

[文章编号] 1674-8603(2020)03-0151-05

基于人口腔上皮细胞颅颌面特异增强子序列机器学习 预测 *IRF6* 位点唇腭裂致病突变

张涵舒¹, 刘欢^{1,2*}(1. 武汉大学口腔医学院口腔基础医学省部共建国家重点实验室培训基地·口腔生物医学教育部
重点实验室, 湖北 武汉 430079; 2. 武汉大学口腔医院牙周科, 湖北 武汉 430079)

[摘要] 目的: 利用基于人口腔上皮细胞颅颌面特异增强子序列机器学习, 优化唇腭裂相关非编码突变研究靶点的筛选。方法: 采用永生化人口腔上皮细胞(HIOEC) 组蛋白 H3 第 27 位赖氨酸的乙酰化(H3K27Ac) 染色质免疫共沉淀测序(ChIP-seq) 整合人类胚胎期颅颌面超级增强子区域, 获得人口腔上皮细胞颅颌面特异增强子区域。采用 gapped k-mer SVM 算法总结增强子序列特征, 并用该特征判断干扰素调节因子 6(*IRF6*) 附近唇腭裂相关单核苷酸多态性(SNP) 位点或突变对增强子活性的影响。结果: 人口腔上皮细胞颅颌面特异性增强子能够涵盖超过半数唇腭裂相关 SNP 位点($P<0.01$), 而基于该序列特征的机器学习预测与范德伍德综合征相关的 350dupA 位点显著降低所在增强子活性。双荧光素酶报告基因显示 350dupA 显著降低增强子在人口腔上皮细胞内的活性($P<0.01$)。结论: 基于人口腔上皮细胞颅颌面特异增强子序列机器学习能够精准判断 *IRF6* 附近唇腭裂相关非编码突变, 优化唇腭裂遗传功能研究。

[关键词] 唇腭裂; 功能基因组; 机器学习; *IRF6***[中图分类号]** R782.2 **[文献标识码]** A **[doi]** 10.3969/j.issn.1674-8603.2020.03.004

Machine learning based human oral epithelium active enhancers predicts a pathogenic variant associated with cleft lip and palate near *IRF6*

ZHANG Hanshu¹, LIU Huan^{1,2}. (1. The State Key Laboratory Breeding Base of Basic Science of Stomatology and Key Laboratory of Oral Biomedicine Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; 2. Department of Periodontics, Hospital of Stomatology, Wuhan University, Wuhan 430079, China)

Corresponding author: LIU Huan, Email: liu.huan@whu.edu.cn

[Abstract] **Objective:** This study aimed to validate whether machine learning based on craniofacial-specific oral epithelium enhancers could be used to non-coding variants associated with cleft lip and palate. **Methods:** Anti H3K27Ac ChIP-seq was performed in human immortalized oral epithelial cells (HIOEC), which were integrated published craniofacial super-enhancers. The sequences of these enhancers were used as training set for gapped k-mer SVM (gkm SVM) machine learning, which could summarize all the DNA sequence features. Delta SVM scores were employed to validate the effects of non-coding variants near *IRF6*, and dual luciferase assays were used for biological validation. **Results:** Published cleft lip and palate-associated SNPs were more enriched in oral epithelium-specific enhancers. gkm SVM showed 350dupA mutation could significantly decrease enhancer activity, which was then validated using dual luciferase assays in HIOEC cells. **Conclusions:** The classifiers based on oral epithelium-specific enhancers are useful in nominating functional SNPs identified in cleft lip and palate genome wide association studies.

[Key words] cleft lip and palate; functional genome; machine learning; *IRF6*

唇腭裂是口腔最常见的由于口腔上皮或间充质发育缺陷造成的先天性颅颌面缺损。我国唇腭裂发病率约为 1.8 : 1000, 近年来呈逐渐增长趋势。根据是否伴发其他全身畸形, 唇腭裂可分为非综合征型唇腭裂(non-syndromic cleft lip or palate, NSCL/P) 和综合征型唇腭裂(syndromic cleft lip or palate, SCL/P), 临床上以 NSCL/P 最为常见。通常临床所述唇

腭裂主要指 NSCL/P, 其发病一方面与环境因素有关, 也与患者的遗传因素相关, 全基因组关联分析(genome-wide association studies, GWAS) 配合重测序是发现唇腭裂相关致病基因的有效手段^[1]。至今, 针对不同国家人群的 GWAS^[2-11]、重测序^[12]、Meta 分析^[6, 13-14] 和遗传关联分析^[15] 显示: 超过 40 个易感基因区域(locus) 与不同人群唇腭裂的发病有统计学相关性。但是, 统计学相关性并不代表这些位点的致病性。因此, 从生物功能上验证这些“相关”的致病突变, 是推动遗传学研究向临床诊断转化的不

基金项目: 国家自然科学基金(81771057)

* 通信作者: 刘欢, Email: liu.huan@whu.edu.cn

可或缺的环节。

根据突变在基因组中的位置,唇腭裂突变致病机制主要分为两类:(1)少量位于编码蛋白质基因外显子序列中的突变:通过影响氨基酸序列从而导致唇腭发育相关基因功能的紊乱,如转录因子 *GRHL3* (grainyhead-like 3) 编码区功能缺失突变^[10,16-17];(2)非编码区 DNA 突变:绝大多数 NSCL/P 相关突变位点均位于非编码区,但由于非编码区 DNA 序列与相应功能缺乏绝对的对对应关系,因此这些突变位点的功能解读十分困难。我们曾整合 DNA 元件百科全书计划(ENCODE)^[18]以及表观组学线路图计划(Roadmap Epigenomics Project)^[8]的结果,利用体内外报告基因实验筛选、染色体构象实验、染色质免疫共沉淀(chromatin immunoprecipitation, ChIP)以及 CRISPR-Cas9 技术,针对 GWAS 所发现的位于 1p22 区域的 10 个唇腭裂相关的单核苷酸多态性(single nucleotide polymorphisms, SNP)位点进行生物学功能验证,结果显示这 10 个候选位点中,仅有 3 个 SNP 通过影响所在增强子对 Rho 三磷酸鸟苷酶激活蛋白 29(Rho GTPase activating protein 29, *ARHGAP29*) 基因转录调控,影响口腔上皮生物学功能进而提高唇腭裂发病率^[19]。该流程虽然能够系统的应用于几乎所有唇腭裂相关的非编码区 DNA 突变的研究,但是第一步利用荧光素酶报告基因对 SNP 的筛选需要耗费大量的时间。

机器学习是克服该难点的可行路线之一,而在多种机器学习方法中,单纯以 DNA 序列为输入的 gapped k-mer SVM(gkm SVM)方法被证明是精确率和准确率较高的方法之一,能够较为稳定的归纳多种组织特异性增强子序列特征,并利用该特征进行遗传病相关位点致病碱基的预测^[20]。以往研究中,我们采用该方法总结了斑马鱼周皮细胞特异增强子 DNA 序列特征,利用 delta SVM 评估了细胞角蛋白 18 编码基因 *KRT18* 附近唇腭裂相关 SNP 位点对增强子活性的影响,并预测出 rs2070875 的致病碱基相对更强的唇腭裂致病性^[21]。但由于物种差异,人来源的口腔上皮细胞与斑马鱼周皮细胞存在一定的差异,人口腔上皮特异性增强子 DNA 序列特征能否更加精准的预测唇腭裂相关突变的致病性仍然未知。

本研究将通过机器学习手段,分析和比较人及斑马鱼来源口腔上皮增强子特征与唇腭裂易感区域的相关性。同时,由于干扰素调节因子 6(interferon regulatory factor 6, *IRF6*) 基因特异性表达于口腔上皮结构,报道显示该基因附近的非编码区存在一定

数量的 SNP 位点和突变,且与 NSCL/P 和 SCL/P 相关^[2-12]。因此,本研究也利用该机器手段预测和验证 *IRF6* 附近相关的唇腭裂相关 SNP 位点和突变的生物学致病性。

1 材料与方法

1.1 主要试剂

含重组表皮生长因子(epidermal growth factor, EGF)和牛脑垂体提取物(bovine pituitary extract, BPE)的无血清角质形成细胞培养基(Keratinocyte-SFM)培养基、0.25%胰酶(含 EDTA)、青霉素和链霉素(Life Technologies, 美国)、蛋白酶抑制剂(Roche, 德国)、双荧光素酶载体 PGL3-Promoter、双荧光素酶报告基因检测试剂盒以及 GloMax[®] 20/20 发光检测仪(Promega, 美国)、组蛋白 H3 第 27 位赖氨酸的乙酰化(histone H3 acetylated on lysine 27, H3K27Ac)抗体(ab4729, Abcam, 英国)、DNA 测序建库试剂盒(ND606-01, 南京诺唯赞生物科技公司)、ChIP 试剂盒(17-371, Millipore, 德国)、SG 4D-Nucleofector[™] X solution 电转试剂盒和 Amaxa[™] 4D-Nucleofector[™] 电转仪(Lonza, 德国)。

1.2 方法

1.2.1 细胞培养 永生化人口腔上皮细胞(human immortalized oral epithelial cells, HIOEC) (原上海第二医科大学附属第九人民医院口腔颌面外科肿瘤生物学实验室馈赠,保存于武汉大学口腔医学院口腔生物医学教育部重点实验室)^[22]培养于含有重组 EGF 和 BPE 的无血清 Keratinocyte-SFM 培养基,并加入 100 U/mL 青霉素和 100 μ g/mL 链霉素,放入 37 $^{\circ}$ C、5% CO₂ 细胞培养箱中培养。在培养基中加入 1.2 mmol/L 的 CaCl₂促进上皮细胞终末分化,培养 3 d 用于后续染色质免疫共沉淀测序(ChIP-Seq)和电转实验。

1.2.2 HIOEC H3K27Ac 的 ChIP-seq 建库、测序及后续分析 HIOEC 培养于直径 100 mm 的培养皿中(每皿作为一个生物学重复),并使用 1.2 mmol/L Ca²⁺诱导培养。诱导前后各 3 个皿的细胞使用 1% 甲醛室温固定 10 min 后,134 mol/L 甘氨酸处理 5 min,预冷 PBS 冲洗后刮出收集每皿细胞。细胞裂解后,冰浴条件下使用超声打断基因组 DNA。根据试剂盒说明采用 H3K27Ac 抗体进行 ChIP 实验,所获得的 DNA 片段采用 DNA 测序建库试剂盒进行建库,交付北京安诺优达医学检验实验室有限公司完成 Illumina 二代测序。测序结果 fastq 数据根据前期已经发表并存储于托管平台 Github 中的相关流

程进行测序后分析(https://github.com/Badgerliu/periderm_ATACSeq),包括测序结果的清理、修剪、比对、ChIP-seq 峰值注释和可视化处理。将最终获得的 HIOEC H3K27Ac ChIP-seq 的峰值进行后续分析。

1.2.3 获得人口腔上皮细胞颌颌面特异性增强子序列 考虑到超级增强子比单独增强子有更强的组织特异性,我们下载了人胚胎发育期颌颌面特异性超级增强子序列(GSE97752)^[23]。但该超级增强子为面突上皮和间充质等细胞的混合信息。故我们利用 Bedtools (2.24.0) intersect 函数获得与 GSE97752 序列重合的 HIOEC 特异增强子,即人口腔上皮细胞颌颌面特异性增强子序列。将已发表的正常人表皮角质形成细胞(normal human epidermal keratinocytes , NHEK)的 H3K27Ac ChIP-seq 结果作为皮肤上皮特异增强子(GSE26320)^[18] ,并获得其与 GSE97752 序列重合区域,作为“其他上皮对照组”。

判断所有与唇腭裂发病相关的 SNP 是否与某种组织细胞特异增强子活性相关,需要检查这些 SNP 在组织细胞特异增强子区域内的富集程度。如果唇腭裂相关 SNP 更多的富集于某种组织/细胞特异增强子区域内,则证明这种组织/细胞相关的增强子序列更适合进行唇腭裂相关 SNP 的功能判断。为了比较两种上皮细胞颌颌面特异性增强子区域与唇腭裂 SNP 的相关性,我们采用 GREGOR 算法^[24] 量化比较了两者与前期发表的唇腭裂 GWAS 发现的 SNP^[11] 和突变的相关性。

1.2.4 基于不同颌颌面特异性增强子序列的机器学习 为提取不同上皮细胞颌颌面特异性增强子序列的特征,我们采用已报道的 gkm SVM 算法($K=6$, $L=10$,其余均采用默认参数)获得不同上皮特异增强子序列对应的训练模型,并采用 delta SVM 预测 *IRF6* 附近已报道的 NSCL/P 和范德伍德综合征(Van der Woude syndrome ,VWS) 相关 SNP 位点不同等位基因对增强子活性的影响。相关计算使用 R 语言(3.6.2)完成,相关函数存储于 GitHub(https://github.com/Badgerliu/EZ_gkmSVM_functions)。

1.2.5 细胞转染及双荧光素酶报告基因实验 为了验证机器学习结果对于 *IRF6* 附近 350dupA 对其所在 DNA 序列增强子活性的影响,本研究合成了该区域(chr1: 209989080-209989980 (hg19))野生型(wild type ,WT) 和已筛选的带有 350dupA 突变的该片段,合成由上海生工完成并插入至 pGL3-promoter 载体中,分别命名 pGL3-promoter-WT 和 pGL3-promoter-350dupA。该载体分别与 phRL-TK 载体以 10 : 1 混合后电转转染入 HIOEC 内。转染后 3 d ,利

用双荧光素酶报告基因检测试剂盒检测增强子活性。电转实验独立重复 3 次,每孔双荧光素酶活性反复读取 3 次。利用 R dplyr 的单因素方差分析对结果进行统计。

2 结果

2.1 人口腔上皮细胞颌颌面特异性增强子和斑马鱼周皮特异性增强子对唇腭裂注释的比较

本研究中,我们采用 GREGOR 算法比较了以往 GWAS 研究在亚洲人群中发现的唇腭裂相关 SNP 位点^[11] 与人口腔上皮细胞颌颌面特异增强子序列、其他上皮对照组以及人基因组内类似斑马鱼周皮特异增强子序列^[21] 内的富集程度。作为对照,我们随机选取基因组内 1000 个 SNP 位点,进行同样的富集。统计显示,与对照组相比,唇腭裂 SNP 位点与三种细胞特异性增强子相关(富集程度 >1 , $P<0.05$,图 1) ,并且在人口腔上皮细胞颌颌面特异性增强子序列中富集程度更高。

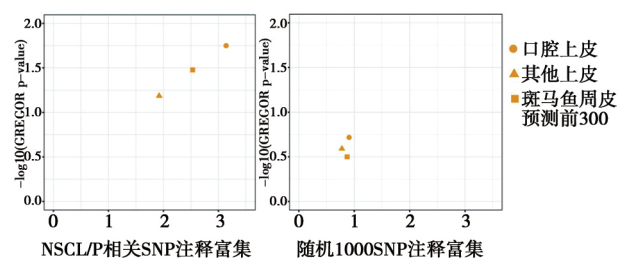


图 1 SNP 与不同上皮特异性增强子区域富集关系
Fig.1 Enrichment of SNPs in the enhancers active in different types of epithelium

2.2 基于不同颌颌面特异性增强子序列机器学习模式的建立

为了实现对唇腭裂相关 SNP 影响上皮增强子活性的预测,我们采用了以往研究显示精确度较高的 gkm SVM 算法,归纳不同颌颌面特异上皮增强子序列对应的训练模型。该训练模式的获得将用于下文对于 SNP 不同等位基因对增强子活性影响的预测。结果显示,基于人口腔上皮细胞颌颌面特异性增强子序列的机器学习能够较为精准的归纳其 DNA 序列特征($AU_{ROC}=0.73$, $AU_{P-RC}=0.71$,图 2)。

2.3 人口腔上皮颌颌面特异性增强子序列机器学习模式优化 *IRF6* 位点附近功能性致病突变预测

为了进一步判断机器学习能否优化唇腭裂功能基因组研究,我们选择以往报道位于 *IRF6* 基因附近的 NSCL/P 相关 SNP 和引起 VWS(更严重的唇腭裂表型) 的非编码区 DNA 突变^[25] 进行机器学习评估(表 1) ,通过 delta SVM 分数来判断 SNP/ 突变不同等位基因对增强子活性的影响,人类基因组内随机

1000 个 SNP 位点的不同等位基因作为对照组。通过比对,我们发现绝大多数 *IRF6* 附近的唇腭裂相关 SNP/突变位点不与上皮特异性增强子区域重合(图 3A),而引起 VWS 的 350dupA 位点位于一个口腔上皮细胞特异增强子内(图 3B)。且在所有选取的 SNP 中,350dupA 位点被预测影响口腔上皮细胞增强子活性的作用最强(图 3C)。双荧光素酶报告基因实验验证发现,350dupA 对所在增强子的活性降低有统计学意义($P<0.01$,图 3D)。

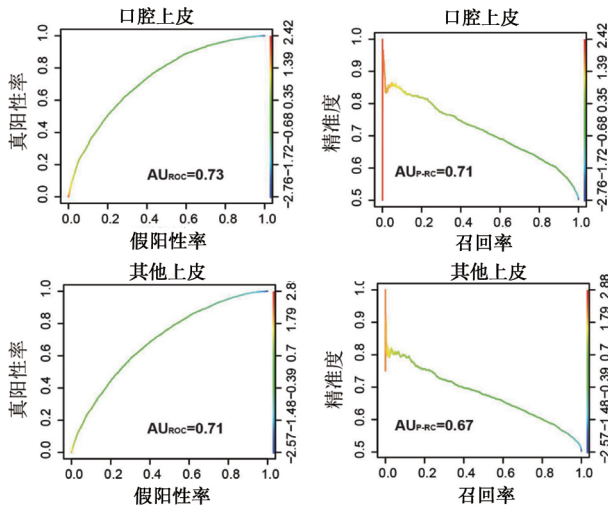


图2 gkm SVM 基于人口腔上皮细胞颅颌面特异性增强子 DNA 序列机器学习效果评估

Fig.2 Analysis of gkm SVM classifier trained with craniofacial-specific oral epithelium enhancers

表1 *IRF6* 附近唇腭裂相关 SNP/突变位点

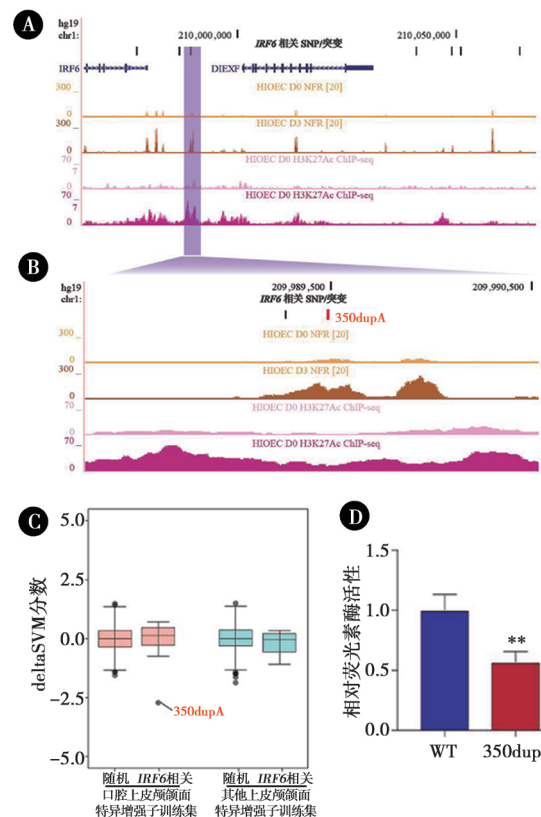
Tab.1 SNPs or mutation associated with cleft lip and palate phenotypes near *IRF6*

染色体	起点位置	SNP/突变位点
chr1	209977111	rs861020
chr1	209986672	rs680923
chr1	209989270	rs642961
chr1	210019824	rs126280
chr1	210040672	rs7548533
chr1	210048819	rs2064163
chr1	210050794	rs9430019
chr1	210064188	rs10863794
chr1	209989269	rs642961
chr1	209989480	350dupA

3 讨论

先天性唇腭裂是口腔最常见的先天发育畸形,发病与遗传背景有关。虽然 GWAS 和全基因组测序等手段已经揭示出 40 多个与唇腭裂发病有关的易感基因位点,但是如何从功能上证明这些具有“统计学相关性”的基因位点的致病机制一直是本领域的难点,也阻碍了遗传学研究向临床精准诊断

的转化。针对这一难点,我们已建立了一系列体内外系统,用于唇腭裂相关非编码区突变的功能研究,并精准判断“相关”突变中真正具有生物学功能的突变位点^[19]。但是该研究系统需要花费大量时间和费用。因此,如果能够事先知道非编码区功能片段 DNA 序列特征,就能够事先判断这些位点不同等位基因能否产生功能上的影响,从而优化研究目标选择,减少时间和成本。我们已经利用机器学习模型手段构建了斑马鱼周皮的特异性增强子序列,该模型能够非常精准的预测唇腭裂易感区域中不同致病位点对所在增强子活性的影响,替代非编码区功能研究平台第一步中对所有突变位点的比对和双荧光素酶报告基因的筛选过程^[21]。



A、B: 分别显示 NSCL/P、VWS 相关 SNP/突变位点附近的增强子区域; C: gkm SVM 评估基因组内随机 1000 个或者 *IRF6* 相关 SNP/突变不同等位基因对增强子活性影响; D: HIOEC 细胞内 350dupA 的双荧光素酶报告基因验证实验; **: 与 WT 相比 $P<0.01$

图3 gkm SVM 预测 *IRF6* 附近的 350dupA 显著降低口腔上皮特异性增强子活性

Fig.3 gkm SVM based on oral epithelium-specific enhancers sequences predict 350dupA decrease enhancer activity

本研究中我们进一步比较了基于斑马鱼周皮、人正常皮肤上皮以及口腔上皮细胞特异性增强子机器学习在唇腭裂非编码区 DNA 功能研究的效果。结果显示,基于人口腔上皮细胞特异性增强子机器

学习所获得的训练模式能更精准的判断唇腭裂相关 SNP 的致病位点。VWS^[25] 作为唇腭裂缺陷更严重的表型,与其相关的非编码突变相比 NSCL/P 相关 SNP 位点,能够更加显著的降低所在增强子活性,进而引起唇腭裂发生。

除了增强子以外,基因组非编码区 DNA 中包含了不同种类的功能元件。与唇腭裂相关的非编码区突变不仅仅影响了增强子序列,还有可能造成了绝缘子、抑制子甚至是“增强子-启动子”拓扑结构的改变。这些元件的功能基础和突变的致病基础均为 DNA 序列,在无法从生物学精准总结这些元件 DNA 序列特征之前,深度学习手段可以较好地“绕过”这一知识缺陷,推动遗传精准诊断的发展。

[参 考 文 献]

- [1] Huang L , Jia Z , Shi Y , *et al.* Genetic factors define CPO and CLO subtypes of nonsyndromic orofacial cleft [J/OL]. PLoS Genet 2019 ,15(10) : e1008357 [2020-07-19]. <https://doi.org/10.1371/journal.pgen.1008357>.
- [2] Beaty TH , Murray JC , Marazita ML , *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4 [J]. Nat Genet 2010 ,42(6) : 525-529.
- [3] Birnbaum S , Ludwig KU , Reutter H , *et al.* Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24[J]. Nat Genet 2009 ,41(4) : 473-477.
- [4] Camargo M , Rivera D , Moreno L , *et al.* GWAS reveals new recessive loci associated with non-syndromic facial clefting [J]. Eur J Med Genet 2012 ,55(10) : 510-514.
- [5] Grant SFA , Wang K , Zhang H , *et al.* A Genome-Wide Association Study Identifies a Locus for Nonsyndromic Cleft Lip with or without Cleft Palate on 8q24[J]. J Pediatr 2009 ,155(6) : 909-913.
- [6] Leslie EJ , Carlson JC , Shaffer JR , *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts , and TP63 and cleft lip with or without cleft palate [J/OL]. Hum Genet 2017 ,136(3) : 275-286 [2020-07-19]. <https://doi.org/10.1007/s00439-016-1754-7>.
- [7] Mangold E , Ludwig KU , Birnbaum S , *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate [J]. Nat Genet 2010 ,42(1) : 24-26.
- [8] Kundaje A , Meuleman W , Ernst J , *et al.* Integrative analysis of 111 reference human epigenomes [J]. Nature 2015 ,518(7539) : 317-330.
- [9] Wolf ZT , Brand HA , Shaffer JR , *et al.* Genome-wide association studies in dogs and humans identify ADAMTS20 as a risk variant for cleft lip and palate [J]. PLoS Genet 2015 ,11(3) : e1005059.
- [10] Leslie EJ , Liu H , Carlson JC , *et al.* A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3 [J/OL]. Am J Hum Genet 2016 ,98(4) : 744-754 [2020-07-19]. <https://doi.org/10.1016/j.ajhg.2016.02.014>.
- [11] Yu Y , Zuo X , He M , *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity [J]. Nat Commun 2017 ,8: 14364.
- [12] Leslie EJ , Taub MA , Liu H , *et al.* Identification of functional variants for cleft lip with or without cleft palate in or near PAX7 ,FGFR2 and NOG by targeted sequencing of GWAS loci [J]. Am J Hum Genet 2015 ,96(3) : 397-411.
- [13] Beaty TH , Taub MA , Scott AF , *et al.* Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study [J]. Hum Genet 2013 ,132(7) : 771-781.
- [14] Ludwig KU , Mangold E , Herms S , *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci [J]. Nat Genet 2012 ,44(9) : 968-971.
- [15] Moreno LM , Mansilla MA , Bullard SA , *et al.* FOXE1 association with both isolated cleft lip with or without cleft palate and isolated cleft palate [J]. Hum Mol Genet 2009 ,18(24) : 4879-4896.
- [16] Eshete MA , Liu H , Li M , *et al.* Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate [J/OL]. J Dent Res 2018 ,97(1) : 41-48 [2020-07-19]. <https://doi.org/10.1177/0022034517729819>.
- [17] Mangold E , Bohmer AC , Ishorst N , *et al.* Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate [J/OL]. Am J Hum Genet 2016 ,98(4) : 755-762 [2020-07-19]. <https://doi.org/10.1016/j.ajhg.2016.02.013>.
- [18] Dunham I , Kundaje A , Aldred S , *et al.* An integrated encyclopedia of DNA elements in the human genome [J]. Nature 2012 ,489(7414) : 57-74.
- [19] Liu H , Leslie EJ , Carlson JC , *et al.* Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting [J/OL]. Nat Commun 2017 ,8: 14759 [2020-07-19]. <https://doi.org/10.1038/ncomms14759>.
- [20] Lee D , Gorkin DU , Baker M , *et al.* A method to predict the impact of regulatory variants from DNA sequence [J]. Nat Genet 2015 ,47(8) : 955-961.
- [21] Liu H , Duncan K , Helverson A , *et al.* Analysis of zebrafish periderm enhancers facilitates identification of a regulatory variant near human KRT818 [J/OL]. eLife 2020 ,9: 51325 [2020-07-19]. <https://doi.org/10.7554/eLife.51325>.
- [22] Sdek P , Zhang ZY , Cao J , *et al.* Alteration of cell-cycle regulatory proteins in human oral epithelial cells immortalized by HPV16 E6 and E7 [J]. Int J Oral Maxillofac Surg 2006 ,35(7) : 653-657.
- [23] Wilderman A , VanOudenhoove J , Kron J , *et al.* High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development [J]. Cell Rep 2018 ,23(5) : 1581-1597.
- [24] Schmidt EM , Zhang J , Zhou W , *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic data-driven approach [J]. Bioinformatics 2015 ,31(16) : 2601-2606.
- [25] Fakhouri WD , Rahimov F , Attanasio C , *et al.* An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects [J]. Hum Mol Genet 2014 ,23(10) : 2711-2720.

(收稿日期: 2020-07-19)

(本文编辑: 顾建雨)