A single nucleotide polymorphism associated with isolated cleft lip and palate, thyroid cancer and hypothyroidism alters the activity of an oral epithelium and thyroid enhancer near FOXE1

Andrew C. Lidral1,*, Huan Liu2,5, Steven A. Bullard3, Greg Bonde4, Junichiro Machida6, Axel Visel7,8, Lina M. Moreno Uribe1, Xiao Li4, Brad Amendt4 and Robert A. Cornell4

1Department of Orthodontics, 2Dows Research Institute, 3Department of Endocrinology and 4Department of Anatomy, University of Iowa, Iowa City, IA, USA, 5State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China, 6Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota City, Aichi, Japan, 7Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA and 8Department of Energy Joint Genome Institute, Walnut Creek, CA, USA

*To whom correspondence should be addressed at: Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA 52241, USA. Tel: +1 3193358498; Fax: +1 3193356790; Email: andrew-lidral@uiowa.edu

Abstract

Three common diseases, isolated cleft lip and cleft palate (CLP), hypothyroidism and thyroid cancer all map to the FOXE1 locus, but causative variants have yet to be identified. In patients with CLP, the frequency of coding mutations in FOXE1 fails to account for the risk attributable to this locus, suggesting that the common risk alleles reside in nearby regulatory elements. Using a combination of zebrafish and mouse transgenesis, we screened 15 conserved non-coding sequences for enhancer activity, identifying three that regulate expression in a tissue specific pattern consistent with endogenous foxe1 expression. These three, located −82.4, −67.7 and +22.6 kb from the FOXE1 start codon, are all active in the oral epithelium or branchial arches. The −67.7 and +22.6 kb elements are also active in the developing heart, and the −67.7 kb element uniquely directs expression in the developing thyroid. Within the −67.7 kb element is the SNP rs7850258 that is associated with all three diseases. Quantitative reporter assays in oral epithelial and thyroid cell lines show that the rs7850258 allele (G) associated with CLP and hypothyroidism has significantly greater enhancer activity than the allele associated with thyroid cancer (A). Moreover, consistent with predicted transcription factor binding differences, the −67.7 kb element containing rs7850258 allele G is significantly more responsive to both MYC and ARNT than allele A. By demonstrating that this common non-coding variant alters FOXE1 expression, we have identified at least in part the functional basis for the genetic risk of these seemingly disparate disorders.
**Introduction**

Isolated cleft lip with or without cleft palate (CL/P) and cleft palate (CP) are common structural birth defects with a strong genetic component that remains incompletely understood. While periconceptional and gestational environmental influences contribute (smoking, alcohol and nutrition), twin, family and population-based studies have clearly established a genetic basis for CL/P etiology. For instance, there is a 40-fold increased prevalence among first-degree relatives compared with within the general population (1). Moreover, genome-wide association studies (GWAS) have identified at least 18 loci in which certain alleles of specific single nucleotide polymorphisms (SNPs) are significantly associated with isolated CL/P (2–5). However, coding mutations have been identified in <10% of patients. This follows the trend of GWAS of other diseases, where a meta-analysis of over 150 GWAS studies revealed only 12% of associated SNPs were in a haplotype block containing coding exons (6). Mutations landing in non-coding DNA are presumed, in many or most cases, to disrupt cis-regulatory elements (CREs) (i.e. promoters and enhancers). However, our understanding of how sequence variation alters the function of regulatory elements remains in its infancy. Consequently, very few mutations that directly elevate risk for CL/P have been identified, hindering progress in the investigation of disease mechanisms (7–8).

We have recently identified a strong peak of association for both CL/P and CP at 9q22, encompassing the gene FOXE1 (9). This association has been replicated in a variety of Caucasian populations (10,11) and Hispanics from Honduras (12). Importantly, FOXE1 is likely to be the relevant gene, because homozygous FOXE1 mutations cause Bamforth–Lazarus syndrome which is characterized by cleft palate, choanal atresia, bifid epiglottis, thyroid agenesis or dysgenesis, hypothyroidism and spiky hair (13–15). Inactivation of FOXE1 in mice causes a similar phenotype (16). We have found FOXE1 coding mutations in only 1% of subjects, despite exhaustive mutation screens (9). This is incongruent with estimates that FOXE1 has a 25–38% attributable risk, suggesting that causal variants occur in nearby cis-regulatory elements.

Consistent with thyroid defects in patients with Bamforth–Lazarus syndrome, a variety of common thyroid diseases also map to FOXE1, but causal variants for these disorders at this locus remain largely unknown. Thyroid disorders mapping to 9q22 include congenital hypothyroidism due to thyroid dysgenesis (17–19), hypothyroidism (20,21), goiters, non-medullary thyroid cancer (22,23), papillary thyroid carcinoma (24–27), radiation induced papillary thyroid carcinoma as a result of the Chernobyl accident (28,29), and thyroid cancer (30–32). Also, a number of biomarkers of thyroid metabolism are also associated at this locus (27,30–33–37). These observations are consistent with evidence that FOXE1 is involved in thyroid homeostasis, negatively regulating thyroglobulin and thyroperoxidase expression (38). Interestingly, a SNP associated with papillary thyroid cancer, rs1867277, is located in the FOXE1 promoter and alters binding of the USF1/USF2 transcription factors (26). However, rs1867277 is unlikely to explain all of the risk for thyroid cancer at this locus, because other SNPs show more significant association, and because the haplotype block structure in the region suggests more than one risk allele is present for these diseases. Although FOXE1 coding mutations may be present in a subset of patients with thyroid conditions mentioned earlier, it is probable that additional causative variants in non-coding DNA remain to be identified.

We hypothesized that causal mutations for orofacial clefting and thyroid diseases occur within cis-regulatory elements at the human FOXE1 locus. To that end, we screened evolutionarily conserved sequences of human DNA at the FOXE1 locus (Fig. 1) for craniofacial and thyroid enhancers using both transgenic zebrafish and mouse strategies that have been shown to effectively identify cis-regulatory elements (39,40). We identified three such elements, including one containing a SNP that is associated with CL/P, hyperthyroidism and thyroid cancer. In vitro quantitative reporter assays in embryonic oral epithelial and thyroid cells demonstrated that different alleles of this SNP altered the level of enhancer activity of the element.

**Results**

**Endogenous foxe1 expression during zebrafish development**

We reasoned that oral and thyroid enhancers near human FOXE1 would be more likely to be functional in zebrafish if the fish ortholog, foxe1, were expressed in oral tissues and thyroid gland. Indeed, RNA in situ hybridization revealed foxe1 expression in the central nervous system at 11.4 h post fertilization (hpf) (Fig. 2A), in oral epithelium at 36 hpf (Fig. 2B, C, E and F, data not shown) and thyroid gland (Fig. 2D and G) at 36 hpf and later stages. We detected foxe1 expression in pharyngeal arch epithelium at 72 and 96 hpf (Fig. 2H and I). Furthermore, we observed foxe1 expression in the developing heart at 48 hpf, the pectoral fin starting at 72 hpf and the gill rakes (data not shown). These observations confirm and extend a previous expression analysis (41).

**Enhancer screen results**

We amplified from human DNA, 15 human conserved non-coding elements (hsCNEs) in a 152 kb region spanning from 101 kb upstream and 51 kb downstream of FOXE1 and tested them for enhancer activity in transient transgenic zebrafish (Fig. 1; Supplementary Table 1). The hsCNEs are named by their position relative to the FOXE1 start codon, negative being upstream and positive downstream. Our zebrafish enhancer screen was based on an eGFP reporter, pT2cfosGW eGFP (39). This reporter construct is designed to clone the human test DNA upstream of the mouse minimal cfos promoter that is unable to independently initiate mRNA transcription. To confirm this, we injected >200 single cell embryos with the reporter lacking any human sequences and did not detect any GFP expression. In contrast, of 15 hsCNEs, three had enhancer activities capable of directing transcription in a pattern consistent with endogenous foxe1 expression (hsCNE+22.6, hsCNE-67.7 and hsCNE-82.4; Supplementary Material, Table S1), as summarized below. These positive hsCNEs acted independently and without the FOXE1 promoter.

**FOXE1 hsCNE+22.6 is active in the developing pharyngeal arches and heart**

In embryos injected with FOXE1 hsCNE+22.6 kb element (hg19 chr9: 100 638 752–100 639 947), we detected a reproducible pattern of eGFP expression in a mosaic pattern within pharyngeal muscles, a subset of cranial muscles (see Fig. 3 legend), heart and pharyngeal arch epithelium (n > 800 embryos scored) (Fig. 3). We established five independent, stable transgenic lines and all exhibited non-mosaic GFP expression in the same structures that
expressed GFP mosaically in transient transgenic G0 embryos. In stable transgenic embryos, GFP was evident in the developing mouth by 24 hpf (data not shown) and by 72 hpf the muscles of the second and branchial arch were labeled along with the branchial arch cartilages including the gill rakes (Fig3A–C). Moreover, this element also directed GFP expression in the developing heart and pectoral fins (Fig. 3A–C). This pattern was confirmed by in situ hybridization with an eGFP probe. While aspects of this pattern matched that of foxe1 expression, for instance in epithelium of ceratobranchial arches (e.g. compare Fig. 3D to Fig. 1H), however...
expression of foxe1 mRNA was not detected in pharyngeal muscles. If human FOXE1 is also absent from pharyngeal muscles, we infer that additional regulatory elements must normally prevent the enhancer at +22.6 kb from driving FOXE1 expression in this tissue. Alternatively, through divergence in tissue specific expression of upstream regulatory factors, this enhancer may be active in pharyngeal muscles only in fish, although ENCODE data show this region has chromatin marks of enhancer activity in human skeletal muscle myoblasts.

While hsCNE+22.6 was able to activate lacZ expression in human embryonic oral epithelial cell lines, no difference was observed between the two rs10984103 alleles (Supplementary Material, Fig. S1).

**FOXE1 hsCNE-67.7 drives expression in the developing oral epithelium, hyoid arch, branchial arches, thyroid and heart**

In transient transgenic embryos injected with the FOXE1 -67.7 element (hg19 chr9: 100 548 460–100 549 280), we reproducibly detected GFP expression in the jaw, hyoid and pharyngeal arches, thyroid and heart (data not shown). The same pattern was observed in five independent transgenic lines and in these, expression in the oral epithelium was more evident than in transient transgenics (Fig. 4A and B). Expression in oral epithelium and thyroid precursors was detectable by 24 hpf, while heart expression became evident at 48 hpf (data not shown). Staining with the thyroid specific T4 antibody demonstrated coincident signal with
eGFP within the thyroid in the hsCNE-67.7 transgenic fish (data not shown).

We next tested an overlapping region, (hg19 chr9: 100 548 557–100 549 640) in transient transgenic mouse embryos using a β-galactosidase reporter with an Hsp68 minimal promotor. This region, Vista Enhancer hs1595, drove reporter expression in a pattern matching endogenous foxe1 expression in the oral epithelium and developing thyroid in three of four mouse embryos (Fig. 4C–F, Supplementary Material, Fig. S2). In summary, the FOXE1 hsCNE-67.7 element has enhancer activity in multiple oral tissues, in the heart and the thyroid gland during development in zebrafish, and an overlapping human DNA element has similar activity in mouse embryos.

**Differential enhancer activity for rs7850258 alleles in FOXE1 hsCNE-67.7**

Interestingly, rs7850258, for which one allele (G) is associated with hypothyroidism and CLP, and another (A), with thyroid cancer, is present within the FOXE1 hsCNE-67 element. We engineered variants of this element, containing one or the other alleles of SNP rs7850258, into a vector containing a minimal promoter and the gene encoding firefly luciferase. Since Foxe1 is expressed in the oral epithelium during facial development (9), an embryonic oral epithelial cell line that endogenously expresses Foxe1, GMSM-K (human fetal oral epithelial cells), was used in these assays. For assessment in the context of thyroid function, the rat FRTL epithelial...
expression in the oral epithelium (white arrowheads). In all over 370 embryos were screened from seven different injection batches. Table 1. Scoring of expression in G0 embryos

<table>
<thead>
<tr>
<th>FOXE1 hsCNE</th>
<th>Total live embryos</th>
<th>Expressers (%)</th>
<th>Expression in oral or pharyngeal regionsa</th>
<th>Number transgenic lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>+22.6</td>
<td>822</td>
<td>41</td>
<td>29 (48 hpf) 18 (72 hpf) 13 (96 hpf)</td>
<td>5</td>
</tr>
<tr>
<td>−67.7</td>
<td>262</td>
<td>30</td>
<td>ND (24 hpf) ND (48 hpf) ND (72 hpf)</td>
<td>5</td>
</tr>
<tr>
<td>−82.4</td>
<td>453</td>
<td>19</td>
<td>17 (48 hpf) 13 (72 hpf) ND (96 hpf)</td>
<td>7</td>
</tr>
</tbody>
</table>

aPercent of live embryos.

Figure 5. FOXE1 hsCNE-82.4 directs GFP expression in the oral epithelium. (A and B) Lateral view line drawings representing a composite of transient G0 embryos at 48 and 72 hpf, showing eGFP labeled cells highlighted in pink. GFP expression is primarily in the oral epithelium. (C) Lateral view of a representative G0 larva at 96 hpf showing expression in the oral epithelium (white arrowheads). In all over 370 embryos were screened from seven different injection batches.


discussion

Zebrafish and mouse orthologs of FOXE1 are expressed in the oral epithelium and in the thyroid gland consistent with role for FOXE1 in diseases of both tissues

We have confirmed in zebrafish that fox1 is expressed in the developing oral and pharyngeal epithelium, thyroid gland and pectoral fins. Furthermore, fox1 is expressed in the heart, and in the developing midbrain to hindbrain. While fox1 expression in central nervous system has not been explicitly mentioned previously, it is evident in an earlier study in zebrafish [see Fig. 1 of (41)]. Similarly, in mice Foxe1 expression in the developing central nervous system and heart has not been previously reported, but a review of the literature shows expression in the midbrain and hindbrain on embryonic Day 10.5 (44). Furthermore, in the public domain via EuroExpress, Foxe1 mRNA expression is evident in the midbrain, cerebellum and heart outflow tracts in E14.5 mouse embryos (Supplementary Material, Fig. S3) (45). This is consistent with RT-PCR data showing in zebrafish adults very strong expression in the heart and brain (41).

Identification of human tissue specific enhancers at the FOXE1 locus

Using an strategy based primarily on sequence conservation, we tested 15 regions of human DNA totaling 14.9 kb, encompassing 9.8% of the human FOXE1 locus as defined by the CL/P critical region (9). Of these, three showed positive tissue specific enhancer activity in patterns consistent with endogenous fox1 expression. All three enhancers, hsCNE+22.6, hsCNE-67.7 and hsCNE-82.4 shared the same capacity to independently regulate expression in the oral and pharyngeal epithelium, and pharyngeal arch structures. Two of these, hsCNE+22.6 and hsCNE-67.7 are also heart and pectoral fin enhancers. Unique is the ability of hsCNE-67.7 to direct expression in the developing thyroid. The oral epithelial and thyroid activity of the -67.7 element was also confirmed in transient transgenic mice. In contrast, the enhancer activity of hsCNE-82.4 is primarily confined to the oral epithelium.

In several previous studies, human DNA, which is not detectably conserved to zebrafish, has been shown to nonetheless possess enhancer activity in zebrafish. Examples include enhancers for RET, a gene for Hirschsprung disease (46), for PHOX2B, which is associated with neuroblastoma and Crohn’s disease (47–49), and
one for NOS1AP that contains a functional SNP associated with cardiac QT interval (50). Similarly, the FOXE1 enhancers that we have identified here lack detectable conservation to zebrafish, the longest stretch of 100% identity being 26 bp and for any identity >85% being 47 bp of which none map to the zebrafish foxe1 locus (Supplementary Material, Table S1). While the interpretation of CNE-82.4 as a facial enhancer is based entirely on transgenic zebrafish embryos, our experience and published work has shown expression patterns in transient transgenic zebrafish embryos accurately reflect that found in founder lines (51,52). Nonetheless, there is no guarantee that every element conserved in mammals will have enhancer activity in zebrafish, and therefore negative results, as we observed for 12 of 15 tested elements, cannot be strongly interpreted. Alternative explanations for the negative results could be those elements act as silencers or insulators, which would not have been detected using the reporter construct in this study. Additional craniofacial enhancers for FOXE1 are likely to exist; criteria other than conservation, for instance chromatin marks in appropriate cell lines, may be of use to identify them.

In retrospect to see if we could improve upon our 20% success rate when choosing human candidate elements based on
conservation, we compared our results to available ENCODE data. Only the hsCNE+22.6 enhancer had chromatin marks and an abundance of ChIP-Seq peaks, clearly indicating regulatory activity. Beyond that, we could not identify a matching pattern for either positive or negative elements with chromatin marks for regulatory function in any cell type. This underscores the need to have such data available on the relevant cell type. This is especially true for studying embryonic structural birth defects for which there is a paucity of data resources.

Relevance to human diseases

The expression of FOXE1 in the developing heart suggests it may be associated with cardiac defects. To date, variations at the FOXE1 locus have not been associated with cardiac anomalies. Yet tangential data exist in that among infants with congenital hypothyroidism there is a 3–5-fold over representation of heart defects (53,54), suggesting they may have shared developmental processes.

At the FOXE1 locus, based on association peaks and LD blocks in different populations, there is evidence for three CL/P risk haplotypes relative to the start codon: −89 to −59 kb; −19.2 to +0.4 kb; and +7.0 to +81 kb (9). We postulate that rs7850258 within the hsCNE-67.7 explains entirely or in part the risk for CL/P associated with the most upstream risk haplotype. The facial enhancer, hsCNE+22.6, containing rs10984103, the SNP most strongly associated with the most upstream risk haplotype. The facial enhancer, hsCNE+22.6, containing rs10984103, the SNP most strongly associated with the most upstream risk haplotype. The facial enhancer, hsCNE+22.6, containing rs10984103, the SNP most strongly associated with the most upstream risk haplotype. However, there was no statistical difference between the two rs10984103 alleles.

The discovery of the hsCNE-67.7 thyroid enhancer properties is noteworthy given the number of thyroid diseases ascribed to the FOXE1 locus. It contains rs7850258, that has the most significant association in GWAS studies of hypothyroidism [P = 2.5 × 10⁻¹⁹ (20); P = 3.96 × 10⁻⁹ (21)] and papillary thyroid cancer (P = 1.7 × 10⁻⁵) (21,29). We show here that rs7850258 alters enhancer function in both oral epithelial and thyroid cell lines. Of interest is that the G allele is associated with increased risk for CLP and hypothyroidism, while the A allele is associated with thyroid cancer. Relative to the pathogenesis of CLP, that the G allele has more enhancer activity is consistent the gain of function mutation characterized in one patient with Bamforth-Lazarus syndrome (55). Additionally, over expression of Foxe1 in mice also leads to cleft palate (56). One possible mechanism proposed for the association with hypothyroidism is that ectopic FOXE1 induces apoptosis in the developing thyroid, leading to agenesis or dysgenesis (55).

The decreased enhancer activity of the A (cancer associated) allele of rs7850258 also implies that FOXE1 is a tumor suppressor. Evidence for this exists in that the RET/PTC3 proto-oncogene caused by chromosomal rearrangements leading to thyroid papillary carcinoma is known to decrease FOXE1 expression (57). Furthermore, somatic loss of the 9q22 region containing FOXE1 is frequent in squamous cell carcinoma (58). Moreover, hypermethylation of the FOXE1 promoter and/or decreased FOXE1 expression have been observed in a number of cancers including squamous cell carcinoma, (59), pancreatic cancer (60), adenoid cystic carcinoma of salivary gland (61), breast cancer (62), colorectal cancer (63) and anaplastic carcinomas (64). Additionally, another risk allele at rs965513 for papillary thyroid carcinoma is associated with decreased FOXE1 expression in thyroid tissue (65). In total, these reports are also consistent with FOXE1 being pro-differentiation in the developing thyroid and differentiation maintenance in the adult thyroid (44,65).

rs7850258 is predicted to disrupt ARNT and MYC transcription factor binding and consistent with this our data shows a significant response to either factor when the binding site is present in contrast to no response when absent. rs7850258 affects an E-box binding site, so it is possible that other E-box binding transcription factors may be involved as well. That being said there is evidence for both ARNT and MYC being involved in orofacial clefting and thyroid diseases.

ARNT and ARNT2 are basic helix-loop-helix transcription factors that are expressed in the palatal shelf epithelium (66–68). Furthermore, ARNT2 is expressed in the developing thyroid (66). ARNT and ARNT2 regulate gene expression through dimerization with the aryl hydrocarbon nuclear receptor (ARH). AHR responds to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), a well-known teratogen that causes cleft palate in mice (69) and is associated with clefting in humans through agricultural chemical exposures (70). TCDD is also found in tobacco smoke that through maternal gestational exposure attributes to 11% of orofacial clefts (71). Moreover, a genetic variation at the ARNT locus is associated with CL/P in humans (72). While there is convincing data in animal models demonstrating TCDD exposure interferes with fetal and newborn thyroid function, epidemiological studies with equivocal results have not shown a consistent correlation in humans (73). There have not been any reports of association between ARNT, ARNT2 or AHR with hypothyroidism or thyroid cancer in any human studies.

c-MYC is also a b-HLH protein that is expressed in both the epithelia and mesenchyme of the facial processes (74,75). Conditional inactivation in cranial neural crest cells results in craniofacial anomalies (76) and deletion of a facial enhancer at the Myc locus results in cleft lip in conjunction with decreased facial Myc expression. Increased MYC expression is associated with undifferentiated thyroid cancer and the loss of thyroid differentiation markers, PAX8 and TTF1 (77). So in theory, the rs7850528 cancer associated allele A would be less responsive to MYC, resulting in relative decrease of FOXE1 expression compared with the non-risk cancer allele G. This model is consistent with FOXE1 being a tumor suppressor.

A final caveat in this discussion is that until there is association data based on comprehensive sequencing across the region in individuals with CLP and thyroid diseases, it is not possible to state whether rs7850258 is the main contributing factor or if there are additional functional variants on a shared haplotype that in total comprise the genetic risk for these disorders at this locus.

In summary, using zebrafish and mouse transgenesis, we have identified craniofacial and thyroid enhancers at the FOXE1, including one that contains functional risk alleles for both CLP and thyroid diseases. These results reveal the pathogenesis of these disorders associated with the FOXE1 locus. We anticipate that there are additional genetic and biological mechanisms at this locus yet to be discovered that contribute to risk for these disorders. Given that thyroid diseases and orofacial clefting converge on the same SNP within a thyroid and facial enhancer, there may a shared risk for both diseases within families or a population.

Materials and Methods

Zebrafish husbandry

Parental fish were housed at room temperature overnight prior to breeding. Embryos were raised at 28.5°C. The ethical use of animals for research was approved by the University of Iowa The Institutional Animal Care and Use Committee.
In situ hybridization

A 870 bp Foxe1 probe was cloned using primers (forward—
ATGCCCTGGTCTTAAAGTGAGACT (chr1: 2566896—2566999; 
Zv9/danRen7) and reverse—GCACCTGCAGATTGGCCCA (chr1: 
25668130—25668149; Zv9/danRen7)) to amplify from a zebrafish 
cDNA library. The probe was sequenced and by BlastN search of the 
Zebrafish genome, 100% identity over the entire 870 bp probe 
was observed for only foxe1 (refXM_600973.3) at chr1: 25668 130— 
2566899 in Zv9/danRen7. The next level of BlastN identity was 
79% with Foxe3 for only 328 base pairs. Based on these results 
and those published by Nakada et al. (41), there is only one zebrafish 
foxe1 gene and that this gene has been given the gene ID 
number ENSDARG00000079266 in Ensembl. Given the stringency 
conditions used, it is highly unlikely that hybridization of the full 
length probe would occur with anything other than the endogen- 
ous foxe1 mRNA.

Selection of potential regulatory elements

To be conservative relative to boundaries established by our CL/P 
association data, we have targeted a 152 kb interval between 
hg19 Chr9: 100514919—100666931 (rs10818023 to rs3780420) 
(Fig. 1). We chose 15 conserved non-coding human sequences 
(hsCNE) using publicly available sequence data (UCSC) and the 
Phastcons algorithm (79) with a mammalian Phastcons score of 
>400 as inclusion criteria (Supplementary Material, Table S1). 
Fourteen of our chosen elements met these criteria. One addi-
tional element, hsCNE-2.9, was added since it contained a region 
overlapping with the rat Fox1 promoter (80). For all 15 regions 
there was no identity within the Zebrafish genome for more than 
30 base pairs.

Zebrafish enhancer screen

Potential regulatory elements were PCR amplified from human 
BAC clones RP11-10012 or RP11-151120, cloned into the Gateway® 
(Invitrogen) pENTR/D-TOPO vector and transferred to the 
pt2cfosGW eGFP reporter construct that is constructed on the 
Gateway® system (39). The pt2cfosGW eGFP reporter is designed 
to clone the human test DNA upstream of the mouse minimal 
cfos promoter that is unable to independently initiate mRNA 
transcription. Also, built into the pT2cfosGW vector are Tol2 re-
combination sites bracketing the entire construct that facilitate 
transcription. Also, built into the pT2cfosGW vector are Tol2 re-
combination sites bracketing the entire construct that facilitate 
integration of the expression construct into the zebrafish genome 
thus reducing cell mosaicism within the injected embryos. All 
PCR products and plasmid constructs were sequenced and the 
sequence results were aligned onto the human genome using 
the UCSC tool BLAT (81) to ensure fidelity of these steps. Tol2 
mRNA was transcribed from the plasmid pCS-TP (82).

Constructs were injected along with the tol2 mRNA into 100– 
200 zebrafish embryos at the 1 cell stage. The developing embryos 
(generation zero; G0) were screened at 24, 48, 72 and 96 hpf for 
eGFP expression patterns consistent with endogenous foxe1 ex-
pression. A consistent pattern of expression in a minimum of 
10% of injected fish was the criteria for tissue specific enhancer 
activity (83). The use of recombinant DNA was approved by Uni-
versity of Iowa Institutional Biosafety Committee.

Mouse enhancer assay

The methods for evaluating conserved non-coding sequences in 
transgenic mice have been previously described as part of the 
Vista Enhancer project (84). Briefly, target regions were amplified 
from human genomic DNA, cloned into an HSP68-LacZ reporter 
vector, and sequence validated. The linearized vector was in-
jected into the pronucleus of FVB oocytes to generate transgenic 
embryos. LacZ activity was assessed at E11.5 for craniofacial 
staining. The results for Vista Enhancer hs1595 have not been 
previously reported in detail.

Luciferase reporter constructs, transfections 
and luciferase assays

To assess differential allelic activity for the rs10984103 CL/P risk 
SNPs within hsCNE+22.6, the alleles (A/C) two were introduced 
by PCR mediated mutagenesis (85) into the hsCNE+22.6 enhancer 
and subsequently subcloned into pTol2-cFos-Fluc (firefly lucifer-
ase vector with a c-fos minimal promoter) using LR reaction ac-
cording to the manufacturer’s protocol (Life Technologies). In a 
similar fashion different combinations of the alleles for 
rs7850258 (A/G) and rs7864322 (C/T) were introduced into 
hsCNE-67.7, keeping the same allele (A) for the other SNP, 
r1006125 that also resides in this enhancer. Sequencing was 
used to ensure clone integrity.Transient transfections were 
performed using X-tremeGENE HP (Roche, Germany) into GMSM-K 
or FRTL cells with every firefly luciferase vector. For each con-
struct, three independent transfections were performed with Re-
nulla luciferase (pTol2-cFos-RLuc) co-transfection, as an internal 
control for transfection efficiency. The Dual-Luciferase Reporter 
Assay System (Promega, Madison, USA) and a luminometer 
were used to measure luciferase activity in cell lysates. All quan-
tified results are presented as mean ± SEM. Three luciferase mea-
surements were made on each of three or four independent 
biological replicates. A two-tailed unpaired Student’s t-test was 
used to determine statistical significance.

Predictions of differential allelic transcription factor binding 
for rs7850258 were performed using JASPAR (43). Ten base pairs 
surrounding rs7850258 were evaluated and relative 
score of >0.85 was used as a threshold for significance.

The human ARNT expression vector is from GeneCopoeia 
(Rockville, MD, USA) plasmid #EX-C0312-M68 and the human 
MYC expression vector is from Addgene (Cambridge, MA, USA) 
plasmid #18773: MSCV MYC TS8A puro (86).

Cell culture

GMSM-K a human fetal oral epithelial cell line (a kind gift from Dr 
Daniel Grenier (87) was maintained in keratinocyte serum-free 
medium (Life Technologies) supplemented with EGF 1–53 and 
bovine pituitary extract (Life Technologies) in a 5% CO2 atmosphere 
at 37°C. GMSM-K cells constitutively express FOXE1. FRTL a rat thy-
roid cell line was purchased from ATCC (CRL-1468; 42) and cultured 
in Ham’s F-12K (Kighn’s, Life Technologies) supplemented with 
10 nM TSH (Sigma, St Louis, USA), 0.01 mg/ml insulin (Sigma), 
10 nM hydrocortisone (Sigma), 0.005 mg/ml transferrin (Sigma), 
10 mg/ml somatostatin (Sigma), 10 mg/ml glycyrl-histidyl-l-lysine 
acetate (Sigma), and 0.5% FBS. In order to induce the expression 
of Foxe1, the concentration of TSH was increased to 30 mU/ml.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Dr Shannon Fisher for sharing the pT2cfosGW eGFP 
reporter construct. Also, Dr Koichi Kawakami of Division of
Molecular and Developmental Biology, National Institute of Genetics and Department of Genetics, Graduate University of Advanced Studies, Mishima, Shizuoka, Japan generously provided the Tol2 plasmid pCS-TP. Dr Daniel Grenier at the Universit´e Laval, Qu´ebec graciously provided the GMSM-K human fetal oral epithelial cell line. Youssef Kousa aided with Figure construction. We greatly appreciate Drs Rick Domann and Adam Dupuy for sharing the ARNT and MYC expression plasmids.

Conflict of Interest statement. None declared.

Funding
This publication was supported by the National Institutes of Health (RO1 DE014667), the University of Iowa Department of Orthodontics and the National Center for Advancing Translational Sciences through Grant 2 UL1 TR000442-06 (A.C.L.). RO1 Orthodontics and the National Center for Advancing Translation-Health (RO1 DE014667), the University of Iowa Department of

References
23. Bonora, E., Rizzato, C., Diiguiovanni, C., Oudot-Mellakh, T., Campa, D., Vargiolu, M., Guedj, M., McKay, J.D., Romeo, G., Canzian, F. et al. (2014) The FOXE1 locus is a major genetic


30. Gudmundsdson, J., Sulem, P., Gudbjartsson, D.F., Jonasson, J. G., Sigurdsson, A., Berthogthorn, J.T., He, H., Blondal, T., Gel-


32. Kohler, A., Chen, B., Gemignani, F., Elisei, R., Romei, C., Figli-


37. Porcu, E., Medici, M., Pistas, G., Volpato, C.B., Wilson, S.G., Cap-


45. Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Ros-


the outgrowth of unpaired skeletal appendages in zebrafish. BMC Dev. Biol., 7, 75.
82. Suster, M., Kitkata, H., Urasaki, A., Asakawa, K. and Kawakami, K. (2009) Transgenesis in zebrafish with the Tol2...


