
| 100101     | 5    | 18,365,795 | 18,382,021| 16,227   | 14        | Loss     | −                                                   |
| 100101     | 5    | 84,822,505 | 84,868,110| 45,606   | 8         | Loss     | −                                                   |
| 100101     | 6    | 29,962,774 | 29,981,888| 19,115   | 78        | Gain     | HLAH, HLAG, HLAJ                                     |
| 100101     | 6    | 67,893,398 | 67,923,322| 29,925   | 13        | Loss     | −                                                   |
| 100101     | 6    | 67,954,304 | 68,004,709| 50,406   | 19        | Loss     | −                                                   |
| 100101     | 7    | 142,157,556| 142,172,768| 15,213   | 13        | Loss     | TCRBV22SIA2N1T, TCRBV5SIA1T                        |
| 100101     | 8    | 130,571,112| 130,581,329| 10,218   | 10        | Loss     | −                                                   |
| 100101     | 9    | 10,384,286 | 10,395,076| 10,791   | 11        | Deletion  | PTPRD                                              |
| 100101     | 11   | 48,284,271 | 48,304,374| 20,104   | 36        | Loss     | OR4X1                                              |
| 100101     | 11   | 51,052,130 | 51,152,453| 100,324  | 8         | Gain     | −                                                   |
| 100101     | 11   | 114,007,895| 114,017,913| 10,019   | 10        | Loss     | ZBTB16                                             |
| 100101     | 12   | 74,069,809 | 74,089,055| 19,247   | 10        | Loss     | −                                                   |
| 100101     | 13   | 17,982,800 | 18,006,081| 23,282   | 7         | Gain     | −                                                   |
| 100101     | 14   | 1,524,714  | 1,537,988 | 13,275   | 8         | Gain     | SIRPD                                              |
| 100101     | 16   | 41,004,182 | 41,016,180| 11,999   | 16        | Gain     | AOC3                                               |
| 100101     | 17   | 62,342,876 | 62,353,618| 10,743   | 5         | Loss     | −                                                   |
| 100101     | 18   | 64,098,920 | 64,110,327| 11,408   | 16        | Loss     | −                                                   |
| 100101     | 20   | 2,267,951  | 2,275,367 | 7,416    | 13        | Gain     | abParts                                            |
| NBF3       | 2    | 34,809,903 | 34,820,073| 10,171   | 15        | Loss     | −                                                   |
| NBF3       | 2    | 91,293,640 | 91,322,549| 28,910   | 12        | Loss     | −                                                   |
| NBF3       | 3    | 198,837,449| 198,871,090| 33,642   | 13        | Loss     | −                                                   |
| NBF3       | 6    | 26,849,823 | 26,860,992| 11,170   | 15        | Loss     | −                                                   |
| NBF3       | 6    | 32,617,395 | 32,633,666| 16,272   | 24        | Gain     | HLA-DQ81                                           |
| NBF3       | 7    | 57,728,536 | 57,767,235| 38,700   | 13        | Gain     | GUSBFP                                             |
| NBF3       | 7    | 64,895,813 | 64,925,393| 29,581   | 15        | Gain     | −                                                   |
| NBF3       | 10   | 46,781,951 | 46,805,985| 24,035   | 7         | Gain     | PTPN20, GLUD1P7                                    |
| NBF3       | 14   | 105,648,434| 105,725,651| 77,218   | 9         | Gain     | BRF1, BTBD6                                        |
| NBF3       | 16   | 34,343,935 | 34,601,761| 257,827  | 27        | Gain     | LINC01566, UBEZM1                                  |
| NBF3       | 16   | 68,615,369 | 68,650,243| 34,875   | 6         | Gain     | −                                                   |
| NBF3       | 18   | 14,211,931 | 14,239,072| 27,142   | 6         | Gain     | ANKRD20A5P                                         |
| NBF3       | 20   | 1,526,976  | 1,541,888 | 14,913   | 9         | Gain     | SIRPD                                              |

CNV: copy number variation.

mutations in 13 MZ pairs. Similarly, Kimani et al.\textsuperscript{10} investigated 25 discordant MZ twin pairs with different genome scale genetic methods; they not only concluded that postzygotic genomic alterations are not a common cause of MZ twin discordance for isolated nsCL/P but also suggested that detection of
discordant events in other MZ twin pairs would be remarkable and of potential disease significance.

A possible limitation of our study was related with the CNV calling method from microarray data. Indeed, discrimination of biologically relevant data from noise CNV is still a bioinformatics challenge and different algorithms produce different results.17 We tried to increase accuracy for CNV calling by setting stringent threshold of CNV size and spanning SNP number. However, this could reduce sensitivity increasing missing calls, while the false positive call remains a concrete possibility, as observed in other investigations.18 There is no clear estimate of the rate of somatic CNVs, and our sample that is limited to discordant twins, in theory should have a higher rate of such events. Considering all these limitations, together with the small size of our sample study, the results of this investigation should be considered with caution and more data obtained with different technical approaches are needed to evaluate the real impact of CNVs in nsCL/P. Further investigations of specifically involved tissue, aimed to screen for epigenetic factors or postzygotic somatic mutation events, could possibly help to explain different trait expression in MZ twins.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Luca Scapoli https://orcid.org/0000-0003-4006-9910
Francesco Carinci https://orcid.org/0000-0001-9639-6676

References
1. Watkins SE, Meyer RE, Strauss RP, et al. (2014) Classification, epidemiology, and genetics of orofacial clefts. *Clinics in Plastic Surgery* 41(2): 149–163.
2. Group IW (2011) Prevalence at birth of cleft lip with or without cleft palate: Data from the International Perinatal Database of Typical Oral Clefts (IPDTOC). *The Cleft Palate-craniofacial Journal* 48(1): 66–81.
3. Mossey PA and Modell B (2012) Epidemiology of oral clefts 2012: An international perspective. *Frontiers of Oral Biology* 16: 1–18.
4. Acuna-Hidalgo R, Veltman JA and Hoischen A (2016) New insights into the generation and role of de novo mutations in health and disease. *Genome Biology* 17(1): 241.
5. Machin GA (1996) Some causes of genotypic and phenotypic discordance in monozygotic twin pairs. *American Journal of Medical Genetics* 61(3): 216–228.
6. Bruder CE, Piotrowski A, Gijsbers AA, et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics* 82(3): 763–771.
7. Mitchell LE and Risch N (1992) Mode of inheritance of nonsyndromic cleft lip with or without cleft palate: A reanalysis. *American Journal of Human Genetics* 51(2): 323–332.
8. Kondo S, Schutte BC, Richardson RJ, et al. (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nature Genetics* 32(2): 285–289.
9. Mansilla MA, Kimani J, Mitchell LE, et al. (2005) Discordant MZ twins with cleft lip and palate: A model for identifying genes in complex traits. *Twin Research and Human Genetics* 8(1): 39–46.
10. Kimani JW, Yoshiura K, Shi M, et al. (2009) Search for genomic alterations in monozygotic twins discordant for cleft lip and/or palate. *Twin Research and Human Genetics* 12: 462–468.
11. Jakobsen LP, Bugge M, Ullmann R, et al. (2011) 500K SNP array analyses in blood and saliva showed no differences in a pair of monozygotic twins discordant for cleft lip. *American Journal of Medical Genetics. Part A* 155(3): 652–655.
12. Takahashi M, Hosomichi K, Yamaguchi T, et al. (2018) Whole-genome sequencing in a pair of monozygotic twins with discordant cleft lip and palate subtypes. *Oral Diseases* 24: 1303–1309.
13. Cura F, Palmieri A, Girardi A, et al. (2018) Possible effect of SNAIL family transcriptional repressor 1 polymorphisms in non-syndromic cleft lip with or without cleft palate. *Clin Oral Investig* 22: 2535–2541.
14. Wang K, Li M, Hadley D, et al. (2007) PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. * Genome Res* November;17(11): 1665–1674.
15. Gringras P and Chen W (2001) Mechanisms for differences in monozygous twins. *Early Human Development* 64(2): 105–117.
16. Shi M, Mostowska A, Jugessur A, et al. (2009) Identification of microdeletions in candidate genes for cleft lip and/or palate. *Birth Defects Res A Clin Mol Teratol* 85: 42–51.
17. Castellani CA, Melka MG, Wishart AE, et al. (2014) Biological relevance of CNV calling methods using familial relatedness including monozygotic twins. *BMC Bioinformatics* 15: 114.
18. Talseth-Palmer BA, Holliday EG, Evans TJ, et al. (2013) Continuing difficulties in interpreting CNV data: Lessons from a genome-wide CNV association study of Australian HNPCC/lynch syndrome patients. *BMC Medical Genomics* 6: 10.