**Research Letter**

 **Reassignment of *HMX1* Indicates Copy Number Variation within 4p16.1 may be an Alternative Cause of Oculoauricular Phenotypes.**

John CK Barber

Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton General Hospital, Southampton, UK.

Correspondence
John C.K. Barber, Department of Human Genetics and Genomic Medicine,
Faculty of Medicine, University of Southampton, Southampton General Hospital,
Tremona Road, Southampton, SO16 6YD, UK.
Email: john.barber@soton.ac.uk
tel +44 (0)2380 796586; +44 (0)2380 794760;

Running title: Copy number variation and *HMX1*

KEYWORDS
copy number variation; 4p16.1; oculoauricular phenotypes; *HMX1*; evolutionarily conserved region; euchromatic variant*.*

Recessive loss of function mutations in the H6 family homeobox 1 transcription factor gene (*HMX1*) have been associated with oculoauricular anomalies (OMIM 142992) and the rare Oculoauricular Syndrome (OMIM 612109) [Schorderet et al., 2008]. Here, it is proposed that the reassignment of *HMX1,* recent evidence from animals and overlapping copy number variations (CNVs) indicate that CNV of a downstream evolutionarily conserved enhancer region (ECR) of *HMX1* and/or the gene itself may be alternative causes of these and other phenotypes.

In man, five copies of a 750 kb amplicon within 4p16.1 were found to co-segregate with microtia and nasolacrimal-duct imperforation in 10 members of a three generation family of whom 3 also had eye coloboma [Balikova et al., 2008]. As this amplification was cytogenetically visible, contained a high proportion of segmental duplications (SDs) and, at the time, no relevant dosage sensitive genes (Figure 1a), this author suggested that it might be a benign euchromatic variant (EV) [Gardner et al., 2012]. However, no CNVs of this amplicon were found in 200 normal individuals [Balikova et al., 2008] and the majority of CNVs in the Database of Genomic Variants (DGV) now coincide with the SDs and the sequence gap between them (Figure 1a).

The *HMX1* gene was originally mapped to 4p16.3 (hg18 chr4:279,194-283,964) but subsequently reassigned to 4p16.1 (hg 19 chr4:8,868,773-8,873,543). Although not peer reviewed, it was suggested in a more recent abstract that a 1.55 Mb duplication of the *HMX1* and carboxypeptidase Z(*CPZ*) genes might be the cause of the microtia and hemifacial macrosomia found in an affected proband and apparently unaffected mother [Dipple et al., 2013](Figure 1a). This family and the reassignment of *HMX1* support the idea that the original amplification was indeed a rare pathological quintuplication CNV of 4p16.1 [Balikova et al., 2008].

In the dumbo rat, misplaced ears mouse and crop eared highland cattle, duplication and deletion of an ~300 base pair non-coding ECR downstream of *HMX1* has also been associated with a number of abnormalities of the ears and eyes [Quina et al., 2012; Koch et al., 2013; Rosin et al., 2016]. Proteins produced by three Homeobox genes involved in embryonic patterning in mice (Homeobox a2, Myeloid ecotropic integration site homeobox and pre-B-cell leukemia homeobox 1) are thought to regulate *Hmx1* expression by acting cooperatively on this ECR via a core 32 bp sequence [Rosin et al., 2016]. Using the UCSC lift over tool, the duplicated ECR in cattle corresponds to a 307 bp region between *HMX1* and the *CPZ* genes at a locus ~ 165 kb downstream of *HMX1* in humans (chr4:8,702,159-8,702,465) (Figure 1a,b). This region contains three conserved transcription factor binding sites (Figure 1b) for proteins of the CCAAT displacement (*CDP*) Cut-like homeodomain (V$CDPCR3HD\_01),pre-B-cell leukemia transcription factor 1(*PBX1*) (V$PBX1 02) and LIM domain only 2 (*LM02*) (V$LMO2COM 02) genes. Of these only the *PBX1* binding site is in common with the mouse and is a gene that has been associated with external ear anomalies and intellectual disability in man [Slavotinek et al., 2017].

CCAAT displacement proteinCNVs that overlap or extend by 500 kb either side of the region defined by the CNVs of Balikova et al. [2008] and of Dipple et al. [2013] have been examined for the presence or absence of oculoauricular phenotypes (Figure 1a). The DECIPHER database contained 6 microdeletions ranging from 83 kb to 2.1 Mb in size of which five had phenotypic information and four had no relevant phenotypes. The only one considered likely to be pathogenic had intrauterine growth retardation, chorioretinal coloboma, iris coloboma, and choroid coloboma (DECIPHER 283751) and a paternally inherited 1.7 Mb duplication that included both *HMX1* and the ECR. However, this patient also had a de novo 9.4 Mb duplication of the long arm of chromosome 2 regarded as pathogenic. Nevertheless, undetected CNV may explain why Li et al. [2014] mapped a microtia locus to a 10 Mb interval of 4p15.32–4p16.2 but could find no mutations after sequencing candidate genes that included *HMX1*. ClinGen had no microdeletions that fitted the same criteria.

Against this proposal are the multiple larger overlapping CNVs with no relevant phenotypes, the apparently benign gain and/or loss CNVs in the DGV that are less than 3 Mb in size and include *HMX1* (5), *HMX1* and the ECR (3) or the ECR without *HMX1* (5). In addition, the single maternally inherited CNV that encompassed the ECR alone was not associated with an oculoauricular phenotype but with delayed speech and language and a low anterior hairline (DECIPHER 288656) (Figure 1a). Thus, if CNV of the ECR is significant, it must be subject to the incomplete penetrance and variable expressivity that is characteristic of other pathogenic non-coding CNVs [Zhang and Lupski, 2015].

In conclusion, (1) high dosage of the downstream ECR and/or *HMX1* may explain the highly penetrant oculoauricular phenotypes associated with the amplification reported by Balikova et al [2008], (2) lower dosage downstream ECR and/or *HMX1* may result in similar phenotypes with incomplete penetrance and (3), it is important to re-examine previous analyses in the light of new builds of the human genome in case re-interpretation of the results is required. Further evidence for or against this proposal may be provided by including *HMX1* and the ECR in targeted arrays [Cox et al., 2014] and using whole genome sequencing as well as conformation capture technologies to determine whether this ECR is part of a distinct regulatory network in man responsible for oculoauricular and/or other phenotypes.

**CONFLICT OF INTEREST**The author declares no conflict of interest.

**WEB RESOURCES**

The Clinical Genome Resource (ClinGen) consortium: <https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/>

DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources): <https://decipher.sanger.ac.uk/>

DGV (Database of Genome Variants): <http://dgv.tcag.ca/dgv/app/home>
OMIM (Online Mendelian Inheritance in Man): <http://omim.org>

UCSC (University of Santa Clara, California) web browser: <http://genome.ucsc.edu/>

**References:**

Balikova I., Martens K., Melotte C., Amyere M., Van Vooren S., Moreau Y.,…& Vermeesch J.R. 2008. Autosomal-dominant microtia linked to five tandem copies of a copy-number-variable region at chromosome 4p16. *Am J Hum Genet*, 82, 181-187.

Cox T.C., Camci E.D., Vora S., Luquetti D.V., & Turner E.E. 2014. The genetics of auricular development and malformation: new findings in model systems driving future directions for microtia research. *Eur J Med Genet*, 57, 394-401.

Dipple K.M., Lee-Barber J., & Quintero-Rivera F. 2013. A case report of hemifacial microsomia with duplication of 4p16.1 including the HMX1 and CPZ1 genes. *ACMG annual meeting* (abstract 214).

Gardner R.J.M., Sutherland G.R., & Shaffer L.G. 2012. Chromosome abnormalities and genetic counseling. 4th edn. *Oxford University Press*, Oxford, UK.

Huang N., Lee I., Marcotte E.M., & Hurles M.E. 2010. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*, 6, e1001154.

Koch C.T., Bruggmann R., Tetens J., & Drögemüller C. 2013. A non-coding genomic duplication at the HMX1 locus is associated with crop ears in highland cattle. *PLoS One*, 8, e77841.

Li X., Hu J., Zhang J., Jin Q., Wang D.M., Yu J.,…& Zhang Y.B. 2014. Genome-wide linkage study suggests a susceptibility locus for isolated bilateral microtia on 4p15.32-4p16.2. *PLoS One*, 9, e101152.

Quina L.A., Kuramoto T., Luquetti D.V., Cox T.C., Serikawa T., & Turner E.E. 2012. Deletion of a conserved regulatory element required for Hmx1 expression in craniofacial mesenchyme in the dumbo rat: a newly identified cause of congenital ear malformation. *Dis Model Mech,* 5, 812-822.

Rosin J.M., Li W., Cox L.L., Rolfe S.M., Latorre V., Akiyama J.A.,…& Cox T.C. 2016. A distal 594 bp ECR specifies Hmx1 expression in pinna and lateral facial morphogenesis and is regulated by the Hox-Pbx-Meis complex. *Development*, 143, 2582-2592.

Schorderet D.F., Nichini O., Boisset G., Polok B., Tiab L., Mayeur H.,…Munier F.L. 2008. Mutation in the human homeobox gene NKX5-3 causes an oculo-auricular syndrome. *Am J Hum Genet*, 82, 1178-1184.

Slavotinek A., Risolino M., Losa M., Cho M.T., Monaghan K.G., Schneidman-Duhovny D.,…& Shieh J. 2017. De novo, deleterious sequence variants that alter the transcriptional activity of the homeoprotein PBX1 are associated with intellectual disability and pleiotropic developmental defects. *Hum Mol Genet*, 26, 4849-4860.

Zhang F. & Lupski J.R. 2015. Non-coding genetic variants in human disease. *Hum Mol Genet* 2015, 24(R1), R102-110.

**TITLES AND LEGENDS TO FIGURES**

**Figure 1** Screen shots from the UCSC web browser (GRCh37/hg19) of (**a**) the 2.1 Mb region of interest (ROI) of 4p16.1 (chr4:8,500,000-10,600,000) and (**b**) the 550 bp region that contains the evolutionarily conserved region (ECR) (chr4:8,702,100-8,702,650). (**a**) The box on the idiogram of chromosome 4 shows the ROI in band 4p16.1; the black boxes frame the RefSeq genes with their Haploinsufficiency likelihood scores beneath (marked n/a where not available) [Huang et al., 2010]; Segmental Duplications (SDs) and a sequence gap are represented by the two blocks in shades of grey with a black bar between them; horizontal bars with dotted vertical lines mark the approximate minimum extent of overlapping gains from the literature and DECIPHER with source, extent and phenotypic summaries below each CNV; the vertical arrowed line indicates the position of the ECR; (**b**) Dotted vertical lines show the approximate minimum extent of the 307 bp ECR; horizontal grey lines illustrate ENCODE regulation data; black boxes indicate the three conserved transcription factor binding sites (V$CDPCR3HD\_01,V$PBX1 02 and V$LMO2COM 02) and the black bar a Weizmann conserved CpG island.
Data generated by the DECIPHER Consortium has been used and a full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

