



# The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes

A Collins<sup>1\*</sup>, L Arias<sup>2</sup>, R Pengelly<sup>1</sup>, J Martínez<sup>2</sup>, I Briceño<sup>2</sup>, S Ennis<sup>1</sup>

## Abstract

### Introduction

Next-generation sequencing is revolutionising the study of genetic variation and its role in disease. Individual DNA samples can now be sequenced cost-effectively enabling analysis of the complete spectrum of genetic variation. This technology has the potential to contribute significantly to the understanding of non-syndromic cleft lip and/or palate. This condition occurs with relatively high frequency and only a proportion of the underlying genetic causal factors have been identified. Many of the genes implicated have been found through genome-wide association studies but further progress is limited because these approaches consider only common genetic variants and neglect rarer variations. Because many of the causal genetic variants remain unknown, the role of gene-environment and gene-gene interaction is difficult to characterise. The identification of novel, low frequency, variants will provide new insights into the biological mechanisms and pathways involved in the condition. Sequence-based analysis will also be invaluable for fine mapping causal variants in the larger regions already identified by linkage and association studies for which positive identification of causal genetic variants has proven difficult. This review considers the available evidence for the genes

involved and current understanding of how genetic variation interacts with environmental factors known to influence risk. Only by characterising the underlying genetic factors will the effort to understand gene-environment interaction and underlying functional processes be successful.

### Conclusion

Success with next-generation sequencing will lead to improvements in prediction, prevention, and treatment for cleft lip and palate patients.

### Introduction

Next-generation sequencing (NGS) is revolutionising genomics<sup>1</sup> and this trend of increasingly significant impact is likely to continue. Most importantly, NGS provides a route to characterising and understanding the role of genetic variation in disease. This is important because evidence from genome-wide association studies (GWAS) suggests that much of the heritability underlying complex disease phenotypes is not explained by common deleterious genetic variants with small effect sizes<sup>2</sup>. NGS enables new analytical strategies, which are not achievable through genome-wide association studies. These include: the identification of the complete complement of DNA variants in samples (across the allele frequency spectrum); tests on the burden of rare variation (do specific genes contain many rare variants which collectively impair gene function?); the identification of de novo mutations and the genes underlying rare Mendelian forms of disease; fine mapping of causal variants within broader regions identified by linkage and/or association and the

characterisation of important structural variation, such as differences in copy number, which may contribute to disease.

Orofacial cleft lip and/or palate (CLP) represents a complex phenotype for which NGS offers the potential to increase understanding. CLP phenotypes are among the most frequent birth defects with rates of between 1/500 and 1/2500 births<sup>3</sup>. The frequency of CLP phenotypes is related to population ancestry, geographical location, maternal age, prenatal exposures and socioeconomic status<sup>4-6</sup>. The frequency of orofacial clefting (OC) is higher in Latin American and Asian countries<sup>7</sup>. CLP phenotypes are classified into syndromic and non-syndromic forms. The former includes many conditions which have simple Mendelian modes of inheritance in families for which a number of causal genes have already been identified through, for example, linkage mapping. However, ~70% of CLP cases occur as isolated phenotypes without any additional cognitive or craniofacial structural abnormalities. These are usually described as isolated non-syndromic cleft lip and/or palate (NSCLP). Understanding the factors underlying NSCLP phenotypes is important to improve prevention, treatment and prognosis of the condition. However, the genetic dissection of NSCLP phenotypes is challenging and progress towards understanding the underlying genetic and environmental factors, and how they are inter-related, has, until recently, been relatively slow, despite decades of research. This review considers the impact of recent work and, in particular, prospects for progress through

\*Corresponding author  
Email: arc@soton.ac.uk

<sup>1</sup>Genetic Epidemiology and Genomic Informatics Group, Faculty of Medicine, University of Southampton, Southampton, UK

<sup>2</sup>Department of Biomedical Sciences, Medical School, Universidad de La Sabana, Bogota, Colombia

Licensee OA Publishing London 2013. Creative Commons Attribution Licence (CC-BY)

**FOR CITATION PURPOSES:** Collins A, Arias L, Pengelly R, Martínez J, Briceño I, Ennis S. The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes. OA Genetics 2013 Sep 01;1(1):10.

the application of NGS to further characterise underlying genetic variation and its role in NSCLP.

### Genetic factors and NSCLP

NSCLP is a genetically complex disorder<sup>8</sup> which results from interactions between multiple genetic and environmental risk factors. The disorder has a significant genetic basis and it is known that first degree relatives of affected individuals have a 30–40 fold increased risk compared to the background population<sup>3,9</sup>. The degree of phenotype concordance for monozygotic (MZ) twins is 40–60% compared to 5% for di-zygotic twins. Murray<sup>10</sup> and Grosen et al.<sup>11</sup> found heritability estimates exceeding 90% for CLP phenotypes. Genetic studies including linkage analysis<sup>12,13</sup>, genome-wide association, and GWAS-based meta-analysis<sup>14</sup>, have yielded reproducible evidence for several genes and gene regions<sup>8</sup>. Results from Ludwig et al.<sup>14</sup> identified four genes and gene regions (IRF6, 8q24, 17q22 and 10q25; Table 1) for which the total population attributable risk is ~55% suggesting that, unusually for a complex trait, a substantial proportion of the variation in NSCLP might be explained by these loci<sup>15</sup>. However, many uncertainties remain. Poor concordance between regions identified by linkage with those found by association mapping (Table 1) must reflect in part the different targets of the techniques. Association mapping is good for detecting common variants contributing small effect sizes in population samples whereas linkage mapping is more powerful where there is allelic heterogeneity, for example where multiple (rare) variants in a particular gene contribute to disease. But, although many of these signals have been replicated in independent samples, several of the linkage regions are broad and the underlying causal gene(s) are poorly established. Incomplete knowledge about gene function presents difficulties for selecting the most likely

**Table 1** Some genes and gene regions implicated in non-syndromic cleft lip and/or palate

Nearest or causal gene	Region	Protein function	Method	Reference
PAX7	1p36	Transcription factor: neural crest development in mouse	Association	14
ARHGAP29	1p22	Regulation of binding proteins involved in craniofacial development	Association	14
IRF6	1q32	Involved in formation of connective tissue	Association/linkage	12–14
THADA	2p21	Possible regulatory functions	Association	14
TGFA	2p13	Involved in signalling pathway for cell proliferation, differentiation and development	Linkage	12
EPHA3	3p11	Regulation of cell shape and cell:cell contacts	Association	14
-	8q21	Intergenic region	Association	14
-	8q24	Gene desert: may contain regulatory elements for craniofacial development	Association	14
FOXE1	9q21	Transcription factor regulating diverse developmental processes	Linkage	12, 13
VAX1	10q25	Transcriptional regulator	Association	14
SPRY2	13q31	Signal is inter-genic, nearest gene is regulator of multiple receptor tyrosine kinases	Association	14
PAX9,TGFB3, BMP4	14q 22-24	BMP4: bone morphogenetic protein involved in bone/cartilage development	Linkage	12, 13
TPM1	15q22	Signal is inter-genic, in regulatory region, gene encodes actin-binding protein	Association	14
FOXC2, CRISPLD2	16q24	FOXC2: transcription factor, possible role in development of mesenchymal tissues	Linkage	12, 13
NOG	17q22	Essential for cartilage morphogenesis and joint formation	Association	14
MAFB	20q12	Transcription factor involved in development of keratinocytes	Association	14

Licensee OA Publishing London 2013. Creative Commons Attribution Licence (CC-BY)

**FOR CITATION PURPOSES:** Collins A, Arias L, Pengelly R, Martínez J, Briceño I, Ennis S. The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes. OA Genetics 2013 Sep 01;1(1):10.

candidate gene(s) in these regions. Several gene regions identified by linkage in earlier studies<sup>12</sup> have not replicated subsequently in independent samples<sup>13</sup>. Successful replication is difficult to achieve and it is perhaps too early to dismiss some of the more uncertain signals. Association mapping frequently reveals variants in inter-genic regions which are suggested to have regulatory functions influencing gene(s) nearby. Such a mechanism has only been firmly established for a small number of regions and identifying the precise causal variant(s) is made more difficult because of extensive linkage disequilibrium. It is particularly difficult to understand the precise functional roles of these apparent regulatory variants. Frequently, the nearest gene and/or gene with the most plausible NSCLP-related function is highlighted (Table 1). Other issues which have not been resolved by linkage and association studies include causes of apparent differences in the underlying genetic basis of NSCLP between populations of different ethnicity. One region which has been extensively studied is 8q24 for which Murray et al.<sup>16</sup> found much stronger evidence in European-derived samples, compared to Asians. In this case, the difference was attributed to reduced haplotype diversity in the Asian sample reducing power, rather than a distinct genetic effect. It is far from clear that such a mechanism accounts for ethnic genetic differences in other candidate gene regions.

Although understanding of the genetic basis of NSCLP phenotypes has advanced considerably in recent years, many unanswered questions remain, for which NGS may offer a route to progress. NGS has the potential to identify novel genes and other sources of causal variation (such as differences in copy number) which contribute to NSCLP. Furthermore, because sequencing can identify most DNA variants (rather than common 'tag' single-nucleotide polymorphisms, as in GWAS), it has the

potential to help determine actual causal variants rather than assignments to a broader region. The sequencing of many NSCLP genomes will be essential to establish models which consider the roles of regulatory sequences and the genes involved.

#### Environmental factors and gene-environment interaction

Although high MZ concordance is consistent with substantial genetic influences, the incomplete concordance suggests non-genetic influences on NSCLP phenotypes. Environmental effects might generate incomplete penetrance through random developmental events or a non-homogeneous in utero environment<sup>10</sup>. Grosen et al.<sup>11</sup> pointed out that MZ twin discordance might reflect genetic, cytogenetic or epigenetic anomalies in the affected twin that are not found in the unaffected twin. Post-zygotic genomic alterations resulting from mitotic recombination have been considered but have been shown by Kimani et al.<sup>17</sup> to not be a common cause of MZ twin discordance in CLP. Their analysis did not exclude rare or balanced genomic alterations, tissue-specific events and small aberrations beyond the resolution of their methods (~1Mb). Sequence-level resolution achieved by NGS might be informative given appropriately designed studies.

Establishing relationships between genetic and environmental factors has proven extremely challenging so far. Skare et al.<sup>9</sup> conducted a large study aimed at detecting interactions between 334 candidate genes and maternal first trimester exposure to smoking, alcohol, coffee, folic acid supplements, dietary folate and vitamin A. This study contrasted 425 case-parent triads with 562 control-parent triads. Very little evidence for gene-environment interaction was found in these data. They noted that 'it is remarkable that OC, a phenotype of supposedly very high heritability, remains so hard to decipher'. The authors consider that larger sample sizes and, therefore, greater power

to establish effects are required. Butali et al.<sup>18</sup> examined interactions between the MTHFR gene C677T variant and folic acid in OC aetiology<sup>19</sup>. They contrasted 1149 isolated cases and 1161 controls and considered maternal peri-conceptual exposure to smoking, alcohol and folic acid. Although folic acid and smoking were found to influence OC outcomes, no significant interaction was demonstrated with the C677T variant. Beatty et al.<sup>20</sup> found some evidence for gene-environment interaction using available data on maternal smoking during pregnancy in European case-parent trios. The genes involved were GRID2 and ELAVL2. However, neither gene showed evidence of association with NSCLP in the absence of the smoking interaction effect.

Efforts have been made to understand the underlying molecular mechanisms behind NSCLP and their relationship to genetic and environmental factors. Studies contrasting the transcriptome of dental pulp stem cells from NSCLP patients with controls suggest that there are alterations in gene networks (differentially expressed genes) functionally relevant to orofacial development, such as collagen metabolism and extracellular matrix remodelling<sup>21</sup>. Because NSCLP is considered to arise through anomalies in cellular migration, proliferation, trans-differentiation and apoptosis<sup>21,22</sup> Kobayahi et al.<sup>23</sup> considered possible overlap between NSCLP and cancer gene pathways. They demonstrated that NSCLP patient-derived stem cells show dysregulation in gene networks controlling cellular defences against DNA damage. The authors speculate that alterations in a small number of upstream genetic or epigenetic regulators, combined with deleterious genetic variants could disrupt the modulating activity of transcription factors such as E2F1. Hence genetic and epigenetic variation underlying regulatory anomalies, combined with environmental factors, may be driving NSCLP. Continuing progress with

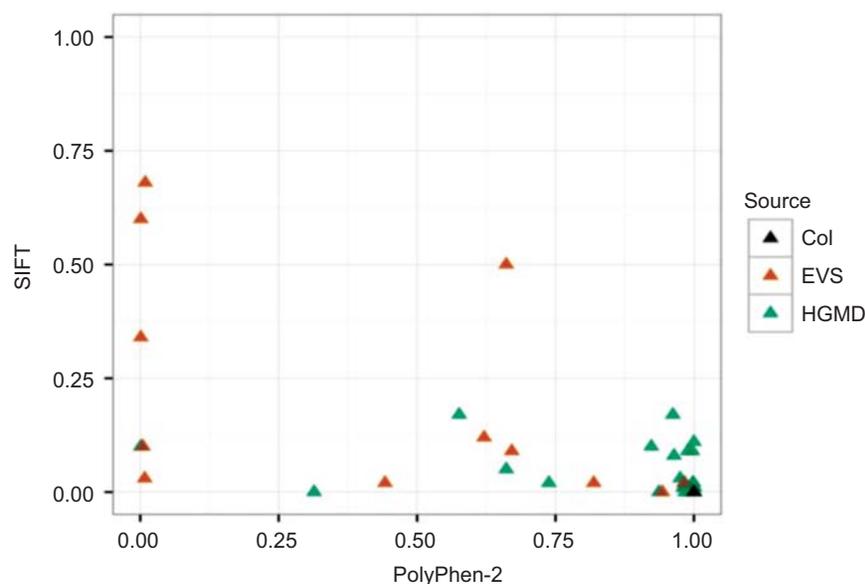
this functional work is hampered, in part, by incomplete understanding of underlying genetic factors. Specifically, it is not clear that the small number of NSCLP variants identified thus far can account for the dys-regulation of cellular functions and pathways identified and none of the differentially expressed genes identified correspond to the known GWAS variants<sup>23</sup>. To investigate how regulatory anomalies underlie the development of NSCLP, it is necessary to characterise the complete spectrum of variation in genome sequences, including all regulatory variants in non-coding regions<sup>24</sup>.

#### NGS: some practical considerations

Although NGS has enjoyed dramatic success through the identification of genes underlying Mendelian disorders<sup>25</sup>, and also de novo disease mutations<sup>26</sup>, complex phenotypes such as NSCLP have proven much more difficult to elucidate. Aside from numerous data quality control, technical and data management issues, a particular difficulty arises from the many, often apparently deleterious DNA variants, identified in each DNA sample. Faced with this complexity, various methods to 'filter' variants lists are undertaken to try to exclude 'neutral' variation. This procedure involves removal of 'common' variants (those represented in high frequency in databases of sequence variants from individuals lacking recognised disease) and removal of implausible disease candidates in genes which are highly mutable<sup>27</sup>. Such genes include those with sensory or immune functions for which high allele diversity is adaptive. Frequently, it is cost-effective to sequence only the protein coding exons of genes (the 'exome'), representing only 1% of the genome. From an exome sequence, non-synonymous variants (those that change an amino acid in the protein) can be selected for further study and other variants

excluded. A disadvantage of using only the exome and extensive filtering is that variants in non-coding regions, which may have regulatory functions, along with much of the structural variants, are excluded. Even a highly filtered list of non-synonymous variants may contain many potentially deleterious variants, which do not in fact influence the phenotype. Various predictive metrics such as SIFT and PolyPhen2 have been developed which help discriminate potentially deleterious variants from those that are neutral. SIFT predicts whether an amino acid substitution affects protein function based on conservation of amino acid residues across species<sup>28</sup>. PolyPhen2 considers impacts of an amino acid substitution on the structure and function of a protein<sup>29</sup>. Low scores (~0) for SIFT and high scores (~1) for PolyPhen2 suggest that the variant may be deleterious and contribute to disease. Figure 1 presents SIFT and PolyPhen2 scores for non-synonymous variants in the

IRF6 gene, which contains variants involved in both syndromic forms of CLP (Van der Woude (VDW) and popliteal pterygium syndrome<sup>30</sup> and NSCLP (Table 1). Scores for variants in this gene from the Exome Variant Server (EVS: a database of 6400 exome sequences)<sup>31</sup> and known disease causal variants from the Human Gene Mutation Database (HGMD)<sup>32,33</sup> are shown. The score for an exome-sequenced patient from Colombia with typical popliteal pterygium syndrome, who has the rs121434226 single-nucleotide polymorphism in IRF6, is also given. Although there is a degree of separation between known neutral and known causal variations (and the Colombian patient score is clearly deleterious by both measures), there is also overlap. These functional predictive methods can be useful for ranking variants worthy of further investigation but are not fully discriminatory, particularly for complex phenotypes where individual variants have reduced penetrance.



**Figure 1:** Polyphen2 and SIFT scores of variants in the IRF6 gene. PolyPhen-2 versus SIFT scores for non-synonymous variants in the IRF6 gene. Presumed neutral variants from the Exome Variant Server (EVS, n = 12), variants reported to cause CLP from the Human Gene Mutation Database (HGMD, n = 80) and the aetiological SNP from an exome sequenced in our Colombian patient (Col) are shown. Variants known to cause CLP are clearly clustered in the bottom right of the plot, representing a predicted deleterious nature by both metrics.

Licensee OA Publishing London 2013. Creative Commons Attribution Licence (CC-BY)

**FOR CITATION PURPOSES:** Collins A, Arias L, Pengelly R, Martínez J, Briceño I, Ennis S. The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes. OA Genetics 2013 Sep 01;1(1):10.

## Discussion

The identification of genetic factors underlying NSCLP has proven extremely challenging although recent progress with GWAS, and subsequent meta-analyses, have firmly implicated a number of genes and variants in NSCLP phenotypes. Although multiple loci identified through GWAS appear capable of explaining a relatively high proportion of the heritability low concordance between linkage and association studies strongly suggests that rarer variations, which can be detected by NGS, will provide additional causal insights. NGS sequencing studies will be invaluable for fine mapping causal variants in linkage and GWAS-identified genes and in pursuing additional, rarer, variations in related genes and pathways, along with novel genes. Only by developing a greater understanding of the underlying genetic basis of NSCLP will efforts to understand gene-environment interaction and functional processes underlying NSCLP be successful.

NGS also has the potential to contribute to understanding of the roles of different genetic factors amongst different ethnic groups and how these interact with diverse environmental influences. NGS is also capable of delimiting distinct disease sub-types within the NSCLP 'umbrella' which is important for refining diagnosis and tailoring treatment. The development of integrated models which consider gene-gene and gene-environment interaction and how these influence the function of key pathways will underpin more complete understanding. Although exome sequencing is valuable, whole genome sequencing of many individuals from different populations, comprehensive phenotyping, and careful consideration of environmental factors, may be required for establishing regulatory roles of some variants. However, NGS presents considerable challenges for data analysis and interpretation. Much effort is now focussed on

addressing these difficulties and, as many more genomes are sequenced, further success in understanding the role of genes in NSCLP phenotypes is expected.

## Conclusion

Although important recent advances have revealed some of the genetic variants underlying NSCLP, NGS has the potential to identify novel genetic factors. Only given more comprehensive understanding of genetic variation underlying NSCLP can interactions between genes, and between genes and environmental variables, be firmly identified. Success with NGS will lead to improvements in prediction, prevention and treatment for cleft lip and palate patients.

## Acknowledgements

The authors gratefully acknowledge funding from the Newlife Foundation for Disabled Children.

## References

1. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*. 2013 Sep;155(1):27–38.
2. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development*. 2009 Jun;19(3):212–219.
3. Wehby GL, Cassell CH. The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral diseases*. 2010 Jan;16(1):3–10.
4. Mossey PA, Little J. Epidemiology of oral clefts: an international perspective. *Cleft lip and palate: from origin to treatment*. 2002;127–158.
5. Clark JD, Mossey PA, Sharp L, Little J. Socioeconomic status and orofacial clefts in Scotland, 1989 to 1998. *The Cleft palate-craniofacial J*. 2003 Sep;40(5):481–485.
6. Durning P, Chestnutt P, Morgan MZ. The relationship between orofacial clefts and material deprivation in Wales. *The Cleft palate-craniofacial J*. 2007 Mar;44(2):203–207.
7. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *The Lancet*. 2009 Nov;374(9703):1773–1785.

8. Leslie EJ, Marazita ML. Genetics of cleft lip and cleft palate. *Am J Med Genetics Part C: Seminars in Medical Genetics*. 2013 Oct;163(4):246–258.
9. Skare Ø, Jugessur A, Lie RT, Wilcox AJ, Murray JC, Lunde A, et al. Application of a Novel Hybrid Study Design to Explore Gene-Environment Interactions in Orofacial Clefts. *Annals of human genetics*. 2012 May;76(3):221–236.
10. Murray JC. Gene/environment causes of cleft lip and/or palate. *Clinical genetics*. 2002 May;61(4):248–256.
11. Grosen D, Bille C, Petersen I, Skytthe A, von Bornemann-Hjelmberg J, Krabbe J et al. Risk of oral clefts in twins. *Epidemiology (Cambridge, Mass.)*. 2011 May;22(3):313.
12. Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, et al. Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Human Genetics*. 2004 Aug;75(2):161–173.
13. Marazita ML, Lidral AC, Murray JC, Field LL, Maher BS, Cooper ME, et al. Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. *Human heredity*. 2009 Jul;68(3):151–170.
14. Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, et al. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nature genetics*. 2012 Aug;44:968–971.
15. Marazita ML. The evolution of human genetic studies of cleft lip and palate. *Annual Review of Genomics and Human Genetics* 2013 Sep;13:263–283.
16. Murray T, Taub MA, Ruczinski I, Scott AF, Hetmanski JB, Schwender H, et al. Examining markers in 8q24 to explain differences in evidence for association with cleft lip with/without cleft palate between Asians and Europeans. *Genetic Epidemiology*. 2012 May;36(4):392–399.
17. Kimani JW, Yoshiura K, Shi M, Jugessur A, Moretti-Ferreira D, Christensen K, et al. Search for genomic alterations in monozygotic twins discordant for cleft lip and/or palate. *Twin Research and Human Genetics*. 2009 Oct;12(5):462–468.
18. Butali A, Little J, Chevrier C, Cordier S, Steegers-Theunissen R, Jugessur A, et al. Folic acid supplementation use and the MTHFR C677T polymorphism in orofacial

Licensee OA Publishing London 2013. Creative Commons Attribution Licence (CC-BY)

**FOR CITATION PURPOSES:** Collins A, Arias L, Pengelly R, Martínez J, Briceño I, Ennis S. The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes. *OA Genetics* 2013 Sep 01;1(1):10.

- clefts etiology: an individual participant data pooled-analysis. *Birth Defects Research Part A: Clinical and Molecular Teratology*. 2013 Aug;97(8):509–514.
19. Martinelli M, Scapoli L, Pezzetti F, Carinci F, Carinci P, Stabellini G, et al. C677T variant form at the MTHFR gene and CL/P: a risk factor for mothers?. *Am J Med Genet*. 2001 Feb;98(4):357–360.
20. Beaty TH, Taub MA, Scott AF, Murray JC, Marazita ML, Schwender H, et al. Confirming genes influencing risk to cleft lip with/without cleft palate in a case–parent trio study. *Human genetics*. 2013 Jul;132(7):771–781.
21. Bueno DF, Sunaga DY, Kobayashi GS, Agüena M, Raposo-Amaral CE, Masotti C, et al. Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. *Stem Cell Reviews and Reports*. 2011 Jun;7(2):446–457.
22. Baroni T, Bellucci C, Lilli C, Pezzetti F, Carinci F, Lumare E, et al. Human cleft lip and palate fibroblasts and normal nicotine-treated fibroblasts show altered in vitro expressions of genes related to molecular signaling pathways and extracellular matrix metabolism. *J Cellular physiology*. 2010 Mar;222(3):748–756.
23. Kobayashi GS, Alvizi L, Sunaga DY, Francis-West P, Kuta A, Almada BVP, et al. Susceptibility to DNA Damage as a Molecular Mechanism for Non-Syndromic Cleft Lip and Palate. *PloS One*. 2013 Jun;8(6):e65677.
24. Ward LD, Kellis M. Interpreting non-coding genetic variation in complex traits and human disease. *Nature Biotechnol*. 2012 Nov;30(11):1095–1106.
25. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nature genetics*. 2009 Nov;42(1):30–35.
26. Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature genetics*. 2010 Jun;42(6):483–485.
27. Fuentes Fajardo KV, Adams D, Mason CE, Sincan M, Tiff C, Toro C, et al. Detecting false-positive signals in exome sequencing. *Human mutation*. 2012 Apr;33(4):609–613.
28. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 2009 Jun;4(7):1073–1081.
29. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010;7(4):248–249.
30. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nature genetics*. 2002 Sep;32(2):285–289.
31. NHLBI Exome Sequencing Project (ESP) Exome Variant Server. <http://evs.gs.washington.edu/EVS/> [accessed 15 October 2013].
32. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, et al. Human gene mutation database (HGMD®): 2003 update. *Human mutation*. 2003 Jun;21(6):577–581.
33. The Human Gene Mutation Database, <http://www.hgmd.org/> [accessed 15 October 2013].

Licensee OA Publishing London 2013. Creative Commons Attribution Licence (CC-BY)

**FOR CITATION PURPOSES:** Collins A, Arias L, Pengelly R, Martínez J, Briceño I, Ennis S. The potential for next-generation sequencing to characterise the genetic variation underlying non-syndromic cleft lip and palate phenotypes. *OA Genetics* 2013 Sep 01;1(1):10.