Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate

Huan Liu1,4,†, Elizabeth J. Leslie5,†, Zhonglin Jia2,6,†, Tiffany Smith1, Mekonen Eshete7, Azeez Butali3, Martine Dunnwald2, Jeffrey Murray2 and Robert A. Cornell1,*

1Department of Anatomy and Cell Biology, College of Medicine, 2Department of Pediatrics, College of Medicine and, 3Department of Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa City, IA, USA, 4State 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, 5Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA, 6State Key Laboratory of Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, Sichuan, 610041, China and 7Department of Burns and Plastic Surgery, Addis Ababa University, Addis Ababa, Ethiopia

*To whom correspondence should be addressed at: 1-400D Bowen Science Building, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA.
Tel: +1 3193358908; Fax: +1 3193357198; Email: robert-cornell@uiowa.edu

Abstract

Non-syndromic (NS) cleft lip with or without cleft palate (CL/P) is a common disorder with a strong genetic underpinning. Genome-wide association studies have detected common variants associated with this disorder, but a large portion of the genetic risk for NSCL/P is conferred by unidentified rare sequence variants. Mutations in IRF6 (Interferon Regulatory Factor 6) and GRHL3 (Grainyhead-like 3) cause Van der Woude syndrome, which includes CL/P. Both genes encode members of a regulatory network governing periderm differentiation in model organisms. Here, we report that Krüppel-like factor 17 (Klf17), like Grhl3, acts downstream of Irf6 in this network in zebrafish periderm. Although Klf17 expression is absent from mammalian oral epithelium, a close homologue, Klf4, is expressed in this tissue and is required for the differentiation of epidermis. Chromosome configuration capture and reporter assays indicated that IRF6 directly regulates an oral-epithelium enhancer of KLF4. To test whether rare missense variants of KLF4 contribute risk for NSCL/P, we sequenced KLF4 in approximately 1000 NSCL/P cases and 300 controls. By one statistical test, missense variants of KLF4 as a group were enriched in cases versus controls. Moreover, two patient-derived KLF4 variants disrupted periderm differentiation upon forced expression in zebrafish embryos,

†These authors contributed equally.
Introduction

Non-syndromic cleft lip with or without cleft palate (NSCL/P) affects about one in 700 newborns in the USA and is the second most common structural birth defect (after heart anomalies) (1). The pathogenesis of NSCL/P involves both environmental and genetic risk factors, as is true for other common diseases (2). Genome-wide association studies (GWASs) (3–7), linkage studies (8) and a meta-analysis of these (9) have identified 18 loci in which specific common alleles are associated with NSCL/P. Despite the success of GWASs, a substantial portion of the heritability for NSCL/P remains unexplained. Additional common variants conferring risk for NSCL/P may have escaped detection by GWASs, for instance, because of a low density of genotyped SNPs, as in the vicinity of FOXE1 (8). By analogy to other complex diseases, another portion of missing heritability is likely derived from rare or de novo sequence variants (10). Candidate genes to carry these variants are those pointed to by the GWASs, i.e. the genes whose expression level or function is altered by common variants associated with NSCL/P (11). However, the results of sequencing 13 such genes/loci in over 1000 patient-parent trios were not consistent with coding variants of these genes, as a group, being over-transmitted to patients; it nonetheless remains possible that specific function-altering variants confer risk for NSCL/P (12). In principle, genes encoding members of the same gene regulatory networks (GRNs) as GWAS-indicated genes are also candidates to harbour risk variants (13). In this context, the GRN regulating differentiation of oral periderm is of special interest.

The periderm is the most superficial layer of embryonic epidermis and oral epithelium. Evidence that periderm development is relevant to pathogenesis of cleft palate came from animal models of Van der Woude syndrome (VWS), which includes CL/P and, usually, lip pits. This syndrome is caused most commonly by mutations in IRF6, which likely inhibits Klf17, Klf2a and Klf2b, recapitulate the phenotype of embryos injected with dnIRF6. Switching to mammalian oral periderm, in which Klf17 is not expressed, we shift focus to a member of the same structural subgroup, Klf4, that is expressed in this tissue (25). Klf4 is essential for differentiation of epidermis in mice (29), and we find human KLF4 functions like zebrafish Klf17 in over-expression assays. We show that IRF6 binds an enhancer of Klf4 that is active in oral epithelium cells, supporting the possibility that KLF4 is directly downstream of IRF6 in a GRN regulating oral periderm differentiation. Finally, we find evidence that rare variants in KLF4 confer risk for NSCL/P.

Results

We previously found that klf17 and klf2b expression is strongly downregulated in embryos injected with dnIRF6 (20). Embryos depleted of klf17 exhibit lower-than-normal expression of simple epithelium keratin genes (e.g. krt4, krt5 and krt17), but they live until at least 24 h post fertilization (hpf) (28). The fact that such embryos do not die during gastrulation similar to dnIRF6-injected embryos may reflect the partially redundant activity of Klf2a and Klf2b in promoting periderm differentiation; expression of these genes is upregulated in klf17 knockdown embryos (28). To determine whether activity of the Klf2a/2b/17 network is essential for the expression of markers of periderm differentiation (20). Mutations in GRHL3 were identified in eight of 45 families with VWS lacking IRF6 mutations, and missense variants of GRHL3 segregating with VWS mimicked the ability of dominant-negative GRHL3 to disrupt zebrafish periderm differentiation (23). These findings highlight the importance of periderm integrity for correct morphogenesis of the face and suggest that mutations in genes encoding additional members of the periderm GRN may influence risk for syndromic and/or non-syndromic forms of CL/P.

In addition to Grhl3 and Irf6, the zebrafish periderm GRN includes Krüppel-like factor 17 (Klf17). The 17 members of KLF superfamily of DNA-binding transcriptional regulators share the structural feature of a C-terminal DNA binding domain (DBD) with three zinc fingers; depending on the family member, the cellular context and the presence of co-factors, they regulate proliferation, differentiation, migration and pluripotency in embryonic development and in cancer (24,25). Among the KLF proteins, KLF2, KLF4 and KLF17 constitute a more-closely related subgroup (25,26). In zebrafish, klf17 [previously annotated as klf4b (27)] is expressed at high levels in the periderm from 30% epiboly onwards (27,28), whereas klf2a and klf2b are expressed in a mosaic, salt-and-pepper pattern in the periderm at this stage; klf4 is not expressed at gastrula stages (27,28). Kotkamp et al. (28) showed that Klf17, Klf2a and Klf2b function in a regulatory network to promote zebrafish periderm differentiation, acting downstream of both Pou5f1 and an additional transcription factor speculated to be Irf6. They found that many keratin genes are direct transcriptional targets of Klf17 and that promoters of these genes were enriched in motifs matching the preferred binding site of the Klf2/4/17 subgroup (28).

Here we show that, as anticipated, Klf17 acts downstream of Irf6 in the GRN governing zebrafish periderm and that embryos injected with RNA encoding dominant-negative Klf17, which likely inhibits Klf17, Klf2a and Klf2b, recapitulate the phenotype of embryos injected with dnIRF6.Switching to mammalian oral periderm, in which Klf17 is not expressed, we shift focus to a member of the same structural subgroup, Klf4, that is expressed in this tissue (25). Klf4 is essential for differentiation of epidermis in mice (29), and we find human KLF4 functions like zebrafish Klf17 in over-expression assays. We show that IRF6 binds an enhancer of KLF4 that is active in oral epithelium cells, supporting the possibility that KLF4 is directly downstream of IRF6 in a GRN regulating oral periderm differentiation. Finally, we find evidence that rare variants in KLF4 confer risk for NSCL/P.
periderm differentiation and viability during gastrulation, we engineered a cDNA encoding only the KIF17 nuclear localization sequence and its DBD. The promoters of deduced direct targets of all three proteins are enriched for the same binding motif (28), so we predicted that KIF17-DBD would compete with endogenous Klf17, Klf2a and Klf2b for chromatin binding and thereby act in a dominant-negative manner (dnKlf17) (Fig. 1A). We also engineered a truncated Klf17 without the DBD (Klf17-ΔDBD), anticipating that this would be non-functional (Fig. 1A). For each cDNA construct, we synthesized mRNA in vitro and injected it into wild-type zebrafish embryos at the one-cell stage. Most embryos injected with a control mRNA (IacZ2) developed normally, the majority of embryos injected with dnklf17 mRNA ruptured at the animal pole starting at ~6 hpf (Fig. 1B), recapitulating the phenotype of embryos injected with dnirf6 mRNA (19). Control RNA-injected embryos had characteristic expression of krt4 in the periderm (Fig. 1C), characteristic polygonal shape and pronounced phalloidin staining, revealing polymerized F-actin, at the cell perimeter (Fig. 1G). In contrast, dnklf17-injected embryos exhibited widespread reduction of krt4 expression (Fig. 1D). In such embryos, periderm cells had weaker-than-normal phalloidin staining and were less regularly shaped (Fig. 1H), again recapitulating the phenotype of dnirf6-injected embryos (19). Embryos injected with klf17-ΔDBD developed normally and exhibited normal krt4 (Fig. 1E) and phalloidin stain (Fig. 1I); those injected with intact klf17 will be discussed subsequently. By quantitative reverse transcriptase–polymerase chain reaction (RT–PCR), we determined that dnklf17-injected embryos also exhibited reduced expression of several periderm markers (i.e. krt4, cldn, ghrl3 and ovol1) in comparison to control RNA-injected embryos (Fig. 1K), whereas the levels of irf6 expression were similar in dnklf17-injected and control mRNA-injected embryos (Fig. 1K). These findings suggest that the Klf2a/2b/17 network is required for periderm differentiation and that it acts downstream of Irf6.

It was previously reported that over-expression of klf17 will induce elevated expression of many markers of the periderm, but the spatial quality of the ectopic expression was not shown (28). To test whether exogenous Klf17 expression is sufficient to induce deep cells to adopt characteristics of periderm cells, we injected embryos with klf17 mRNA. Such embryos exhibited ectopic expression of krt4 in deep cells at 6 hpf (Fig. 1F inset, arrowheads); upregulation of other periderm markers was detected by immunoprecipitation with anti-IRF6 and anti-H3K27Ac, an indicator of active chromatin (29), and transgenic mice with epithelium-specific over-expression of Klf4 exhibit a cleft palate (37). We therefore pursued the possibility that KLF4 functions downstream of IRF6 in a mammalian oral epithelium cell line. Supporting this possibility, knockdown of IRF6 expression in primary human keratinocyte reduced KLF4 expression (38). In wild-type mouse embryos at E14.5, we detected anti-KLF4 immunoreactivity in the outer surface of the palate and tongue (i.e. oral periderm), whereas in the irf6-null mutant mice such immunoreactivity was undetectable (Fig. 2A). We also found that over-expression of full-length human IRF6 in a human oral epithelium cell line (i.e. GMSM-K) led to the upregulation of KLF4 mRNA (Fig. 2B). These results place IRF6 upstream of KLF4 expression in mammalian oral epithelium.

To test the possibility that IRF6 directly regulates KLF4 expression, we examined published catalogues of peaks of chromatin immunoprecipitation with anti-IRF6 and anti-H3K27Ac, an indicator of active chromatin, detected in primary normal human epidermal keratinocytes (NHEKs) (38,39). Whereas there were no overlapping peaks near the KLF4 promoter, there were five present near 1 Mb downstream from KLF4, with no intervening genes (Fig. 2C). The sequences of these elements are evolutionarily conserved among mammals (www.genome.ucsc.edu, date last accessed, August 30, 2015), supporting the notion that they are enhancers; hereafter, we refer to them as KLF4 candidate enhancers 1–5 (KLF4-E1 to -E5). To test whether they are indeed enhancers of KLF4 expression (i.e. in the dorsal and ventral periderm), this phenotype was partially rescued in embryos that were co-injected with klf17 and dnirf6 mRNAs (Fig. 1M and N). These results further support a model whereby KIF17 acts downstream of IRF6 in the GRN that governs periderm differentiation.
expression, we conducted chromatin conformation capture (3C) in a human oral keratinocyte cell line (i.e. GMSM-K), anchored with a primer in the KLF4 promoter, and found that KLF4-E1 and KLF4-E5 interact with the KLF4 promoter (Fig. 2C). Notably, four of the five putative KLF4 enhancers are located within an extended segment (>12 kb) of chromatin marked with the H3K27Ac in NHEKs. This is a feature of so-called ‘super-enhancers’, which are proposed to regulate the expression of key cell-identity genes (40,41) (Fig.2C and Supplementary Material, Fig. S2 and Table S1).

We again performed 3C, with the anchor primer in KLF4-E1, and found that KLF4-E1 interacts with the other four elements in NHEK cells (Supplementary Material, Fig. S2). A published ChIA-PET analysis of MCF7 breast epithelium cells supports the findings of our 3C studies (42).

On the basis of the results from the model systems, we asked whether mutations in KLF4 predispose individuals to cleft lip with or without cleft palate. As mentioned earlier, dominant-negative variants of GRHL3, encoding another member of the periderm GRN, are present in a subset of VWS patients who lack IRF6 mutations (23). We sequenced KLF4 exons in 40 individuals who had been diagnosed with VWS and lacked mutations in either IRF6 or GRHL3. However, we detected no
non-synonymous mutations, indicating that KLF4 mutations are not a common cause of VWS.

Next, to test hypothesis that common sequence variants near KLF4 elevate risk for NSCL/P, we genotyped common variants near KLF4 exons in 184 multiplex families from the Philippines. We found no association with NSCL/P (the lowest P-value was $P = 0.228$ for rs149109998, Supplementary Material, Table S4). Similarly, we did not detect association of common variants near the distal enhancer with NSCL/P (E.L., unpublished data). This argues that common variants near KLF4, at least within the linkage disequilibrium units that we analysed, are unlikely to contribute risk for NSCL/P.

Finally, we identified KLF4 missense variants and tested for their association with NSCL/P using several statistical methods. We sequenced KLF4 exons in 1034 patients with NSCL/P from the Philippines, USA and Ethiopia and 338 controls from the Philippines (not all exons were sequenced in every sample, for details see Supplementary Material, Table S2). We detected one common coding variant, previously submitted to dbSNP (rs45520942/ss70352771, see Materials and Methods), in two US cases. This variant, a 51 bp deletion resulting in the loss of 17 amino acids from the repression domain of KLF4, is present at a frequency of 2.6% in Western European samples (i.e. CEPH) catalogued at HapMap. Thus, it is unlikely to have a strong pathogenic
effect. We also detected five novel missense variants: three in the Filipinos, p.S284L, p.H427Y (in two different families) and p.R94G, all predicted to be ‘probably damaging’ by Polyphen and ‘deleterious’ by SIFT; and two in the American cases, p.R309P and p.V261M, both predicted to be probably damaging by Polyphen but ‘tolerated’ by SIFT (Supplementary Material, Table S2). In three families, we sequenced all affected and unaffected individuals and found that the respective KLF4 variant was present in all affected individuals (Fig. 3B). In each family, the missense variant was inherited from an unaffected parent, as opposed to being de novo. Next we employed two burden tests: the cohort allelic sums test (CAST) (43) and the variable threshold (VT) method (44). These tests revealed an apparent trend towards association that did not reach formal significance (i.e. CAST, \( P = 0.058 \) and VT, \( P = 0.071 \)). Of note, burden tests assume that all variants in the analysis have the same effect (i.e. all increase risk of disease) and have limited power when there are variants with opposite effects. We also applied two statistical analyses that do not make this assumption: the replication-based test (RBT) (45) and the C-alpha test (46). One but not the other revealed association that reached nominal significance (i.e. RBT, \( P = 0.026 \) and C-alpha, \( P = 0.362 \)). These statistical analyses constitute mild support for the possibility that missense KLF4 variants contribute to NSCL/P pathogenesis.

Next, to test for potential dominant-negative activity of patient-derived KLF4 variants, we synthesized KLF4 mRNA that matched the reference sequence, the common missense variant (i.e. 51 bp deletion) or the variants predicted to be probably damaging by Polyphen and deleterious by SIFT (i.e. R94G, S284L and H427Y). As expected, >90% of the embryos injected with the reference allele KLF4 survived through 24 hpf, and none ruptured during gastrulation. Similarly, the frequency of survival for embryos injected with the 51 bp deletion and R94G variants was comparable to that for those injected with KLF4. In contrast, >30% of the embryos injected with the S284L variant and >70% of those injected with H427Y exhibited widespread reduction of krt4 expression in the periderm at 5 hpf and ruptured at ∼6 hpf (Fig. 3C and D). These observations support the possibility that S284L and H427Y variants are pathogenic for NSCL/P. The fact that they were inherited from unaffected parents suggests that at normal expression levels they are not as severely damaging as when over-expressed.

**Discussion**

To identify candidates for genes that when mutated cause CL/P, we first employed the zebrafish gastrula-stage periderm as a model of human oral periderm, then pursued the results from zebrafish with functional tests in an oral epithelium cell line and finally tested a promising candidate gene by sequencing its exons in patients with NSCL/P. Previously, we found that in comparison to control embryos, embryos injected with dnIrf6 exhibit reduced expression of several transcription-factor-encoding genes, including klfl7 (20), in addition to genes encoding periderm differentiation effectors such as keratin genes. Here, we found that forcing expression of klfl7 could partially reverse the effects of dnIrf6, restoring expression of keratin genes. A promoter analysis suggested that this gene is directly regulated by

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Rare NSCL/P alleles of KLF4 disrupt development of the zebrafish periderm. (A) Schematic representation of KLF4 protein. Positions of the rare, patient-derived variants predicted to be damaging by Polyphen and SIFT are indicated above, and that of the common variant is indicated below. (B) Pedigrees of families with KLF4 p.S284L and p.H427Y variants. (C–I) Dorsal views of embryos, injected with the indicated mRNA and fixed at 5 hpf and processed to reveal (C–G) krt4 expression or (H and I) phalloidin stain. Arrows indicate periderm lacking phalloidin stain. (J) Quantification of embryos that died around 6 hpf when injected with the indicated mRNA. The numbers in parentheses indicate the total number of embryos for each group.
Irf6. We found no published evidence that Klf17 is expressed in mammalian periderm or epidermis. In contrast, KLF4 is a close homologue of KLF17 and is known to be necessary for epidermal differentiation (29). Referring to published IRF6-ChIP-SEQ (47) data, we found that IRF6 binds several distal elements. We established that these elements interact with the KLF4 promoter and that they have epidermal enhancer activity that depends on the presence of the IRF6 binding sites. Sequencing within all exons for KLF4 in more than 1000 CL/P probands revealed five rare coding variants presenting in different CL/P pedigrees, and with zebrafish periderm as in vivo model, two of them were functionally identified exhibiting dominant-negative effect. Together with earlier work, this study reveals that mutations in genes encoding three elements of the oral periderm GRN—i.e. IRF6, GRHL3 and KLF4—can all predispose individuals to cleft lip with or without cleft palate.

GWASs (3–7), linkage studies (8) and a meta-analysis of these (9) have identified at least 18 loci in which specific common alleles are associated with NSCL/P at genome-wide significance. The high ratio of the number of risk-associated loci identified versus the number of samples sequenced makes NSCL/P one of the most successful applications of GWASs; together, the loci account for more than half of the overall population-attributable risk for this disorder (48). The remaining ‘missing heritability’ may reside in rare variants of large effect size, among other possibilities (e.g. gene interactions, structural variation and misdiagnosis (49)). Supporting this view, two rare KLF4 variants we identified in NSCL/P cases have the potential to have a large effect size, based on the in silico analysis and on zebrafish-based functional studies. The fact that they were inherited from unaffected parents indicates they are not sufficient to cause disease on their own, consistent with the diagnosis of NSCL/P in the patients. Additional rare variants pathogenic for NSCL/P may be found in other candidate loci with strong biological support either through direct sequencing (as we have done here) or through exome and whole-genome sequencing efforts.

This work indicates that fish Klf17 and mammalian Klf4 share the function of promoting differentiation of superficial epithelia. Phylogenetic analysis shows that klf17 groups with klf2 and klf4 (25,26). In zebrafish, klf2a, klf2b and klf17 function in a network to promote periderm differentiation (28). Our work extends the earlier study to show that excess klf17 has the potential to repress expression of some periderm markers on the ventral side, but not to disrupt cell adhesion in the periderm as is induced by dnklf17, perhaps by repressing klf2b expression (28). Zebrafish klf4 is not expressed in gastrula-stage embryos (28), but embryos depleted of klf4 exhibit abnormal differentiation of intestinal epithelium (50), as do murine Klf4 mutants (51-53). Indeed, in mouse, Klf4 has an essential role in the differentiation of many epithelia, including epidermis, by stimulating expression of Cdkn1a (25). We found that a human KLF4 enhancer drives a periderm expression pattern that resembles zebrafish klf17, not zebrafish klf4, and that human KLF4 functioned equivalently to zebrafish klf17 in an overexpression assay. A parsimonious evolutionary scenario consistent with these data is that an ancestral gene, encoding a KLF orthologue that promoted periderm differentiation, was duplicated, giving rise to KLF2, KLF4 and KLF17. Differential loss of regulatory elements or changes in protein function (both forms of subfunctionalization) (54) in different taxa resulted in retention of periderm differentiation function in distinct paralogues in distinct taxa. Supporting this model, in the mammalian but not in the fish lineage, KLF17 has evolved rapidly, consistent with protein subfunctionalization (25,26). It has been proposed that Klf family members are especially prone to evolvability by virtue of their quality of being co-regulated by the same upstream regulators (28).

Exploring the zebrafish periderm GRN has led us to two genes that appear to contribute risk for orofacial clefting (i.e. GRHL3 and KLF4); we anticipate that other members of this GRN will contain the rare variants that contribute to the missing heritability of NSCL/P. It is noteworthy that among the 18 genes presumed to be associated with NSCL/P in genome-wide studies, at least six are expressed in the oral epithelium (i.e. ARHGAP29, FOXE1, IRF6, MAFB, NOG1, and SPRY2). VAX1, also identified in such studies, is expressed in the brain but is essential for normal patterning of oral epithelium (55). Zebrafish embryonic periderm is a tractable model for deducing the architecture of the regulatory networks containing these proteins, and for testing the function of patient-derived variants. Combined with ChIP and analyses of chromatin interactions in cell lines or primary periderm cells, it should be possible to assemble the periderm GRN, including the direct regulatory interactions, at high resolution.

**Materials and Methods**

**Human samples and study approval**

Patients with NSCL/P were recruited from sites in the Philippines, Ethiopia and the USA. Individuals with additional congenital malformations or developmental delays suggestive of a syndrome were excluded. Patients in the US cohort with self-reported race/ethnicity as Asian, Hispanic or Black were also excluded. Controls with no known history of craniofacial anomalies were recruited from the Philippines. Informed consent was obtained for all participants, and the study was approved by the Institutional Review Board at the University of Iowa (approval numbers 199804080, 199804081 and 201102785).

**Animal models**

Breeding Danio rerio were maintained as described (56) in the University of Iowa Animal Care Facility. Zebrafish embryos were staged according to Kimmel et al. (57) at 28.5°C by hours or days post fertilization (hpf or dph).

The Irf6 mutant allele (irf6<sup>−/−</sup>), which was maintained in a C57BL/6 background, has been described previously (58). C57BL/6 mice were used as wild-type control. Murine embryos were obtained from pregnant female mice at 14.5 days after conception (E14.5). Animal use protocols were approved by the Institutional Animal Care and Use Committees at the University of Iowa.

**Zebrafish KLF17 and human KLF4 mRNA injection and in situ hybridization**

Full-length, wild-type zebrafish klf17 cDNA (RefSeq: NM_131723.1) was cloned from zebrafish 6 hpf cDNA and inserted into pcS2+ plasmid. Different functional domains of Klf17 were determined by comparing its amino acid sequence with human KLF4 (UniProtKB: O43474). Different truncated forms of zebrafish KLF17 were generated by overlapping PCR. Sanger sequencing confirmed that no variants were present in the clones. Full-length, wild-type and p.S284L mutant human KLF4 cDNAs (RefSeq: NM_004235.4) were obtained from Stratagene company. Other mutations from NSCL/P patients were generated via PCR-mediated mutagenesis. All the KLF4 variants were engineered into pcS2+ plasmid.

Capped mRNA was synthesized in vitro (mMESSAGE mMACHINE SP6 kit, Ambion) and purified with RNasy<sup>®</sup> mini kit (Qiagen, Valencia, CA, USA), and ~1 ng of mRNA was injected...
into wild-type zebrafish embryos at one-cell stage. Whole-mount in situ hybridization for krt4 was performed as described before (19).

**RNA isolation and quantitative RT–PCR**

Total RNA was isolated using RNeasy mini kit (Qiagen) and treated with RNease-free DNase I (Promega, Madison, WI, USA). RT was performed by High Capacity cDNA Reverse Transcription Kits (Applied Biosystems, Foster City, CA, USA). Real-time PCR was performed in CFX96 Touch™ Real-Time PCR Detection System. For zebrafish and human mRNA expression assays, beta-2-microglobulin and glyceraldehyde 3-phosphate dehydrogenase were used as an internal control, respectively.

**Cell culture, luciferase reporter constructs, transfections and dual luciferase assay**

The 293FT cell line (human embryonic kidney cell line) was purchased from Invitrogen (Carlsbad, CA, USA) and was maintained in Dulbecco’s modified Eagle’s medium (Life Technologies, Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (Life Technologies). The 293FT cell line was used as an internal control, respectively. For zebrafish using Lipofectamine® 3000 (Life Technologies), according to manufacturer’s instructions. Transient transfection in GMSM-K cells was performed with pCS2+ as described previously (19). Confocal images of periderm cells in Phalloidin-stained embryos were collected with a 20×/0.5 NA objective lens (Carl Zeiss).

Immunodetection for mouse Klf4 was performed on coronal sections of 4% formaldehyde-fixed, paraffin-embedded E14.5 mouse heads. After blocking with 3% goat serum (Vector Laboratories, Burlingame, CA, USA), sections were incubated with rabbit anti-human KLF4 (Abcam, Cambridge, MA, USA, ab72543), washed in PBS and incubated in goat anti-rabbit Alexa-Fluor 488. DAPI (LifeTechnologies) was used as a nuclear stain. Images were viewed with a Nikon Eclipse E800 (Melville, NY, USA) and acquired with a SPOT RT Slider Diagnostic Instruments CCD camera using Spot Advanced software (Diagnostic Instruments, Sterling Heights, MI, USA). Black and white images were pseudocolourized and merged.

**Chromatin conformation capture (3C)**

The 3C-qPCR assay was performed as described (61) for at least two replicates. The 1×10^6 GMSM-K oral epithelial cells were cross-linked with 1% formaldehyde at room temperature for 10 min, followed by glycine quenching and cell lysis. Nuclei were resuspended in 500 µl 1× EcoRI Buffer and then lysed with 0.2% sodium dodecyl sulphate (SDS), followed by SDS sequestration with 1.2% Triton X-100. Lysates were digested overnight at 37°C with EcoRI (NEB, Ipswich, MA, USA) and the restriction enzyme then inactivated for 1 h at 65°C with 1.6% SDS. Ligation was performed with NEB T4 ligase in 1.15× T4 Ligase buffer at 16°C for 4 h. Finally, the ligation products were purified with phenol-extraction protocol. 3C ligation products were quantified with three replicates by SYBR Green qPCR. The 3C signals at the each locus were normalized first to a loading control (primers located in the GAPDH gene and that do not flank any EcoRI digestion site), and second to a random collision control, a region 8.1kb from the KLF4 TSS, chosen because it lacks chromatin marks in NHEK. Primers were designed using Primer3 Plus online software (http://www. bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and tested against EcoRI-digested, ligated BAC templates (RP11-150F11 and RP11-589C8) by serial dilution and melt-curve analysis to ensure specific and linear amplification. Digestion efficiencies were determined with SYBR Green qPCR by comparison of aliquots taken pre- and post-digestion using primer pairs that amplified a region in each interrogated fragment spanning an EcoRI digestion site.

**Genomic interval manipulation**

Identification of super-enhancers of NHEK cells was performed according to the algorithm described before (41). Briefly, constituent enhancers in NHEKs were defined as ChIP-seq peaks of H3K27ac in NHEK (39), with exclusion of 3 kb in the vicinity of TSS for each gene. H3K27ac ChIP-seq raw reads were sorted and indexed using SAMtools (v. 0.1.18) (62). About 1470 super-enhancers in NHEKs were defined. Intersections between NHEK super-enhancers and IRF6 ChIP-seq peaks (38) were carried out in Bedtools (v. 2.24.0) (63), and the positive intersection flanking KLF4 was visualized in UCSC Genome Browser.

**Zebrafish enhancer in vivo reporter assay**

Human KLF4 potential enhancers KLF4-E1 to KLF4-E5 were cloned as described earlier and engineered into the zebrafish enhancer

---

**Immunofluorescent staining**

Zebrafish embryos were fixed at dome stage in 4% paraformaldehyde/phosphate-buffered saline (PBS) at 4°C overnight. For F-actin staining, 0.165 µm Phalloidin Alexa 546 (University of Iowa Central Microscopy Research Facility, Iowa City, IA, USA) was used as described previously (19). Confocal images of periderm cells in Phalloidin-stained embryos were collected with a 20×/0.5 NA objective lens (Carl Zeiss).

Immunodetection for mouse Klf4 was performed on coronal sections of 4% formaldehyde-fixed, paraffin-embedded E14.5 mouse heads. After blocking with 3% goat serum (Vector Laboratories, Burlingame, CA, USA), sections were incubated with rabbit anti-human KLF4 (Abcam, Cambridge, MA, USA, ab72543), washed in PBS and incubated in goat anti-rabbit Alexa-Fluor 488. DAPI (LifeTechnologies) was used as a nuclear stain. Images were viewed with a Nikon Eclipse E800 (Melville, NY, USA) and acquired with a SPOT RT Slider Diagnostic Instruments CCD camera using Spot Advanced software (Diagnostic Instruments, Sterling Heights, MI, USA). Black and white images were pseudocolourized and merged.
detection (ZED) vector (64). ZED vectors (25–30 ng/μl) were injected along with the tol2 mRNA (20–30 ng/μl) into 200 one-cell stage embryos. The embryos were screened at 6 and 72 hpf for enhancer activity in periderm and oral epithelium, respectively. A consistent GFP expression pattern in a minimum of 10% of injected fish (i.e. F0) was the criterion for tissue-specific enhancer activity, as this usually predicts a given element’s expression pattern in F1 animals. Embryos with periderm enhancer activity were photographed in bright field, epi-fluorescent illumination on a Leica DMRA2 compound microscope with a colour 12 bit ‘QIClick’ camera (QImaging). Embryos with oral epithelium enhancer activity were photographed in a Zeiss 700 confocal microscope with a 20×/0.5 NA objective lens (Carl Zeiss).

Human sample genotyping

Three common SNPs (rs2236599, rs149109998 and rs77275043) were genotyped using Taqman SNP Genotyping Assays (Life Technologies, Grand Island, NY, USA) on the ABI Prism 7900HT and analysed with SDS 2.4 software (Applied Biosystems). Families with Mendelian errors were excluded from analyses.

Human sample sequencing

Primers were designed with Primer3 (http://biotools.umassmed.edu/biapps/primer3_www.cgi) to cover all exons of KLF4 (Supplementary Material, Table S1). For technical reasons, only the 375 bp of the 5′-UTR upstream of the start codon and the first 500 bp of the 3′-UTR were sequenced. PCR products were Sanger-sequenced using ABI BigDye v3.1 on an ABI 3730×l instrument (Applied Biosystems, Madison, WI, USA). The chromatograms were transferred to a Unix workstation, base-called with PHRED (v. 0.961028), assembled with PHRAP (v. 0.960731), scanned by POLYPHRED (v. 0.970312) and visualized using CONSED program (v. 4.0). We used the Variant Effect Predictor tool from Ensemble (http://www.ensembl.org/tools.html) to predict the functional effects of missense variants (65).

The rs45520942 variant was deleted from the dbSNP database because of mapping or clustering errors. The original dbSNP submission can be found with the submission identifier ss70352771. It has also been found in the exomes of the Exome Aggregation Consortium, although it has been flagged as a lower quality site. This variant appears to be a polymorphism, although the frequency is unclear, which may be due to difficulty mapping this region of the genome.

Genetic statistical analysis

Association with rs2236599, rs149109998 and rs77275043 in the Filipino families was tested using the family-based association test (FBAT) (66). We used the -e option in FBAT, which allows for the adjustment for multiple affected individuals or multiple nuclear families within a pedigree by adjusting the variance used in the Z-score calculation.

From the sequencing data, we considered a variant ‘novel’ if it was absent from dbSNP137, the 1000 Genomes Project and the NHLBI Exome Variant Server. We performed case–control analyses on two groups of rare variants (MAF <1%): (i) all coding variants and (ii) all missense variants. Association with rare variants was tested using several burden and non-burden statistical methods. Burden tests collapse all of the rare variants in a region into a single ‘score’ and test for association between that score and the phenotype. We employed two burden tests: the CAST (43) and the VT method (44). Burden tests assume that all variants in the analysis have the same effect (i.e. all increase risk of disease) and have limited power when there are variants with opposite effects. We also employed two non-burden tests: the C-alpha test (46) and the RBT (45). All rare variant analyses were implemented in R using the Assosotester R software package.

The Exome Aggregation Consortium catalogue (URL: http://exac.broadinstitute.org) was accessed on July 23, 2015.


Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgement

We are grateful to Gregory Bonde for excellent animal husbandry and technical support of this project.

Conflict of Interest statement. None declared.

Funding

This work was supported by the National Institute of Health grants HD073107 (R.A.C.), K99-DE025060 (E.J.L.), K99/R00 DE022378 (A.B.), DE08559 (J.M.), by the Robert Wood Johnson Foundation grant 72429 (A.B.), by the National Natural Science Foundation of China 81400477 (H.L.), and by Scientific Research Funds for Young Teachers of Sichuan University grant 2015SCJU1999 (Z.L.). Funding to pay the Open Access publication charges for this article was provided by a grant from the A.E. and Martha Michelbacher Trust administered by the Marin Community Foundation.

References


