EPIGENETICS
2019, VOL. 14, NO. 2, 198–213
https://doi.org/10.1080/15592294.2019.1581591

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS

Epigenomic profiling of newborns with isolated orofacial clefts reveals widespread DNA methylation changes and implicates metastable epiallele regions in disease risk

Semira Gonseth[a,b], Gary M. Shaw[c], Ritu Roy[d], Mark R. Segal[e], Kripa Asrani[f], Jasper Rine[f], Joseph Wiemels[g], and Nicholas J. Marini[f]

[a]School of Public Health, University of California, Berkeley, CA, USA; [b]Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland; [c]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA; [d]Cancer Research Institute, University of California, San Francisco, CA, USA; [e]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; [f]California Institute for Quantitative Biosciences, University of California, Berkeley, CA, USA; [g]Center for Genetic Epidemiology, University of Southern California School of Medicine, Los Angeles, CA, USA

ABSTRACT
Cleft lip with or without cleft palate (CL/P) is a common human birth defect whose etiologies remain largely unknown. Several studies have demonstrated that periconceptional supplementation of folic acid can reduce risk of CL/P in offspring. In this study, we tested the hypothesis that the preventive effect of folic acid is manifested through epigenetic modifications by determining whether DNA methylation changes are associated with CL/P. To more readily observe the potential effects of maternal folate on the offspring epigenome, we focused on births prior to mandatory dietary folate fortification in the United States (i.e. birth year 1997 or earlier). Genomic DNA methylation levels were assessed from archived newborn bloodspots in a 182-member case-control study using the Illumina® Human Beadchip 450K array. CL/P cases displayed striking epigenome-wide hypomethylation relative to controls: 63% of CpGs interrogated had lower methylation levels in case newborns, a trend which held up in racially stratified sub-groups. 28 CpG sites reached epigenome-wide significance and all were case-hypomethylated. The most significant CL/P-associated differentially methylated region encompassed the *VTRNA2-1* gene, which was also hypomethylated in cases (FWER p = 0.014). This region has been previously characterized as a nutritionally-responsive, metastable epiallele and CL/P-associated methylation changes, in general, were greater at or near putative metastable epiallelic regions. Gene Set Enrichment Analysis of CL/P-associated DMRs showed an over-representation of genes involved in palate development such as *WNT9B, MIR140* and *LHX8*. CL/P-associated DNA methylation changes may partly explain the mechanism by which orofacial clefts are responsive to maternal folate levels.

ARTICLE HISTORY
Received 6 December 2018
Revised 31 January 2019
Accepted 8 February 2019

KEYWORDS
Orofacial cleft; birth defect; epigenetics; DNA methylation; metastable epiallele; folic acid

Introduction

Although their etiologies are largely unknown, isolated orofacial clefts are suspected of being etiologically heterogeneous with both genetic and non-genetic risk factors [1]. Clinical observations have indicated that isolated orofacial clefts should be classified into at least two distinct phenotypic groups: cleft lip with or without cleft palate (CL/P), and cleft palate alone (CPO) [2]. Collectively, these are among the most common birth defects with a worldwide prevalence of approximately 1 in 700 live births [3].

Genetic associations for CL/P have been observed for both common and rare variants in numerous genes from either linkage analysis, candidate gene, or genome-wide association studies [e.g. see 1,4,5]. The relatively large number of associated loci is probably indicative of the complex genetic etiologies underlying the CL/P phenotype and explain only a fraction of the population burden of these human birth defects. Several non-genetic factors also contribute to cleft phenotypes [6]. Of particular relevance to this study, maternal use of multivitamin supplements containing folic acid in early pregnancy has been associated with decreased risk of CL/P [7,8]. Although there are reports of no such reduction in risk [see ref. 9], the benefit of periconceptional folic acid has been reinforced in subsequent meta-analyses [10,11].

CONTACT Nicholas J. Marini nmarini@berkeley.edu 324 Stanley Hall California Institute for Quantitative Biosciences University of California Berkeley, CA 94720
Supplementary data for this article can be accesed here.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

The underlying mechanisms by which folic acid may reduce CL/P risks are unknown. The essential role of folate-dependent reactions in the synthesis of the cellular methyl group donor S-adenosylmethionine (AdoMet) has pointed to epigenetic alterations (DNA and histone methylation). Several observations support this hypothesis. First, multiple studies suggest a critical role for DNA methylation changes in normal murine palatogenesis [12,13] as well as in murine models of isolated orofacial clefts [14,15]. Second, like folate, increased maternal intake of choline has been associated with decreased CL/P risk [16]. The conversion of choline to glycine is a second route toward AdoMet synthesis and metabolic imbalances as well as genetic variation in this pathway have been associated with CL/P [5,17]. Thus, more than one nutrient affecting methyl donor status has been associated with CL/P. Third, the critical period for reduction of risk by periconceptional supplementation of folate and/or choline coincides with the reprogramming of the fetal epigenome in early embryogenesis [18,19]. Indeed, some epigenomic regions (the so-called Metastable Epialleles), are particularly sensitive to maternal nutritional status around the time of conception and, thus, display inter-individually variable levels of DNA methylation [20-22]. These regions are more likely to be associated with enhancers or transcription start sites and so may be enriched for epigenetic changes with functional consequence [21].

More directly, some reports have demonstrated DNA methylation changes associated with orofacial clefts in humans [23,24]. Sharp et al [23] described differentially methylated regions that may be characteristic for different subtypes of clefts (e.g. isolated cleft palate vs. cleft palate with cleft lip). Alvizi et al [24] described methylation changes at individual CpGs from a CL/P case-control study and suggested that DNA methylation changes may interact with genetic risk factors to increase penetrance. However, the relevance of these sites to metastable epiallelic regions in the epigenome is not yet known.

In this report, we further evaluated maternal folic acid status in fetal epigenetic modifications relevant to orofacial clefts by examining DNA methylation changes in a CL/P case-control study. However, we restricted our analysis to only births prior to mandatory dietary folate fortification in the United States. We imposed this criterion with the rationale that, by eliminating the global epigenomic impact of folate fortification, nutrient-sensitive epigenomic differences between CL/P cases and controls would be more easily observable. We found that CL/P case newborns from pre-fortification births displayed a striking epigenome-wide trend toward hypomethylation. Moreover, the most extensive case-associated methylation changes occurred at or near putative metastable epialleles. CL/P-associated differentially methylated regions implicated genes previously known to be involved in orofacial clefts and/or palatogenesis.

Results

This study evaluated the extent to which epialleles of CpG dinucleotides might account for the well-established ability of folate supplementation to reduce risk of isolated CL/P. Seventy-four Hispanic and 20 non-Hispanic white neonates with CL/P were included, as well as 69 Hispanic and 19 non-Hispanic white nonmalformed controls (Table 1). The distribution of other clinical variables (e.g. maternal age, gestational age, birth weight, etc.) was comparable between ethnic groups and between cases and controls. The most noteworthy feature of the sample population was that all bloodspots used for analysis were collected from newborns born prior to mandatory dietary fortification of folic acid in the United States (birth years 1997 or earlier). We imposed this criteria hypothesizing that doing so would allow us to more easily observe folate (or other nutrient)-sensitive epigenomic differences between CL/P cases and controls should they exist.

Table 1. Study population.

Race/Ethnicity	Sex	Cases *(Collective Percentage)*	Controls *(Collective Percentage)*
White, Hispanic	M	44 *(46.8)*	42 *(47.7)*
	F	30 *(31.9)*	27 *(30.7)*
White, non-Hispanic	M	13 *(13.8)*	9 *(10.2)*
	F	7 *(7.4)*	10 *(11.4)*
TOTAL		**94**	**88**

Widespread hypomethylation effects

We observed a fairly extensive CL/P-associated hypomethylation across the epigenomic profile. Out of the 319,253 CpGs queried on the array (after applying filters), 202,672 (63.5%) displayed a negative regression coefficient (indicating a lower methylation level in cases) in the entire population (Figure 1). Furthermore, the most significantly changing CpGs (lowest p value) were almost exclusively hypomethylated. Stratification of the sample population by ethnicity showed a nearly identical phenomenon in both sub-group populations. Case hypomethylation was observed for most CpGs in Hispanic (61.2%) and non-Hispanic white (61.5%) newborns.

The significance of this hypomethylation profile was further supported by the high degree of concordance observed between Hispanic and non-Hispanic white subgroups with respect to the directionality of case-associated changes. The subset of CpG sites with nominally significant changes ($p < 0.05$ or $p < 0.01$) in both ethnic subgroups were strikingly concordant (χ^2 $p <$ 1E-30) and heavily biased toward hypomethylation (Figure 2). Indeed, there was substantial concordance throughout the profiles even for the many CpG sites that did not reach statistical significance in either subgroup (Fig. S1). These data indicate that, despite sample size and ethnicity difference, methylation signatures for these subgroups were similar and that widespread hypomethylation is related to the CL/P phenotype. Thus, in most subsequent analyses, we combined both ethnic sub-groups to enhance statistical power, while potentially mitigating any confounding from ethnically-sensitive positions. Summary statistics for all 319,253 CpG sites in the combined population and ethnically-stratified sub-groups are in supplementary Tables S1-S3.

EWAS analysis

Twenty-eight individual CpGs reached genome-wide significance for methylation changes (Bonferroni cut-off p = 1.57E-07) when considering all samples (Table 2, Fig. S3). Interestingly, all 28 CpG positions were hypomethylated in cases. This directionality was also seen in Hispanic and non-Hispanic white cases when these subgroups were analyzed separately (not shown). While most of these sites were closely associated with a gene, there was no statistically significant enrichment of functional elements (e.g. DNaseI hypersensitive sites, gene promoters) as has been suggested previously [24; data not shown]. A cross-validation of the top differentially methylated CpGs in the Hispanic and non-Hispanic white subsets resulted in the replication of the top two (*cg19689947* in ZNF77 and *cg02718229* in EHD2; Table 2) as being consistently and significantly less methylated in CL/P cases than in controls (cross-validation $p < 0.05$ in both groups).

Overall, CL/P-associated EWAS profiles displayed more significantly associated CpG sites than expected by chance, as illustrated by the left deviation of the p value histogram and the upper deviation of the quantile-quantile (q-q)

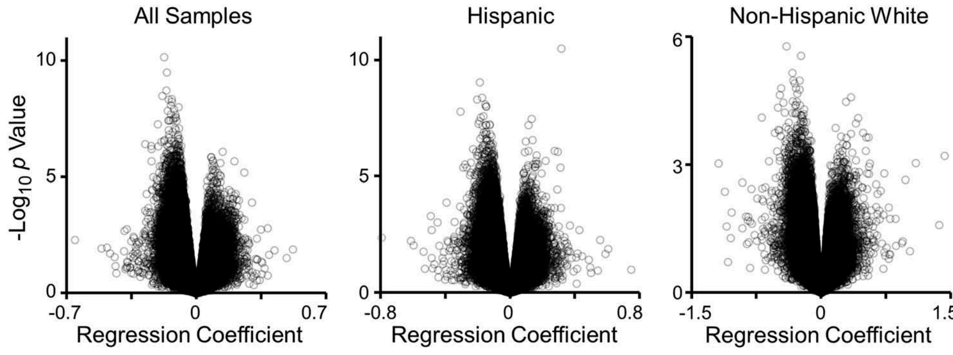


Figure 1. Regression coefficients for 319,253 CpGs in the final, filtered analytical dataset are plotted versus p value (expressed as $-\log_{10} p$). **Left panel**: All samples (94 cases, 88 controls). **Middle panel**: Hispanic-only samples (74 cases, 69 controls). **Right panel**: non-Hispanic white-only samples (20 cases, 19 controls).

Figure 2. Directional concordance of regression coefficients between Hispanic and non-Hispanic white sample sets for subsets of CpG sites. The numbers of CpG sites in each quadrant are indicated. **Left panel**: All 319,253 CpG in the analytical dataset. **Middle panel**: 2,694 CpG sites with p value < 0.05 in both sample sets. **Right panel**: 262 CpG sites with p value < 0.01 in both sample sets.

Table 2. The 28 differentially methylated positions meeting epigenome-wide significance (p = 1.57E-07).

CpG ID	Chrom	Position	Associated Gene[a]	Regression Coefficient	p value	Case Avg β val	Ctrl Avg β val	Location in Gene[b]
cg19689947	19	2941933	*ZNF77*	−0.173	7.08E-11	0.782	0.804	Body
cg02718229	19	48230862	*EHD2*	−0.160	3.32E-10	0.831	0.849	Body
cg13113642	17	997200	*ABR*	−0.163	1.97E-09	0.913	0.924	Body
cg00401471	19	38782834	*SPINT2*	−0.183	3.32E-09	0.718	0.746	3′UTR
cg17399234	5	110428454	*WDR36*	−0.133	4.48E-09	0.066	0.076	Body
cg27052183	11	66411986	*RBM4*	−0.150	4.78E-09	0.827	0.842	3′UTR
cg22906451	19	1994824	*BTBD2*	−0.132	6.03E-09	0.843	0.857	Body
cg24007247	7	6410669		−0.138	8.42E-09	0.869	0.882	
cg05911749	16	30956151	*FBXL19*	−0.096	1.02E-08	0.906	0.912	Body
cg22726273	8	145268244	*HEATR7A*	−0.121	1.28E-08	0.886	0.896	3′UTR
cg23072571	19	54291978	*MIR373*	−0.154	1.45E-08	0.804	0.822	Body
cg15471981	19	54291942	*MIR373*	−0.138	2.05E-08	0.853	0.865	TSS200
cg15089892	3	101568188	*NFKBIZ*	−0.113	2.13E-08	0.029	0.032	5′UTR
cg07805699	3	159481815	*SCHIP1*	−0.111	2.66E-08	0.047	0.052	Body
cg15333818	12	46766724	*SLC38A2*	−0.113	2.94E-08	0.051	0.057	TSS200
cg09610578	3	129407542	*TMCC1*	−0.122	3.16E-08	0.828	0.838	TSS200
cg11855842	1	23807492	*ASAP3*	−0.141	3.65E-08	0.883	0.895	Body
cg20801637	1	1795408	*GNB1*	−0.126	3.67E-08	0.822	0.837	5′UTR
cg02466815	2	177054573	*HOXD1*	−0.150	3.71E-08	0.083	0.093	Body
cg23987548	17	73401912	*GRB2*	−0.106	3.82E-08	0.025	0.027	TSS200
cg11390978	19	16986831	*SIN3B*	−0.150	3.87E-08	0.870	0.885	Body
cg05612966	5	95583220		−0.209	5.57E-08	0.915	0.931	
cg24406116	1	109940036	*SORT1*	−0.093	8.17E-08	0.040	0.043	Body
cg10057318	1	91966336	*CDC7*	−0.115	8.38E-08	0.069	0.077	TSS1500
cg25137403	2	177022172		−0.111	9.56E-08	0.036	0.039	
cg07779777	19	4047755	*ZBTB7A*	−0.150	9.83E-08	0.891	0.903	Body
cg14125003	11	67978396	*SUV420H1*	−0.127	1.18E-07	0.837	0.852	5′UTR
cg02756655	2	65663632		−0.148	1.53E-07	0.241	0.266	

[a] According to UCSC genome browser via Illumina Genome Studio software

[b] Gene-associated CpGs are localized to gene body (between start/stop codons, including introns), untranslated regions (UTR), and upstream proximity (kb) to the transcription start site (TSS).

plot (Fig. S2). The q-q plot presented a general inflation factor = 1.69 (based on median $\chi 2$). For the top two CpGs, the p value corrected for genomic inflation = 6.4E-05 (cg19689947) and 0.00012 (cg02718229).

Differentially methylated region (DMR) analysis

CpG sites in close proximity in the genome are often highly correlated and tend to overlap functional elements. Therefore, differential methylation signals in these regions may be more biologically relevant than

individual CpGs. The most significantly CL/P-associated DMRs are described in Table S4. The most significant association was seen for the gene region surrounding the small, non-coding RNA *VTRNA2-1*, which was significantly less methylated in CL/P cases (Figure 3). This was the sole DMR that remained significant following correction for multiple testing (FWER p value = 0.014 for the entire population). In support of this finding, we performed 100 bump hunting analyses following permutation of the case/control variable; 87 identified *VTRNA2-1* as a DMR resulting in a permutation-based corrected p value of 0.02 (initial p value ranked 2nd over 100).

In analyses following ethnic stratification, the *VTRNA2-1* region was also the most significantly associated DMR in both Hispanics and non-Hispanic whites, though in neither case did the significance survive the FWER-correction for multiple testing (Table S4). Sensitivity analyses in all participants and in stratified sub-populations produced similar results with *VTRNA2-1* consistently presenting as one of the most significantly CL/P-associated DMRs, independent of the parameters of the bump hunting function (Table S5). Interestingly, a survey of the top DMRs in Tables S4 and S5 does not indicate a similarly strong trend toward case hypomethylation as was seen from individual CpG sties in EWAS analysis. The significance of this is discussed below.

Metastable epiallele regions and disease risk

The appearance of the *VTRNA2-1* region as the most significant DMR associated with CL/P is notable since this CpG-rich segment has been previously characterized as a 'metastable epiallele' [20]. Metastable epialleles are regions whose methylation states are established in early embryogenesis and are sensitive to periconceptional environment, particularly maternal nutrition status [20-22]. Since maternal periconceptional intake of folate and choline, both of which contribute to the synthesis of methyl donors necessary for DNA methylation, can mitigate risk for CL/P, it is plausible that epigenomic regions sensitive to in utero environment in early embryogenesis (i.e. metastable epialleles) may contribute to disease risk. Although *VTRNA2-1* has yet to be implicated in CL/P risk, a role is still formally possible. Alternatively, changes at this locus may simply be a sentinel reflecting more widespread metastable epiallele-related changes, in general, in newborn case epigenomes. In support of this hypothesis, we observed a trend of greater case-associated DNA methylation changes at CpG sites near putative metastable allele regions (Figure 4, Figure S4). Metastable epiallele regions in the epigenome have thus far been defined based on CpG clusters with inter-individual variability and systemic intra-individual similarity [20,21]. We observed that the average magnitude of CL/P-associated methylation changes for array CpGs (both positive and negative

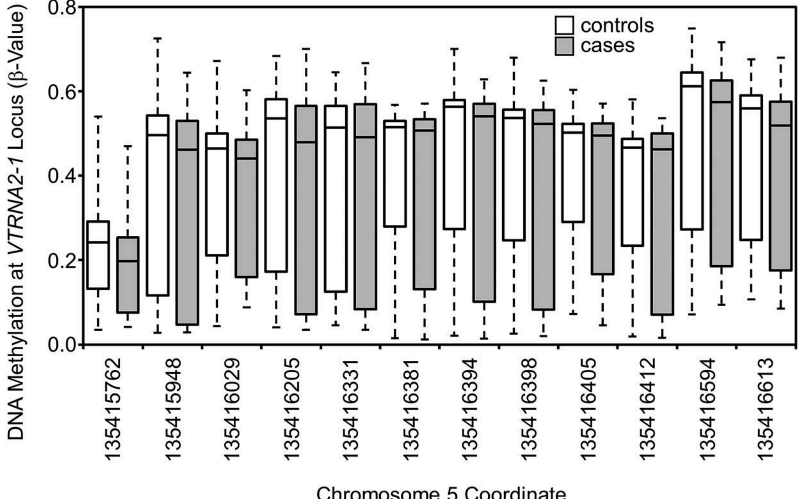
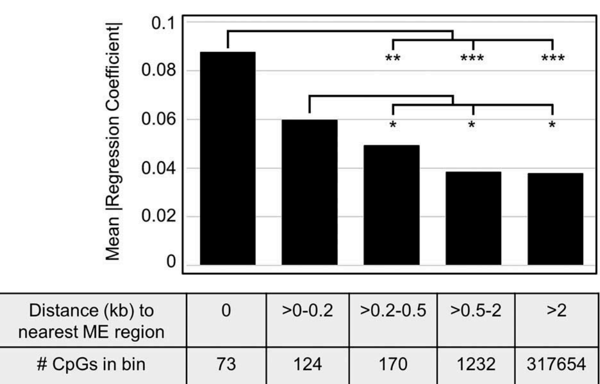


Figure 3. Distributions of DNA methylation β-values for each of the 12 CpG sites (identified by chromosomal coordinate) within the ~1 kb region surrounding the *VTRNA2-1* gene. Data is from all cases (N = 94, gray-filled boxes) and all controls (N = 88, white-filled boxes). *VTRNA2-1* transcript coordinates are chr5:135416187–135416286.



Distance (kb) to nearest ME region	0	>0-0.2	>0.2-0.5	>0.5-2	>2
# CpGs in bin	73	124	170	1232	317654

Figure 4. All CpGs were binned based on proximity (kb) to putative metastable epiallele regions as defined and localized in Table S2 from ref. 20. For each distance bin shown in the figure, the average extent of case-associated methylation changes were calculated by averaging the absolute value of the regression coefficients (to equally account for hypo- and hypermethylation events). # CpGs averaged per bin is indicated. A distance of '0' indicates CpGs that fall within the borders of a metastable epiallele region. Significance of pair-wise distribution comparisons was determined by the Mann Whitney U test. * $p < .05$; ** $p < 0.005$; *** $p < 0.0005$.

as measured by the absolute value of the regression coefficient) increased significantly based on nearness to these metastable epiallele regions (Figure 4; reference dataset was table S2 in ref. 20 defining 4,853 200bp genomic segments with metastable epiallele characteristics). In the same study, these authors also define a more discernible subset of putative epialleles (N = 109) with dense CpG clustering (Fig. S4a), and a recent study [21] further defined 687 epigenome-wide metastable regions with different clustering/ length criteria (Fig. S4b). Using any of these 3 datasets as reference, we saw a significantly greater extent of methylation changes at or near metastable epiallele regions in general. Furthermore, besides *VTRNA2-1*, two additional CL/P-associated DMRs (*PAX8-AS1*, chr 2; *ZFP57*, chr6; Table S5) overlap metastable epialleles whose methylation levels are influenced by season of conception in populations where maternal nutrient availability varies widely by season [20]. Thus, some of the strongest CL/ P-associated DNA methylation changes occurred at epigenomic regions known to be sensitive to periconceptional environment.

Gene set enrichment analysis

To understand the link between methylation changes at cleft-associated DMRs or individual CpGs, and how these might be implicated in disease risk, we performed Gene Set Enrichment Analysis of associated genes (see Methods). The geneset highlighted by DMRs displayed significant enrichment for processes involved in craniofacial development, palatogenesis and clefting. Although *VTRNA2-1* was the only DMR to maintain significance following FWER multiple comparison corrections, the bump-hunting analysis identified a total of 472 CL/ P-associated DMRs encompassing 427 genes with p values ranging from 5E-06 to 0.08. Some DMRs from this analysis were nearby or adjacent, thus multiple DMRs highlight the same region and associated gene. All DMRs were considered in this analysis regardless of p value (geneset in Table S6). Function enrichment queries with this unfiltered gene set returned nominally significant enrichments ($p < 0.05$) for relevant processes/functions (Table S7) such as genes implicated in Head Development (p = 1.3E-04), Cleft Lip (p = 8.7E-03), and Midline Defects (p = 7.9E-03). Thus, these analyses suggested that relevant biological processes may be affected by CL/P-associated methylation changes that may provide at least some of the mechanistic link underlying the nutritional modification of CL/P risk.

In contrast, the genes associated with individual CpGs displaying genome-wide significance in EWAS analysis (Table 2) show no significant enrichment of relevant developmental processes, nor have any been previously implicated in orofacial cleft risk or palatogenesis. Expanding the relevant gene set to identify cleft-related signals by considering highly ranked sites that fell short of genome-wide significance (e.g. top 500 CpGs), similarly showed no functional enrichment in processes related to craniofacial development or dysmorphology.

Orofacial cleft-relevant DMRs

Overall, the bump hunting analyses revealed 10 DMR-linked genes with evidence for their involvement in CL/P risk (Table 3). These genes have been implicated in clefting by either human genetic association studies, or by their roles in palatogenesis and/ or cleft development in vertebrate model organisms. Although the biological impact of altered methylation patterns at these sites is unknown, there were several noteworthy DMRs that occurred at or near the transcription start sites for their respective genes

Table 3. Ten differentially methylated regions that contain at least 2 nearby CpGs and are closely associated with genes implicated in orofacial clefts.

Chromosome	Start	DMR length	# CpG in DMR	Value	p value[a]	Nearest Gene	mRNA RefSeq	Location in Gene[b]	Reference[c]
16	69966902	161	4	0.043	0.017	*MIR140*	NR029681	TSS overlap	25,26
1	75590912	442	4	−0.034	0.046	*LHX8*	NM001001933	TSS −2766	27
1	3239916	312	3	0.029	0.021	*PRDM16*	NM022114	Intron 3, TSS +254174	46
17	44928210	307	3	0.034	0.042	*WNT9B*	NM003396	Promoter, TSS −451	14,47
1	164545553	147	3	−0.034	0.044	*PBX1*	NM002585	Intron 2, TSS +16956	48
16	84860766	152	3	0.030	0.045	*CRISPLD2*	NM031476	Intron 1, TSS +7179	49
14	95330880	105	2	0.031	0.058	*GSC*	NM173849	Intergenic, TSS −94381	50
14	54418728	77	2	−0.03	0.067	*BMP4*	NM001202	Exon 3, TSS +4826	51,52
1	23279554	132	2	−0.028	0.074	*EPHB2*	NM004422	Intergenic, TSS +242222	53
1	12656232	84	2	0.026	0.082	*DHRS3*	NM004753	Intron 1, TSS +21504	54

[a] Uncorrected for multiple comparisons
[b] Gene feature location with distance (bp) upstream or downstream of the transcription start site (TSS)
[c] Reports that gene is associated with isolated orofacial clefts (human or vertebrate models) and/or palatogenesis.

(Figure 5). Measuring the difference in mean beta values (i.e. the fraction of methylated sites) between cases and controls, we observed a DMR hypermethylated in cases surrounding the transcription start site of the microRNA *MIR140* (Figure 5, top panel), which has been implicated in orofacial cleft risk and palatogenesis in both humans [25] and zebrafish [26]. We also observed a region of case hypermethylation in the promoter of *WNT9B*, a signaling protein influencing embryonic development and implicated in cleft risk based on a murine model [14]. Figure 5 (bottom panel) also shows a DMR hypomethylated in CL/P cases that overlaps a CpG island, DNaseI hypersensitive clusters and binding sites for multiple transcription factors, ~ 2.8kb upstream of the *LHX8* TSS. Deletion of *LHX8*, a homeobox member transcription factor, results in a cleft palate phenotype in lhx8-mutant mice [27].

Interestingly, the DMRs associated with the set of 10 CL/P-related genes (Table 3) are more likely to be near putative metastable epiallele regions than other DMRs identified (Figure S5). When compared against other DMRs near other genes (either those implicated in craniofacial development but not obviously related to clefts (N = 45), or those not annotated as craniofacially related (N = 372)), this set of 10 was significantly more enriched for regions within 100, 50 and 10kb of putative metastable epiallele regions, using the same reference dataset as that in Figure 4. However, it should be noted that we did not see the same enrichment for epiallele proximity using the metastable epiallele regions defined in ref. 21. Thus, while providing further evidence for the implication of metastable epiallele regions in cleft risk, the significance of these results await further study.

Discussion

We explored the possible mechanism by which maternal periconceptional folic acid intake reduces risk for offspring with isolated orofacial clefts by investigating fetal DNA methylation changes at birth. To that end, we measured methylation levels in newborns at individual CpG sites or at loci where clustered CpG sites behaved coordinately to define a Differentially Methylated Region (DMR). We restricted analysis to only births that occurred prior to mandatory dietary folate fortification in the United States (birth year 1997 or earlier) with the expectation that, in the absence of widespread dietary fortification, the case-control differential in folate intake might be more significant and, therefore, folate-related events would be easier to observe. Our focus on pre-fortification births is supported by the observation that rates of occurrence of isolated orofacial clefs in California decreased in the post-fortification period [28]. In this population, CL/P case newborns displayed a striking epigenome-wide

Figure 5. The difference in mean β-values between CL/P cases relative to controls for the individual CpGs (delineated as chromosomal coordinates) that comprise DMRs described in Table 3. Plots depict the DMR CpGs plus 6 adjacent CpGs that flank the DMR on either end. **Top panel**: *MIR140*; transcription start site at coordinate 69966984. **Middle panel**: *WNT9B* promoter; transcription start site at coordinate 44928968. **Bottom panel**: CpG island (Chr1:75590818–75591354) upstream of *LHX8* transcription start site (coordinate 75594119).

trend toward hypomethylation. Moreover, the most extensive case-associated methylation changes occurred at or near putative metastable epiallele regions. CL/P-associated DMRs were near genes previously implicated to be involved in orofacial clefts and/or palatogenesis.

We observed a fairly extensive hypomethylation in CL/P cases: nearly two-thirds (63.5%) of the CpG sites interrogated had lower methylation levels in case newborns, a tendency which held up in racially stratified sub-populations (61% in Hispanics, 62% in non-Hispanic whites; Figure 1). Furthermore, the most significantly CL/P-associated sites were heavily skewed toward hypomethylation: the 28 CpG sites that reached epigenome-wide significance in EWAS analysis all had lower methylation levels in cases (Table 2). Indeed, this trend was present throughout the methylation profile even for sites with nominal *p* values. For example, 78% of CpG sites with p value < 0.05 (total N = 41,429) had lower methylation levels in cases, as did 74% of CpG sites with p value < 0.2 (total N = 102,312). These trends are seen equally well in each ethnic sub-group. Indeed, sites with greater significance in both sub-populations are predominantly hypomethylated (Figure 2).

It is tempting to postulate that such a widespread decrease in DNA methylation levels in CL/P cases reflects exposure to a limiting amount of methyl donors, which result from folate (or other 1-carbon-related metabolite) insufficiencies. Thus, CL/P-associated DNA methylation changes at nutritionally-sensitive epigenomic loci may partly explain the mechanism by which orofacial cleft risk is responsive to maternal folate supplementation, particularly if such changes affect genes involved in craniofacial development. However, the relationship between maternal folate intakes and epigenetic patterning in offspring is not necessarily a straightforward one, as evidenced by epigenome-wide studies. For example, Gonseth et al [29] observed an inverse correlation between periconceptional folate supplementation and methylation in newborns in that mostly hypomethylation events were observed with increasing maternal folate intake. This trend, though counter-intuitive, is consistent with observations made on methylation sites in newborn DNA that associate with maternal plasma folate levels during pregnancy [30] and within offspring DMRs associated with low- and high-dose maternal folate supplementation [31].

Nevertheless, different experimental designs in these studies make it difficult to generalize the relationship between maternal folate levels and fetal DNA methylation. For example, Amerasekera et al [32] identified 7 newborn DMRs significantly hypo- or hyper-methylated relative to maternal serum folate levels during the 3rd trimester of pregnancy. However, at 5 of these loci, Gonseth et al [29] observed the opposite directional effects with respect to maternal periconceptional folate intakes. Two of these regions were identified as CL/P-associated DMRs in this study (*ZFP57*, chr6:29,648,507–623; *FZD7*, chr2:202,901,352–428) where their methylation status indicated lower periconceptional folate exposure in cases [according to the results of ref. 29].

While some of these inconsistencies can be attributed to ascertaining folate levels at different times during pregnancy (i.e. early vs. late), ascribing methylation changes to a single folate exposure variable may capture only a portion of the relevant nutritional information. For example, maternal vitamin B12 levels can influence infant global and site-specific DNA methylation patterns, possibly through its role as cofactor in the Methionine Synthase (MTR) reaction necessary for AdoMet synthesis [33]. In addition, choline is an essential micronutrient involved in one-carbon metabolism whose oxidative derivative, betaine, also converts homocysteine to methionine, which provides an additional synthetic route for AdoMet. Thus, multiple nutritional inputs (and their interaction) probably contribute toward setting fetal epigenomic programming. Importantly, maternal levels of vitamin B12 and choline have been implicated in risk of CL/P in offspring [5,17,34].

The most significantly CL/P-associated DMR was the region encompassing the *VTRNA2-1* gene, which was hypomethylated in cases (Figure 3). This CpG-rich locus has been previously characterized as a nutritionally-responsive, metastable epiallele [20,31]. Silver et al [20] examined DNA methylation profiles in infants in Gambia, where seasonality is linked to significant differences in maternal diet, and found that *VTRNA2-1* methylation levels in offspring positively correlated with maternal serum folate levels in the season of conception. On the other hand, Richmond et al [31] found an inverse correlation of methylation levels at *VTRNA2-1* in adult children and pre-natal maternal folic acid supplementation, but only at very high doses of folate supplementation (5 mg/day).

As part of the explanation for the complex nature of the relationship between maternal folate and newborn methylation, Gonseth et al [29] suggested that the fetal epigenomic response to periconceptional folate intake

may be bi-phasic. Their analyses, when stratified by maternal folate status, indicated a *positive* correlation between newborn methylation levels and folate intake in births from folate-insufficient mothers (consumption of <200 μg/d), whereas negative correlations become more pronounced with higher folate intakes. They hypothesize that folate may be limiting for DNA methylation below a certain threshold, while increasing folate in folate-replete populations may produce compensatory or negative feedback effects. In this regard, the population used in the current study, which examines a stark folate-sensitive clinical phenotype in a pre-fortification population, may accentuate effects of folate-limiting conditions.

Along with the widespread hypomethylation changes, we also observed a significant genomic inflation on the q-q plots indicating more statistically significant case-control differences across the epigenome than expected (Fig. S2). This may be due to unaccounted population substructure (e.g. from a genetically admixed subpopulation of Hispanic participants) or it may result from a global effect of an environmental factor on DNA methylation, such as folate intake. We believe that the latter is more probable since genomic inflations have been commonly observed in previous studies determining the impact of maternal folate levels on offspring DNA methylation states [29-31]. This, in fact, may be expected since it seems reasonable to assume that under such conditions, much of the epigenome may be affected.

Our results point toward metastable epialleles as regions that epigenetically distinguish CL/P cases from nonmalformed controls. Because epigenetic programming at metastable epialleles and risk for CL/P intersect at periconceptional nutrition status, particularly nutrients involved in one-carbon metabolism, they are logical candidates at which to look for cleft-related signals. It is therefore noteworthy that we observed: 1) several CL/P-associated DMRs occurring at previously annotated metastable epialleles (Figure 3, Table S4), 2) the extent of methylation changes was greater at or very near previously described metastable epiallele regions (Figure 4, Fig. S4), and 3) Out of all CL/P-associated DMRs, the most compelling candidates were significantly closer to epiallelic loci (Fig. S5). To make these claims, we have used datasets from initial studies annotating metastable epiallele regions in the human epigenome [20,21]. It is likely, however, that these regions will be more comprehensively defined in future studies and so it will be of particular interest to see if these relationships are supported. It is worth noting that metastable epialleles have also been implicated as sites in offspring epigenomes that are particularly sensitive to maternal folic acid supplementation [31].

CL/P-associated DMRs were associated with genes that were enriched for relevant functional categories and highlighted 10 genes that have previously been implicated in orofacial clefts and/or palatogenesis (Table 3,Table S7). In the absence of direct measurement, it is difficult to predict the effect of methylation events on gene expression, especially since some transcription factors prefer to bind methylated sites [35]. Nevertheless, there were 2 DMRs that are particularly deserving of attention. We detected a case-hypermethylated region surrounding the transcription start site of the microRNA *MIR140* (Figure 5). *MIR140* was first implicated in orofacial cleft formation from experiments in zebrafish demonstrating that altered expression levels resulted in craniofacial abnormalities (including cleft palate), through its role in modulation of *PDGFRA* [26]. This relationship was subsequently observed in murine cells where *MIR140* expression is reduced by environmental smoke exposure [25,36]. Importantly, a human polymorphism that affects *MIR140* processing is associated with CL/P [37] and shows a significant interaction for CL/P risk with passive smoke exposure in the first trimester [25]. We hypothesize that environmental impacts of *MIR140* expression, which contribute to CL/P risk, could be mediated via DNA methylation changes at the DMR reported in this study.

We observed a second case-hypermethylated region in the promoter of *WNT9B* (Figure 5), a secreted protein essential for murine craniofacial development and implicated in malformations including CL/P [38]. The potential for DNA methylation changes to affect *WNT9B* transcription thereby leading to CL/P risk is reminiscent of a murine model for isolated orofacial clefts with an epigenetic basis [14]. In the A/WySn mouse strain, a DNA retrotransposon of the IAP (intracisternal A particle) family has inserted downstream of the

Wnt9b gene and drives an anti-sense transcript that interferes with Wnt9b transcription. The effect is variable, based on the degree of methylation and silencing of the IAP, in which embryos with the lowest levels of Wnt9b transcript develop cleft lip. Thus, the IAP appears to be acting as a metastable epiallele. Although the mechanism in this particular mouse model is one of antisense interference rather than promoter function, it does demonstrate that a decrease in *WNT9B* transcription can contribute to CL/P risk. It is worth noting at this DMR, as well as those at *MIR140* and *LHX8*, may be specific for orofacial clefts: none of these were observed in a separate Neural Tube Defect case population (also pre-fortification) when compared against the same controls used in this study (manuscript in preparation).

Recent studies have also interrogated DNA methylation changes as part of the underlying mechanism for orofacial cleft occurrence [23,24]. In an effort to identify methylation changes that are specific to cleft subtypes, Sharp et al [23] identified DMRs in direct case:case comparisons (e.g. Cleft Palate Only vs. Cleft Lip Only infants). The authors found significant differences relative to each phenotypic subtype, with the most significant changes seen when comparing infants with a cleft lip (CLO, CL/P) to those without (CPO). Because this was not a case: control design, results are somewhat difficult to compare to those in the current study. Nevertheless, we observed a significant overlap between these cleft lip-specific DMRs (N = 268) and the CL/P-associated DMRs reported here (N = 472): 14 DMRs co-localized to within 1kb, 12 of which were overlapping (see Table S8 for details). Thus, some DMRs that distinguish cleft lip from cleft palate are also observed in this case-control study, though none of the associated genes are obviously involved in craniofacial development.

Alvizi et al [24] queried DNA methylation changes at individual CpGs in a CL/P case-control population of Brazilian children. These authors found 578 significantly associated CpGs that were enriched for functional regions of the genome (e.g. promoters, DNaseI hypersensitive sites, etc.) and developmentally important pathways, such as WNT/β-catenin signaling. Furthermore, the majority of these sites (N = 361; 62.3%) were case-hypomethylated. Beyond this, however, individual CpGs significantly associated in the Brazilian study were strikingly discordant with data in the current study at the same positions. At 420 sites (of the 578) that overlap both studies 280 were hypomethylated in Brazilian cases whereas we observed hypermethylation at 279 sites (77% discordance; Fisher's exact p value = 1.4E-22). Furthermore, only 11 of these sites were represented in the most significant 10,000 sites from our EWAS analysis and all were discordant: 11/11 were hypermethylated in Brazilian case children yet hypomethylated in our case newborns. It is worth noting that Brazil adopted mandatory folate fortification in 2004, thus folate status may be a critical difference between these populations. This raises the important question as to how epigenetic changes affect disease risk in folate-limiting versus folate-replete populations. Importantly, these authors also presented evidence that promoter DNA methylation at *CDH1*, a gene often mutated in familial cases of CL/P, can interact with genetic variation at *CDH1* in a way whereby decreased expression may increase penetrance of pathogenic variants. It will be of interest to see if this a more general mechanism for the contribution of cleft-associated methylation changes.

Limitations/strengths

The strengths of this study lie in the case-control design and the inclusion of different ethnicities to mitigate race/ethnic-specific differences, conducted prior to mandatory folic acid fortification. We have postulated that folate-sensitive events may be more obvious during this time period since the overall levels of folate intake were lower and more susceptible to consequential insufficiencies. Furthermore, we have proven the effectiveness of genome-scale methylation analysis starting with a limiting quantity (≤ 35 ng) of gDNA from newborn bloodspots archived up to 30 years.

The study also has several limitations. We ascertained DNA methylation profiles from newborn whole blood, with the expectation that informative differences presumably occurring in developing orofacial tissues would be preserved in this lineage. Although probable, especially at metastable epiallele regions where DNA methylation levels are established prior to cell differentiation, it is possible that there may be additional or more pronounced

changes in embryonic lip or palatal tissue. Indeed, this may be the reason that the CL/P-associated methylation signals pointed our attention toward metastable epiallele regions. Previous epigenetic studies on orofacial cleft have established a correlation between methylation patterns seen in postnatal blood versus lip or palate tissue, particularly in differentially methylated regions associated with this phenotype [23,24]. The DMRs identified in these and the current study suggest that blood methylation patterns will reflect at least some of the clinically relevant changes.

We also have no direct measurements of maternal folate status in this population. We have assumed that maternal folate status would be different between cases and controls and, based on the epidemiology, would likely favor controls. The observation of CL/P-associated DMRs at or near metastable epiallele regions supports a case-control nutritional difference, but may not allow unequivocal assignment of lower folate status to cases. Other nutrients that impinge on one-carbon precursors, such as choline and vitamin B12, may also underlie case-control differences in this population. Alternatively, some methylation changes may be influenced by local genetic variation (methylation QTLs), though the widespread hypomethylation trends observed here argues against localized genetic influences. Nevertheless, the unequivocal implication of maternal folate status in CL/P-associated epigenetic changes can only be made with accompanying maternal folate measurements or a comparison before/after dietary folate fortification.

Lastly, although starting with low gDNA inputs is adequate, this does result in a mild (yet measurable) increase in variability at some loci [39]. This, combined with a small sample size, may have generated spurious associations or, more likely, underestimated significant associations. Therefore, replication of these results in a second, larger population is essential.

Methods

Study population

This case-control study included data on infants delivered between 1988–97. The study included live-born infants with isolated CL/P (cases; N = 97) or without any structural malformation (controls; N = 93). Case information was abstracted from hospital reports and medical records following established procedures by the California Birth Defects Monitoring Program [40]. Each medical record was further reviewed by a medical geneticist (Edward Lammer, MD). Infants with trisomies were ineligible. Nonmalformed controls were selected randomly to represent the population from which the cases were derived in selected counties and birth periods. This study, including the collection and use of archived newborn bloodspots, was approved by the California State Committee for the Protection of Human Subjects as well as Institutional Review Boards at Stanford University and the University of California, Berkeley.

DNA methylation assay from archived bloodspots

Genomic DNA (gDNA) from each case or control newborn bloodspot was isolated from a single, surgical punch (2mm dia.) using the QIAamp DNA Micro kit (QIAGEN) according to the manufacturer, with a final column elution volume of 25 µl. Average yield was 40 ng gDNA/punch. For each sample, 30–35 ng gDNA were bisulfite converted using the EpiTect Bisulfite kit (QIAGEN) and eluted in 16 µl buffer EB. Four µl of eluate was processed for epigenome-wide profiling on the HumanMethylation450 BeadChip® array (Illumina) as per manufacturer's instructions. We have previously reported that this process is robust and reproducible for profiling limiting starting quantities of bloodspot gDNA [39].

Data quality filtering

Array CpGs with detection p values > 0.01 in 15% of samples were excluded from analysis. Likewise, samples in which >15% of CpGs had detection p values > 0.01 (of the 450k loci) were also excluded. Two samples had mismatched reported and predicted gender and were removed from the analysis. In addition, cross-reactive, poor quality and SNP-associated probes were removed from the analysis [41]. The final analytical dataset was comprised of 319,253 CpG sites queried in 94 mostly-Hispanic cleft cases and 88 controls (see

Table 1 for population breakdown). Methylation data pre-processing was performed according to Fortin et al [42]. All genomic coordinates are based on human genome hg19, build 37.

Epigenome-wide association analysis

Epigenome-wide association studies (EWAS) were carried out to agnostically discover differentially methylated CpG sites between case and control newborns. Linear regression models (by the function 'lmfit' in the 'limma' R package) were carried out, with DNA methylation beta-values (normalized into M-values) as the outcome, and case/control status as the variable of interest. Adjustments were made to account for the following potential confounding factors: 1) gestational age, 2) gender, 3) estimated cell-mixture (by the *Refactor* method which uses an unsupervised feature selection step followed by a sparse principal component analysis [43]), 4) an estimation of ancestry (by the *Epistructure* method which uses the principal components of CpG sites known to be influenced by genetic variation [44]), 5) batch effects, and 6) ethnicity as reported on the birth certificates. The lambda genomic inflation factor was calculated on p values using the 'estlambda' function of the 'GenABEL' R package. T-statistics resulting from the linear regressions were divided by the lambda genomic inflation factor to obtain p values corrected for multiple testing. The Bonferroni-corrected $p = 1.57 \times 10^{-7}$ was used as the genome-wide level of statistical significance. Manhattan plots and quantile-quantile (q-q) plots were drawn according to the 'qqman' R package. Additional analyses stratified by ethnicity were carried out. Additionally 1,000 bootstrap EWAS analyses were performed in both ethnic groups separately to decrease potential influences of outliers. Bootstrap-resulting p values were calculated by taking the geometric mean of the distribution of bootstrap p values for each association. A bootstrap p value below 0.05 and a concordance in the direction of the association were the criteria that were applied to cross-validate CL/P-associated sites.

Bump-hunting analysis

To discover differentially methylated regions (DMRs) between cases and controls, we used bump-hunting methods [45] and the corresponding function 'bumphunter' from the 'bumphunter' R package. We carried out genome-wide bump hunting analysis in all participants together, and also following ethnic stratification, while setting DNA methylation beta-values as the outcome and case/control status as the variable of interest. The same adjustments were made as for EWAS analysis. A thousand bootstraps were performed to calculate Family-Wise Error Rate (FWER) p values within the 'bumphunter' function. The maxGap (i.e., number of base pairs to include in a 'bump') was set at 500. Additional sensitivity analyses were performed by varying maxGap (250 to 2000 bp), and setting the smoothing function on/off. Finally, we ranked the p value of the most associated DMR against a distribution of p values generated under the null by 100 bump hunting analyses while making permutations of case/control status. The rank of the original p value was divided by 100 to obtain a permutation-based corrected p value. All statistical analyses were performed using the statistical software R, version 3.2.1.

Gene set enrichment analysis

To identify biologically relevant mechanisms (and potential cleft risk genes) indicated by the differential methylation data, we evaluated genes associated with individual CpGs, as well as the DMRs defined in the bump-hunting analyses. CpGs or DMRs were filtered based on p value, overlap of functional elements (e.g. – promoter, first intron, etc.) and/or nearness (bp) to putative metastable epiallele regions as partially defined previously [20,21; see Results]. Associated genes were defined as those with transcription start sites nearest the CpG or DMR of interest. Enrichment of Gene Ontology (GO) or pathway terms was determined using Ingenuity Pathway Analysis (IPA, version #42012434, www.qiagenbioinformatics.com). Within the IPA package, Fisher's exact test was used to measure the probability that a category was randomly associated. In all cases, terms/processes with p values for enrichment < 0.05 were considered of interest. A second software package (DAVID: Database for Annotation, Visualization and Integrated Discovery, v6.8; www.david.ncifcrf.gov) was also used primarily for verification of the annotations and enrichments revealed from the IPA analysis.

Acknowledgments

This work was supported by research grants from the SICPA foundation (S.G.), the National Institutes of Health (HD074695 to N.J.M.), and by a cooperative agreement from the Centers for Disease Control and Prevention (U01 DD001033 to Stanford University). We thank the California Department of Public Health, Maternal Child and Adolescent Health Division for providing surveillance data from California for this study. We thank Gavin Schlissel for help with statistical analyses.

Disclaimers: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the California Department of Public Health.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Centers for Disease Control and Prevention [U01 DD001033]; National Institute of Child Health and Human Development (NIH) [HD074695]; SICPA Foundation [none].

ORCID

Nicholas J. Marini http://orcid.org/0000-0001-8782-2573

References

[1] Dixon MJ, Marazita ML, Beaty TH, et al. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet. 2011;12:167–177.

[2] Fogh-Anderson P. Genetic and non-genetic factors in the etiology of facial clefts. Scand J Plastic Reconstr Surg. 1967;1:22–29.

[3] WHO. Global registry and database on craniofacial anomalies. Report of a WHO registry meeting on craniofacial anomalies. Vol. 12. Baura Brazil; 2001. p. 1–3. World Health Organization (pub.)

[4] 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. Am J Hum Genet. 2015;96:397–411.

[5] Marini NJ, Yang W, Asrani K, et al. Sequence variation in folate pathway genes and risks of human cleft lip with or without cleft palate. Am J Med Genet A. 2016;170:2777–2787.

[6] Mossey PA, Little J, Munger RG, et al. Cleft lip and palate. Lancet. 2009;374:1773–1785.

[7] Shaw GM, Lammer EJ, Wasserman CR, et al. Risks of orofacial clefts in children born to women using multivitamins containing folic acid periconceptionally. Lancet. 1995;346:393–396.

[8] Wilcox AJ, Lie RT, Solvoll K, et al. Folic acid supplements and risk of facial clefts: national population based case-control study. BMJ. 2007;334:464.

[9] Little J, Gilmour M, Mossey PA, et al. Folate and clefts of the lip and palate-a U.K.-based case-control study: part I: dietary and supplemental folate. Cleft Palate Craniofac J. 2008;45:420–427.

[10] Butali A, Little J, Chevrier C, et al. Folic acid supplementation use and the MTHFR C677T polymorphism in orofacial clefts etiology: an individual participant data pooled-analysis. Birth Defects Res A Clin Mol Teratol. 2013;97:509–514.

[11] Jahanbin A, Shadkam E, Miri HH, et al. Maternal folic acid supplementation and the risk of oral clefts in offspring. J Craniofac Surg. 2018;29:e534-e541.

[12] Seelan RS, Appana SN, Mukhopadhyay P, et al. Developmental profiles of the murine palatal methylome. Birth Defects Res A Clin Mol Teratol. 2013;97:171–186.

[13] Seelan RS, Mukhopadhyay P, Warner DR, et al. Epigenetic regulation of Sox4 during palate development. Epigenomics. 2013;5:131–146.

[14] Juriloff DM, Harris MJ, Mager DL, et al. Epigenetic mechanism causes Wnt9b deficiency and nonsyndromic cleft lip and palate in the A/WySn mouse strain. Birth Defects Res A Clin Mol Teratol. 2014;100:772–788.

[15] Kuriyama M, Udagawa A, Yoshimoto S, et al. DNA methylation changes during cleft palate formation induced by retinoic acid in mice. Cleft Palate Craniofacial J. 2008;45:545–551.

[16] Shaw GM, Carmichael SL, Laurent C, et al. Maternal nutrient intakes and risk of orofacial clefts. Epidemiology. 2006;1:285–291.

[17] Shaw GM, Vollset SE, Carmichael SL, et al. Nested case-control study of one-carbon metabolites in mid-pregnancy and risks of cleft lip with and without cleft palate. Pediatr Res. 2009;66:501–506.

[18] Okae H, Chiba H, Hiura H, et al. Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. 2014;10:e1004868.

[19] Smith ZD, Chan MM, Humm KC, et al. DNA methylation dynamics of the human preimplantation embryo. Nature. 2014;511:611–615.

[20] Silver MJ, Kessler NJ, Hennig BJ, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol. 2015;16:118.

[21] Kessler NJ, Waterland RA, Prentice AM, et al. Establishment of environmentally sensitive DNA methylation states in the very early human embryo. Sci Adv. 2018;4:eaat2624.

[22] Dominguez-Salas P, Moore SE, Baker M, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. Nat Commun. 2014;5:3746.

[23] Sharp GC, Ho K, Davies A, et al. Distinct DNA methylation profiles in subtypes of orofacial cleft. Clin Epigenetics. 2017;9:63.

[24] Alvizi L, Ke X, Brito LA, et al. Differential methylation is associated with non-syndromic cleft lip and palate and contributes to penetrance effects. Sci Rep. 2017;7:2441.

[25] Li L, Zhu GQ, Meng T, et al. Biological and epidemiological evidence of interaction of infant genotypes at Rs7205289 and maternal passive smoking in cleft palate. Am J Med Genet A. 2011;155A:2940–2948.

[26] Eberhart JK, He X, Swartz ME, et al. MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis. Nat Genet. 2008;40:290–298.

[27] Zhao Y, Guo YJ, Tomac AC, et al. Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene lhx8. Proc Natl Acad Sci. 1999;96:15002–15006.

[28] Yang W, Carmichael SL, Shaw GM. Folic acid fortification and prevalences of neural tube defects, orofacial clefts, and gastroschisis in California, 1989 to 2010. Birth Defects Res A Clin Mol Teratol. 2016;106:1032–1041.

[29] Gonseth S, Roy R, Houseman EA, et al. Periconceptional folate consumption is associated with neonatal DNA methylation modifications in neural crest regulatory and cancer development genes. Epigenetics. 2015;10:1166–1176.

[30] Joubert BR, Den Dekker HT, Felix JF, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun. 2015;7:10577.

[31] Richmond RC, Sharp GC, Herbert G, et al. The long-term impact of folic acid in pregnancy on offspring DNA methylation: follow-up of the Aberdeen Folic Acid Supplementation Trial (AFAST). Int J Epidemiol. 2018;47:928–937.

[32] Amarasekera M, Martino D, Ashley S, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. FASEB J. 2014;28:4068–4076.

[33] McKay JA, Groom A, Potter C, et al. Genetic and non-genetic influences during pregnancy on infant global and site specific DNA methylation: role for folate gene variants and vitamin B12. PLoS One. 2012;7(3):e33290.

[34] Wallenstein MB, Shaw GM, Yang W, et al. Periconceptional nutrient intakes and risks of orofacial clefts in California. Pediatr Res. 2013;74:457–465.

[35] Yin Y, Morgunova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356:6337.

[36] Izzotti A, Calin GA, Arrigo P, et al. Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. FASEB J. 2009;23:806–812.

[37] Li L, Meng T, Jia Z, et al. Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. Am J Med Genet A. 2010;152A:856–862.

[38] Jin YR, Han XH, Taketo MM, et al. Wnt9b-dependent FGF signaling is crucial for outgrowth of the nasal and maxillary processes during upper jaw and lip development. Development. 2012;139:1821–1830.

[39] Asrani K, Shaw GM, Rine J, et al. DNA methylome profiling on the infinium humanMethylation450 array from limiting quantities of genomic DNA from a single, small archived bloodspot. Genet Test Mol Biomarkers. 2017;21:516–519.

[40] Croen LA, Shaw GM, Jensvold NG, et al. Birth defects monitoring in California: a resource for epidemiological research. Paediatr Perinat Epidemiol. 1991;5:423–427.

[41] Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina infinium humanMethylation450 microarray. Epigenetics. 2013;8:203–209.

[42] Fortin JP, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503.

[43] Rahmani E, Zaitlen N, Baran Y, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. Nat Methods. 2016;13:443–445.

[44] Rahmani E, Shenhav L, Schweiger R, et al. Genome-wide methylation data mirror ancestry information. Epigenetics Chromatin. 2017;10:1.

[45] Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012;41:200–209.

[46] Bjork BC, Turbe-Doan A, Prysak M, et al. Prdm16 is required for normal palatogenesis in mice. Hum Mol Genet. 2010;19:774–789.

[47] 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. Nat Commun. 2017;8:14364.

[48] Ferretti E, Li B, Zewdu R, et al. A conserved Pbx-Wnt-p63-Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. Dev Cell. 2011;21:627–641.

[49] Chiquet BT, Henry R, Burt A, et al. Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection. Birth Defects Res A Clin Mol Teratol. 2011;91:44–49.

[50] Kuratani S, Satokata I, Blum M, et al. Middle ear defects associated with the double knock out mutation of murine goosecoid and Msx1 genes. Cell Mol Biol (Noisy-Le-Grand). 1999;45:589–599.

[51] Zhang Z, Song Y, Zhao X, et al. Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. Development. 2002;129:4135–4146.

[52] Suzuki S, Marazita ML, Cooper ME, et al. Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet. 2009;84:406–411.

[53] Dravis C, Henkemeyer M. Ephrin-B reverse signaling controls septation events at the embryonic midline through separate tyrosine phosphorylation-independent signaling avenues. Dev Biol. 2011;355:138–151.

[54] Billings SE, Pierzchalski K, Butler Tjaden NE, et al. The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. FASEB J. 2013;27:4877–4889.