

中国人群非综合征型唇裂伴或不伴腭裂的单核苷酸多态性遗传度

薛恩慈¹, 陈 曦¹, 王雪珩¹, 王斯悦¹, 王梦莹¹, 李 劲¹, 秦雪英¹, 武轶群¹, 李 楠², 李 静³, 周治波², 朱洪平², 吴 涛^{1△}, 陈大方¹, 胡永华¹

(1. 北京大学公共卫生学院流行病学与卫生统计学系, 北京 100191; 2. 北京大学口腔医学院口腔颌面外科, 北京 100081; 3. 北京大学口腔医学院儿童口腔科, 北京 100081)

[摘要] **目的:** 探索非综合征型唇裂伴或不伴腭裂 (non-syndromic cleft lip with or without cleft palate, NSCL/P) 全基因组常见遗传变异对 NSCL/P 风险的影响。 **方法:** 利用全基因组关联研究 (genome-wide association study, GWAS) 数据, 以全基因组单核苷酸多态性 (single nucleotide polymorphism, SNP) 遗传度和基因组不同分区 SNP 遗传度评估基因组上常见变异的效应。对 GWAS 汇总数据进行质量控制, 标准包括数据中无缺失值、弱势等位基因频率 $\geq 1\%$ 、 P 值在 0~1、SNP 正负链明确等。利用连锁不平衡得分回归计算 NSCL/P 的 SNP 遗传度, 采用分层的连锁不平衡得分回归计算基因组编码区、启动子区、内含子区、增强子区和超级增强子区的分区 SNP 遗传度, 并评估不同分区内的富集度, 分析工具为 LDSC (v1.0.1) 软件。 **结果:** 纳入中国人群 806 个 NSCL/P 核心家系 (2 418 人) 的 GWAS 数据, 490 593 个 SNP 通过质量控制, 被纳入到 SNP 遗传度的计算中。观测样本中 NSCL/P 的 SNP 遗传度为 0.55 (95% CI: 0.28~0.82), 由于观测样本患病率较高, 按中国人群患病率转换为一般人群后 SNP 遗传度为 0.37 (95% CI: 0.19~0.55)。SNP 遗传度在增强子区的富集度为 15.70 ($P=0.04$), 在超级增强子区的富集度为 3.18 ($P=0.03$)。 **结论:** 基因组常见变异有助于解释一部分中国人群 NSCL/P 目前未被解释的遗传度, 同时中国人群 NSCL/P 的 SNP 遗传度在增强子分区和超级增强子分区中显著富集, 提示该区域中可能存在未被发现的遗传致病因素。

[关键词] 非综合征型唇裂伴或不伴腭裂; 单核苷酸多态性遗传度; 核心家系

[中图分类号] R394.1 **[文献标志码]** A **[文章编号]** 1671-167X(2024)05-0775-06

doi: 10.19723/j.issn.1671-167X.2024.05.004

Single nucleotide polymorphism heritability of non-syndromic cleft lip with or without cleft palate in Chinese population

XUE Enci¹, CHEN Xi¹, WANG Xueheng¹, WANG Siyue¹, WANG Mengying¹, LI Jin¹, QIN Xueying¹, WU Yiqun¹, LI Nan², LI Jing³, ZHOU Zhibo², ZHU Hongping², WU Tao^{1△}, CHEN Dafang¹, HU Yonghua¹

(1. Department of Epidemiology and Biostatistics, Peking University School of Public Health, Beijing 100191, China; 2. Department of Oral and Maxillofacial Surgery, Peking University School of Stomatology, Beijing 100081, China; 3. Department of Pediatric Dentistry, Peking University School of Stomatology, Beijing 100081, China)

ABSTRACT Objective: To delve into the intricate relationship between common genetic variations across the entire genome and the risk of non-syndromic cleft lip with or without cleft palate (NSCL/P).

Methods: Utilizing summary statistics data from genome-wide association studies (GWAS), a thorough investigation to evaluate the impact of common variations on the genome were undertaken. This involved assessing single nucleotide polymorphism (SNP) heritability across the entire genome, as well as within specific genomic regions. To ensure the robustness of our analysis, stringent quality control measures were applied to the GWAS summary statistics data. Criteria for inclusion encompassed the absence of missing values, a minor allele frequency $\geq 1\%$, P -values falling within the range of 0 to 1, and clear SNP strand orientation. SNP meeting these stringent criteria were then meticulously included in our analysis. The SNP heritability of NSCL/P was calculated using linkage disequilibrium score regression. Additionally, hierarchical linkage disequilibrium score regression to partition SNP heritability within coding regions, promoters, introns, enhancers, and super enhancers were employed, and the enrichment levels within different genomic regions using LDSC (v1.0.1) software were further elucidated. **Results:** Our study drew upon GWAS summary statistics data obtained from 806 NSCL/P trios, comprising a total of 2 418 individuals from the Chinese population. Following rigorous quality control procedures, 490 593

out of 492 993 SNP were deemed suitable for inclusion in SNP heritability calculations. The observed SNP heritability of NSCL/P was 0.55 (95% CI: 0.28 - 0.82). Adjusting for the elevated disease prevalence within our sample, the SNP heritability scaled down to 0.37 (95% CI: 0.19 - 0.55) based on the prevalence observed in the general Chinese population. Notably, our enrichment analysis unveiled significant enrichment of SNP heritability within enhancer regions (15.70, $P=0.04$) and super enhancer regions (3.18, $P=0.03$). **Conclusion:** Our study sheds light on the intricate interplay between common genetic variations and the risk of NSCL/P in the Chinese population. By elucidating the SNP heritability landscape across different genomic regions, we contribute valuable insights into the genetic basis of NSCL/P. The significant enrichment of SNP heritability within enhancer and super enhancer regions underscores the potential role of these regulatory elements in shaping the genetic susceptibility to NSCL/P. This paves the way for further research aimed at uncovering novel genetic pathogenic factors underlying NSCL/P pathogenesis.

KEY WORDS Non-syndromic cleft lip with or without cleft palate; Single nucleotide polymorphism heritability; Case-parent trios

非综合征型唇腭裂 (non-syndromic oral clefts, NSOC) 是常见的出生缺陷, 其中非综合征型唇裂或不伴腭裂 (non-syndromic cleft lip with or without cleft palate, NSCL/P) 是最常见的亚型^[1]。中国活产儿唇腭裂患病率约为 1.4/1 000^[2]。遗传度是评价遗传与环境对表型贡献的指标, 它衡量的是遗传因素所能解释表型方差的占比。双生子研究和家系研究均表明遗传因素是 NSCL/P 重要的危险因素, 其遗传度为 70%^[3-4]。NSCL/P 是一种多基因疾病, 尽管国内外多个全基因组关联研究 (genome-wide association study, GWAS) 发现了 NSCL/P 的阳性位点, 但这些潜在致病位点所能解释的遗传度合计仅为 20% 左右^[5], 其遗传病因尚不完全清晰, 仍存在未被解释的遗传度, 即尚有致病性遗传因素有待挖掘。目前学界认为尚未被解释的遗传度可能是因为许多常见变异因为 GWAS 效能不足还没有被发现。如果能够得到全基因组所有常见变异所能解释的遗传度, 将有助于理解常见变异在疾病遗传病因中的作用。同时, 如果发现常见变异在遗传度中的贡献较大, 说明还存在许多相关的常见变异位点未被发现, 可以通过进一步增大 GWAS 效能的方式发现更多致病位点^[6-7]。因此, 探索 NSCL/P 的 GWAS 中所有常见变异所能解释的遗传度将为进一步开展遗传病因研究提供基础。

传统上主要采用基于家系^[8]或双生子设计^[9]对遗传度进行估计。随着生物技术的发展, 在人群中开展的 GWAS 可以获得基因数据。依赖单核苷酸多态性 (single nucleotide polymorphism, SNP) 位点对遗传度贡献大小的一些前提假设, 已经可以利用人群的基因数据计算遗传度, 得到全基因组所有常见变异所能解释的遗传度, 即 SNP 遗传度^[7]。连锁不平衡得分回归是一种计算 SNP 遗传度的方法, 通过对 GWAS 汇总数据进行统计建模, 其不仅适用于许多以前不能用于 SNP 遗传度估计的数据, 而且

大大提高了计算速度^[10-11]。此外, 目前随着生物学技术的飞速发展, 基因组不同分区的功能正在不断被发现^[12-15]。由于生物学功能不同, 不同分区在疾病发生中发挥的作用大小可能不同。可针对基因组不同分区开展 SNP 遗传度研究, 如果发现某些功能分区中遗传度更为富集, 提示该分区中可能存在遗传致病因素, 例如 Won 等^[16]在心肌梗死和冠状动脉心脏病的研究中发现调控区域的 SNP 能解释大部分的 SNP 遗传度, 提示基因表达调控区域可能存在重要的致病因素。量化不同分区中的 SNP 遗传度大小可以作为一种新型的候选区域策略, 为进一步精确定位和机制研究提供新的思路和线索。

全基因组范围内常见变异以及不同功能分区范围内常见变异在 NSCL/P 遗传病因学中的作用尚不清晰, 本研究拟利用 806 个中国人群核心家系的 GWAS 数据, 计算 NSCL/P 的 SNP 遗传度和分区 SNP 遗传度, 探索全基因组常见变异和不同功能分区常见变异所能解释的遗传度。

1 资料与方法

1.1 GWAS 数据

本研究所依据的 GWAS 包括 806 个汉族 NSCL/P 核心家系 (包括患儿及其父母双亲, 共计 2 418 人), 其地区和性别分布见表 1。GWAS 收集了患者及其父母的生物标本并提取 DNA, 使用 Illumina Human610-Quad v. 1_B BeadChip 芯片进行基因型检测, 检测工作于 2009 年在美国约翰霍普金斯大学医学院 McKusick-Nathans 遗传医学研究所遗传病研究中心 (Center for Inherited Disease Research, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, USA) 完成, 详见 Beaty 等^[17]的研究。本研究获得了北京大学生物医学伦理委员会批准 (批准号: IRB00001052-15081)。患儿唇腭裂分型由临床遗传学家和医生

共同确定,确保纳入患儿均为 NSCL/P,纳入家系的父母均签署了书面知情同意书。

表1 806个中国 NSCL/P 核心家系的地区和性别分布

Table 1 Regional and gender distribution of 806 NSCL/P trios in China

Site	Male	Female	Total
China Taiwan, <i>n</i>	139	94	233
Shandong, <i>n</i>	193	81	274
Hubei, <i>n</i>	132	55	187
Sichuan, <i>n</i>	75	37	112
Total, <i>n</i>	539	267	806

NSCL/P, non-syndromic cleft lip with or without cleft palate.

GWAS 采用传递不平衡检验 (transmission disequilibrium test, TDT) 探索基因位点与疾病的关联关系,其结果详见 Beaty 等^[17] 的研究。本研究以 GWAS 汇总统计数据为基础,包含每个 SNP 位点的等位基因频率、物理位置及其与 NSCL/P 关联强度。

1.2 质量控制

对 GWAS 进行了严格的质量控制,条件包括:(1) 缺失率 $\leq 10\%$; (2) 弱势等位基因频率 $\geq 1\%$; (3) 父母人群符合 Hardy-Weinberg 平衡 ($P > 0.000\ 01$); (4) 孟德尔遗传错误率 $\leq 5\%$ 。经质量控制后, GWAS 数据中共纳入 492 993 个 SNP 位点。

本研究在采用 LDSC 软件进行 SNP 遗传度计算之前,对 GWAS 汇总数据进行了二次质量控制,条件包括:(1) 无缺失值;(2) 弱势等位基因频率 $\geq 1\%$;(3) 关联分析 P 值在 0 ~ 1;(4) SNP 位点正负链明确;(5) 数据中 SNP 能够与参考人群 SNP 一一对应。

1.3 统计学分析

1.3.1 SNP 遗传度分析 采用基于 Python 语言开发的 LDSC (v1.0.1) 软件计算 SNP 遗传度^[10]。以 GWAS 汇总数据为基础,并以千人基因组计划亚洲人群(481 人,1 208 050 个 SNP 位点)的连锁不平衡得分为参考^[10],计算 GWAS 中所有 SNP 位点所能解释的 SNP 遗传度。Bulik-Sullivan 等^[10] 经过严格的数学推导证明,可以对连锁不平衡得分进行统计建模:

$$E[\chi_j^2] = \frac{Nh_g^2}{M} l_j + Na + 1 \quad (1)$$

其中,下标 j 表示第 j 个 SNP。 h_g^2 是 SNP 遗传度, M 为染色体固定窗口内的 SNP 总数目, N 为总样本量, a 为人群分层及潜在混杂因素等。因变量 χ_j^2 表示其连锁不平衡得分回归统计量,是第 j 个 SNP 与

表型关联的卡方统计量。对全基因组所有 SNP 位点的期望卡方统计量和连锁不平衡得分进行回归分析,可得到 SNP 遗传度 h_g^2 。截距为 1 (在 1 附近),表示无混杂,偏离 1 较远,表示存在混杂。 l_j 为连锁不平衡得分,是一个重要的自变量。利用 GWAS 汇总数据及单倍型参考数据,对同一条染色体一段区间内(一般为 1 厘摩尔遗传距离)的两两 SNP (第 j 个 SNP _{j} 与第 k 个 SNP _{k}) 相关系数 r 进行计算,该区间内合计得到的 r^2 即为某个 SNP 的连锁不平衡得分 l_j 。 l_j 可表示为:

$$l_j = \sum_{k=1}^M r_{jk}^2 \quad (2)$$

LDSC 软件提供了观测尺度和易患性阈值尺度下的 SNP 遗传度估计结果。由于本研究中 NSCL/P 患病率较高会相对高估 SNP 遗传度,所以为了与一般人群实际情况相符,同时便于与其他研究比较,本研究根据一般人群的患病率将其从观测尺度转换到易患性阈值尺度^[11]。

1.3.2 分区 SNP 遗传度分析 本研究所含分区为基因组编码区、启动子区、增强子区、超级增强子区和内含子区。编码区、启动子区和内含子区的注释来源于 UCSC 基因组数据库^[18],增强子区的注释来源于 Hoffman 等^[19] 的研究,超级增强子注释来源于 Hnisz 等^[20] 的研究。采用分层的连锁不平衡得分回归计算各分区 SNP 遗传度^[21],并评估各分区 SNP 遗传度的富集度。SNP 遗传度的富集度是指某分区所能解释的 SNP 遗传度的比例与其所占 SNP 数目的比例的比。研究使用刀切法评估富集度的统计学意义。

2 结果

2.1 SNP 遗传度分析

经过二次质量控制后,因为基因的正负链不明确排除了 2 400 个 SNP 位点,剩余 490 593 个 SNP 位点被纳入到 SNP 遗传度的计算中。在观测尺度下 NSCL/P 的 SNP 遗传度为 0.55 (95% CI: 0.28 ~ 0.82),在易患性阈值尺度下 NSCL/P 的 SNP 遗传度为 0.37 (95% CI: 0.19 ~ 0.55),连锁不平衡得分回归截距为 1.004 (95% CI: 1.019 ~ 0.989)。

2.2 分区 SNP 遗传度分析

基于 GWAS 汇总数据,本研究对基因组编码区、启动子区、内含子区、增强子区和超级增强子区的分区 SNP 遗传度进行了计算,其结果见表 2。NSCL/P 的 SNP 遗传度在增强子区的富集度为 15.70 ($P = 0.04$),表明增强子区内平均每个 SNP

所能解释的遗传度是其他分区(除增强子区外)的 15.70 倍,在超级增强子区的富集度为 3.18 ($P = 0.03$),表明超级增强子区内平均每个 SNP 所能解

释的遗传度是其他分区(除超级增强子区外)的 3.18 倍。SNP 遗传度在编码区和启动子区虽然也有一定的富集度,但差异无统计学意义。

表 2 NSCL/P 的分区 SNP 遗传度分析

Table 2 Partitioning SNP heritability results of NSCL/P

Annotation	Proportion of SNP	Proportion of SNP heritability	Enrichment	P for enrichment
Coding	0.014	0.035	2.44	0.91
Promoter	0.046	0.144	3.12	0.70
Intron	0.388	0.289	0.74	0.63
Enhancer	0.042	0.654	15.70	0.04
Super-enhancer	0.167	0.532	3.18	0.03

NSCL/P, non-syndromic cleft lip with or without cleft palate; SNP, single nucleotide polymorphism.

3 讨论

本研究发现中国人群 NSCL/P 的 SNP 遗传度为 0.37,同时研究发现 SNP 遗传度在增强子区和超级增强子区显著富集,富集度分别达到了 15.70 和 3.18。利用 GWAS 数据,计算 SNP 遗传度有助于理解常见变异在其病因学中的作用,通过对基因组不同功能分区的 SNP 遗传度计算有助于进一步的精细定位和生物学机制研究。

常见复杂疾病和表型的 SNP 遗传度与传统家系研究发现的遗传度仍有一定差距,这部分仍然无法用所有常见变异所解释的遗传度,也被称为“仍然消失的遗传度(still-missing heritability)”^[22]。Ludwig 等^[23]利用 Bonn 队列 1 717 名研究对象的 GWAS 数据,计算出 NSCL/P 全基因组常见变异所能解释的遗传度为 0.32(95% CI: 0.24 ~ 0.40),与本研究结果较为接近,而家系研究和双生子研究发现 NSCL/P 的遗传度约为 70%^[3-4],参照本研究所得到的 NSCL/P 的 SNP 遗传度为 37%,提示除常见遗传变异外,尚有致病性遗传因素未被发现。“仍然消失的遗传度”存在以下几种可能的解释:(1)存在效应较大的罕见变异,不能通过 GWAS 识别;(2)存在交互作用,包括基因-基因交互作用、基因-环境交互作用等;(3)通过家系研究所得到的遗传度可能被高估了;(4)表观遗传学等因素会影响疾病表型。未来有必要对这些可能的遗传致病因素进行深入探索。

另外,对于大多数复杂表型来说,GWAS 的阳性位点合计解释的遗传度仍低于 SNP 遗传度。对于 NSCL/P,现有阳性结果可解释约 20% 的遗传度^[5],而本研究的结果提示 SNP 遗传度为 37%,即尚有未

被发现的潜在致病性常见遗传变异。由于 GWAS 往往会检测数十万甚至上百万个位点,为了避免出现假阳性结果,设置的统计学显著性阈值非常严苛。既往身高和精神分裂症等众多疾病和表型的 SNP 遗传度研究提示,GWAS 发现的阳性位点合计解释的遗传度低于 SNP 遗传度,主要是因为某些致病性常见变异由于效应较小而被严格的显著性阈值过滤掉^[24-28]。基因组范围内可能还存在更多与 NSCL/P 有关的常见变异未被发现,可通过增加样本量和减少多重比较等增加研究效能的方法来识别这类常见变异。

本研究通过计算分区 SNP 遗传度,发现 NSCL/P 的遗传度在增强子区和超级增强子区较为富集。既往 NSCL/P 的 GWAS 阳性结果多位于非编码区域^[24-28]。*IRF6* 基因被众多候选基因研究和 GWAS 证实与 NSCL/P 有关^[29-31],Rahimov 等^[32]通过多序列比对识别到 *IRF6* 基因一个增强子的常见变异(rs642961),通过关联分析发现其变异与 NSCL/P 有关,且风险等位基因数量与 NSCL/P 发生风险存在剂量反应关系,在功能验证中发现该风险等位基因能够破坏转录因子 AP-2 α 的结合位点,且发现其小鼠的颅面结构中表达,这表明增强子这一调控原件在 NSCL/P 的遗传病因中可能起到重要作用。2020 年, Morris 等^[33]研究了增强子变异是否与 NSCL/P 有关,通过对 20 个增强子进行目标区域测序并开展关联分析,发现位于 *MSX1* 基因和 *SPRY* 基因附近的增强子与 NSCL/P 存在显著关联,并进一步通过功能预测发现这些增强子可能会破坏重要的转录因子结合位点。2013 年 Whyte 等^[34]在 *Cell* 期刊上提出了超级增强子的概念,指出超级增强子是具有转录活性增强子的集合,其可以驱动控制细

胞身份基因的表达。Hnisz 等^[20]对 1 675 个 GWAS 中鉴定的 5 303 个阳性 SNP 分析发现,93% 的阳性 SNP 位于非编码区域,其中有 64% 的位点在增强子区域富集,并且富集于超级增强子区域的变异显著多于增强子区域。目前超级增强子的研究主要集中于肿瘤领域^[35],通过开展关联分析推断出与肿瘤相关的超级增强子,从而为进一步确定肿瘤易感基因提供依据。VISTA Enhancer Browser^[36]和 EnhancerAtlas^[37]等大型数据库的开发也为开展增强子和超级增强子研究提供了便利。目前尚未发现有针对超级增强子开展的 NSCL/P 的病因学研究,本研究发现 NSCL/P 的 SNP 遗传度在超级增强子区富集,为进一步开展相关易感基因定位研究提供了依据。

计算 SNP 遗传度有助于量化常见遗传变异对表型风险的贡献,明确了进一步定位致病性常见变异的必要性,也有助于理解除常见变异外其他遗传因素的贡献程度。SNP 遗传度分区较为富集的区域可作为进一步研究的目标区域,可以作为一种新型的候选区域策略。本研究采用连锁不平衡得分回归计算 SNP 遗传度,其基于 GWAS 汇总数据,通过每个 SNP 位点关联研究结果即可计算 SNP 遗传度,不需要个体的基因型信息,计算效率高。Evans 等^[38]证明无论 GWAS 分析时是否很好地控制了环境变量或亲缘关系等因素,连锁不平衡得分回归都能对 SNP 遗传度进行较为准确的估计。本研究依据的 GWAS 通过家系设计提高了研究效能^[17],得到了每个 SNP 位点与表型关联的可靠结果,可以作为计算 SNP 遗传度的数据基础。虽然本研究使用连锁不平衡得分回归计算 SNP 遗传度发现截距为 1.004(在 1 附近),表明本研究潜在的混杂偏倚较少,并且研究结果与现仅有的一篇 NSCL/P 的 SNP 遗传度结果相近^[23],但本研究仍存在以下不足:本研究采用的连锁不平衡得分来源于千人基因组亚洲人群,尽管中国人群属于亚洲人群,但两者之间仍存在一定的不同,所以用此数据来评估中国人群 SNP 遗传度时可能会有一定的偏倚;本研究依据的 GWAS 基于核心家系设计,尽管在分析时采用了正确的估计方法控制了家系相关等因素,但在使用连锁不平衡得分回归估计 SNP 遗传度时仍可能有一定的残余混杂;Evans 等^[38]发现当致病性位点是常见变异时,连锁不平衡得分回归一般会低估约 5% ~ 10% 的遗传度,当致病位点包含很多罕见变异时则可能低估更多,因此,本研究可能低估了中国人群 NSCL/P 的 SNP 遗传度。综上,本研究是对中国人群 NSCL/P SNP 遗传度的初步探索,研究者可以根据所拥有的

数据类型选择相应的模型和算法来对 SNP 遗传度进行估计。未来也需要应用更大样本量、独立人群、覆盖更多 SNP 位点的 GWAS 数据,开发利用不同的统计分析模型对中国人群 NSCL/P 的 SNP 遗传度进行更加准确的估计。

通过评估中国人群的 SNP 遗传度和分区 SNP 遗传度,本研究提示尚有致病性常见变异未被发现,并且 SNP 遗传度主要富集在增强子区和超级增强子区。未来可以采用更大样本量等增加研究效能的方法来识别未被发现的常见变异,也可以针对遗传度较为富集的分区进行大样本的靶向捕获测序,探索该目标分区内的遗传变异与 NSCL/P 的关联。

利益冲突 所有作者均声明不存在利益冲突。

作者贡献声明 薛恩慈:提出研究思路,设计研究方案,撰写论文;陈曦、王雪珩、王斯悦、王梦莹:设计研究方案;李劲、秦雪英、武轶群、李楠、李静、周治波:收集、分析、整理数据;朱洪平:收集、分析、整理数据,总体把关和审定论文;吴涛:提出研究思路,撰写论文,总体把关和审定论文;陈大方、胡永华:总体把关和审定论文。

参考文献

- [1] Mossey PA, Little J, Munger RG, et al. Cleft lip and palate [J]. *Lancet*, 2009, 374(9703): 1773-1785.
- [2] Wang M, Yuan Y, Wang Z, et al. Prevalence of orofacial clefts among live births in China: A systematic review and meta-analysis [J]. *Birth Defects Res*, 2017, 109(13): 1011-1019.
- [3] Sivertsen A, Wilcox AJ, Skjaerven R, et al. Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives [J]. *BMJ*, 2008, 336(7641): 432-434.
- [4] Grosen D, Bille C, Petersen I, et al. Risk of oral clefts in twins [J]. *Epidemiology*, 2011, 22(3): 313-319.
- [5] Leslie EJ, Carlson JC, Shaffer JR, et al. Association studies of low-frequency coding variants in nonsyndromic cleft lip with or without cleft palate [J]. *Am J Med Genet A*, 2017, 173(6): 1531-1538.
- [6] Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease [J]. *Nat Rev Genet*, 2010, 11(6): 446-450.
- [7] Yang J, Zeng J, Goddard ME, et al. Concepts, estimation and interpretation of SNP-based heritability [J]. *Nat Genet*, 2017, 49(9): 1304-1310.
- [8] Eaves LJ, Last KA, Young PA, et al. Model-fitting approaches to the analysis of human behaviour [J]. *Heredity (Edinb)*, 1978, 41(3): 249-320.
- [9] Keller MC, Coventry WL. Quantifying and addressing parameter indeterminacy in the classical twin design [J]. *Twin Res Hum Genet*, 2005, 8(3): 201-213.
- [10] Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies [J]. *Nat Genet*, 2015, 47(3): 291-295.
- [11] Zhu H, Zhou X. Statistical methods for SNP heritability estimation and partition: A review [J]. *Comput Struct Biotechnol J*, 2020, 18: 1557-1568.
- [12] Consortium GP. An integrated map of genetic variation from 1 092 human genomes [J]. *Nature*, 2012, 491(7422): 56-65.

- [13] Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants [J]. *Nat Genet*, 2014, 46(3): 310–315.
- [14] GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans [J]. *Science*, 2015, 348(6235): 648–660.
- [15] Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes [J]. *Nature*, 2015, 518(7539): 317–330.
- [16] Won HH, Natarajan P, Dobbyn A, et al. Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease [J]. *PLoS Genet*, 2015, 11(10): e1005622.
- [17] 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.
- [18] Haeussler M, Zweig AS, Tyner C, et al. The UCSC genome browser database; 2019 update [J]. *Nucleic Acids Res*, 2019, 47(D1): D853–D858.
- [19] Hoffman MM, Ernst J, Wilder SP, et al. Integrative annotation of chromatin elements from ENCODE data [J]. *Nucleic Acids Res*, 2013, 41(2): 827–841.
- [20] Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease [J]. *Cell*, 2013, 155(4): 934–947.
- [21] Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics [J]. *Nat Genet*, 2015, 47(11): 1228–1235.
- [22] Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler [J]. *Nat Rev Genet*, 2014, 15(11): 765–776.
- [23] Ludwig KU, Böhmer AC, Bowes J, et al. Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only [J]. *Hum Mol Genet*, 2017, 26(4): 829–842.
- [24] Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height [J]. *Nat Genet*, 2014, 46(11): 1173–1186.
- [25] Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology [J]. *Nature*, 2015, 518(7538): 197–206.
- [26] Pantelis C, Papadimitriou GN, Papiol S, et al. Biological insights from 108 schizophrenia-associated genetic loci [J]. *Nature*, 2014, 511(7510): 421–427.
- [27] Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery [J]. *Nature*, 2014, 506(7488): 376–381.
- [28] Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations [J]. *Nat Genet*, 2015, 47(9): 979–986.
- [29] 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.
- [30] Suazo J, Recabarren AS, Marín NR, et al. Association between IRF6 variants and nonsyndromic cleft lip with or without cleft palate in Chile [J]. *Reprod Sci*, 2020, 27(10): 1857–1862.
- [31] Zuccherro TM, Cooper ME, Maher BS, et al. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate [J]. *N Engl J Med*, 2004, 351(8): 769–780.
- [32] Rahimov F, Marazita ML, Visel A, et al. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip [J]. *Nat Genet*, 2008, 40(11): 1341–1347.
- [33] Morris VE, Hashmi SS, Zhu L, et al. Evidence for craniofacial enhancer variation underlying nonsyndromic cleft lip and palate [J]. *Hum Genet*, 2020, 139(10): 1261–1272.
- [34] Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes [J]. *Cell*, 2013, 153(2): 307–319.
- [35] Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVII and GATA2 deregulation in leukemia [J]. *Cell*, 2014, 157(2): 369–381.
- [36] Visel A, Minovitsky S, Dubchak I, et al. VISTA enhancer browser: A database of tissue-specific human enhancers [J]. *Nucleic Acids Res*, 2007, 35(Database issue): D88–D92.
- [37] Gao T, Qian J. EnhancerAtlas 2.0: An updated resource with enhancer annotation in 586 tissue/cell types across nine species [J]. *Nucleic Acids Res*, 2020, 48(D1): D58–D64.
- [38] Evans LM, Tahmasbi R, Vrieze SI, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits [J]. *Nat Genet*, 2018, 50(5): 737–745.

(2021-03-18 收稿)

(本文编辑:刘淑萍)