



POPULATION STUDY ARTICLE

Epigenome-wide analysis identifies genes and pathways linked to acoustic cry variation in preterm infants

Ghazal Aghagoli^{1,2,3}, Stephen J. Sheinkopf^{1,2,3,4}, Todd M. Everson⁵, Carmen J. Marsit⁵, Hannah Lee^{1,3}, Amber A. Burt⁵, Brian S. Carter⁶, Jennifer B. Helderan⁷, Julie A. Hofheimer⁸, Elisabeth C. McGowan^{1,2,3}, Charles R. Neal⁹, T. Michael O'Shea⁸, Steven L. Pastyrnak¹⁰, Lynne M. Smith¹¹, Antoine Soliman¹², Lynne M. Dansereau^{1,3}, Sheri A. DellaGrotta^{1,3}, James F. Padbury^{2,3} and Barry M. Lester^{1,2,3,4}

BACKGROUND: Preterm birth places infants at higher risk of adverse long-term behavioral and cognitive outcomes. Combining biobehavioral measures and molecular biomarkers may improve tools to predict the risk of long-term developmental delays.

METHODS: The Neonatal Neurobehavior and Outcomes in Very Preterm Infants study was conducted at nine neonatal intensive care units between April 2014 and June 2016. Cries were recorded and buccal swabs collected during the neurobehavioral exam. Cry episodes were extracted and analyzed using a computer system and the data were summarized using factor analysis. Genomic DNA was extracted from buccal swabs, quantified using the Qubit Fluorometer, and aliquoted into standardized concentrations. DNA methylation was measured with the Illumina MethylationEPIC BeadArray, and an epigenome-wide association study was performed using cry factors ($n = 335$).

RESULTS: Eighteen CpGs were associated with the cry factors at genome-wide significance ($\alpha = 7.08E - 09$). Two CpG sites, one intergenic and one linked to gene *TCF3* (important for B and T lymphocyte development), were associated with acoustic measures of cry energy. Increased methylation of *TCF3* was associated with a lower energy-related cry factor. We also found that pitch (F_0) and hyperpitch ($F_0 > 1$ kHz) were associated with DNA methylation variability at 16 CpG sites.

CONCLUSIONS: Acoustic cry characteristics are related to variation in DNA methylation in preterm infants.

Pediatric Research (2021) 89:1848–1854; <https://doi.org/10.1038/s41390-020-01172-0>

IMPACT:

- Preterm birth is a major public health problem and its long-term impact on health is not well understood.
- Cry acoustics, related to prematurity, has been linked to a variety of medical conditions.
- Biobehavioral measures and molecular biomarkers can improve prediction tools for long-term developmental risks of preterm birth.
- Variation in epigenetic modulation in preterm infants provides a potential link between preterm birth and unfavorable developmental outcomes.

INTRODUCTION

Preterm birth is a major public health concern, accounting for nearly one out of every ten births in the United States.¹ Even though survival of infants born before 30 weeks postmenstrual age (PMA) has increased due to improved medical technologies and interventions,² preterm birth still places infants at risk of developing chronic diseases, brain injuries, and adverse long-term behavioral and cognitive outcomes such as learning disabilities, autism spectrum disorders, cerebral palsy, epilepsy, and various other psychopathologies.^{3–5} These potential outcomes further lead to an increased need (and government spending) for healthcare, education, and supportive resources in addition to

the emotional and financial challenges for the children's family members.

Preterm birth is also associated with a range of medical conditions that could affect the developmental integrity of these infants. Acoustic characteristics of the infant's cry have been related to prematurity and to a broad spectrum of medical conditions, including, asphyxia, hyperbilirubinemia, trisomy anomalies, sudden infant death syndrome, prenatal drug exposure, longer-term developmental outcome in preterm infants, and, more recently, risk for autism.⁶ The acoustic characteristics related to these medical conditions famously include measures related to vibration of the vocal folds responsible for the fundamental

¹Brown Center for the Study of Children at Risk, Providence, RI, USA; ²Department of Pediatrics, Warren Alpert Medical School of Brown University, Providence, RI, USA; ³Department of Pediatrics, Women and Infants Hospital of Rhode Island, Providence, RI, USA; ⁴Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, USA; ⁵Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA; ⁶Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO, USA; ⁷Department of Pediatrics, Wake Forest School of Medicine, Winston Salem, NC, USA; ⁸Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, USA; ⁹Department of Pediatrics, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA; ¹⁰Department of Pediatrics, Spectrum Health-Helen DeVos Hospital, Grand Rapids, MI, USA; ¹¹Department of Pediatrics, Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA, USA and ¹²Department of Pediatrics, Miller Children's and Women's Hospital Long Beach, Long Beach, CA, USA
Correspondence: Barry M. Lester (Barry_Lester@Brown.edu)

Received: 13 April 2020 Revised: 23 July 2020 Accepted: 1 September 2020

Published online: 23 September 2020

frequency or “pitch” of the cry. Some of these characteristics are audible to humans such as the high pitched in *cry in cri du chat*, named for its distinctive sound. These acoustic characteristics are largely due to brainstem (cranial nerves 9–12) innervation of the larynx and pharynx, suggesting that acoustic cry analysis could provide a biobehavioral barometer of the central nervous system (CNS) integrity.^{6,7}

Biological markers, such as epigenetic modification, can provide information about the internalization of environmental and experience-based stimuli. The inclusion of biological markers that reflect the molecular underpinnings of biobehavioral measures, such as acoustic cry characteristics, could help identify infants at risk for long-term developmental delays. Epigenetics refers to changes in gene expression potential that are not explained by DNA sequence changes and that are heritable across cellular divisions. DNA methylation is one of the most studied epigenetic modifications and involves the addition of a methyl group to cytosine-phosphate-guanine (CpG) islands. DNA methylation can be affected by environment and life experiences, and is thought to be one mechanism through which environmental experience can confer long-term disease risk. In preterm infants, DNA methylation variability has been associated with conditions such as pain-related stress, sepsis, and medical and neurobehavioral risk.^{8–10} Differential methylation has also been related to cry acoustics in term, healthy neonates.¹¹ Combining both biobehavioral measures (such as acoustic cry characteristics) and molecular biomarkers could lead to a new understanding of biological mechanisms through which medical problems such as prematurity influence neurobehavioral outcomes. These insights could improve our ability to detect the risk of long-term developmental delays as early as possible. In previous work, we related variability in genome-wide DNA methylation to performance on the neonatal intensive care unit (NICU) Network Neurobehavioral Scale (NNS) in a US multisite cohort of infants born <30 weeks PMA.¹² While the field of biobehavioral epigenetics continues to grow, the only prior epigenetic studies of cry acoustics have focused specifically on the glucocorticoid receptor. Thus, we aimed to perform the first untargeted epigenome-wide study of cry acoustics to examine whether novel genes or pathways may be related to infant cry characteristics. Here, we profiled genome-wide DNA methylation to explore whether differences in DNA methylation are related to acoustic cry characteristics in these infants.

METHODS

Subjects

The Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) study was conducted at nine university-affiliated NICUs in Providence, RI, Honolulu, HI, Grand Rapids, MI, Torrance and Long Beach, CA, Kansas City, MO, and Winston-Salem, NC from April 2014 to June 2016. These NICUs also participated in the Vermont Oxford Network. All participating mothers provided written informed consent. The enrollment and consent procedures of this study were approved by the institutional review boards of Women and Infants Hospital, Spectrum Health, Children’s Mercy Office of Research Integrity, Wake Forest University Health Sciences, John F. Wolf, M.D., Human Subjects Committee at Los Angeles BioMed, Emory University, and Western Institutional Review Board; all methods used in the study were performed in accordance with the relevant regulations and guidelines. Eligibility was based on the following inclusion criteria: (1) birth at <30 weeks gestation; (2) parental ability to read and speak English or Spanish; and (3) residence within 3 h of the NICU and follow-up clinic. Exclusion criteria included maternal age <18 years, maternal cognitive impairment, maternal death, infants with major congenital anomalies, including CNS, cardiovascular, gastrointestinal, genitourinary, chromosomal, and nonspecific

anomalies,¹³ and NICU death. Parents of eligible infants were invited to participate in the study when survival to discharge was determined to be likely by the attending neonatologist. Overall, 704 infants were enrolled and buccal samples were collected from 624 of these infants for epigenomic screening. Demographic variables, including infant gender, race, ethnicity, maternal education, and partner status were collected from the maternal interview. The Hollingshead Index was used to assess socioeconomic status (SES), with Hollingshead level V, indicating low SES.¹⁴ Neonatal medical variables including birthweight, gestational age, length of NICU stay, weight at discharge, and gestational age at discharge were abstracted from medical records. Buccal swabs were collected during the week of discharge from the NICU. Thus, PMA at buccal swab collection represents the combination of gestational age at birth plus the length of NICU stay, and since DNA methylation has been associated with aging metrics, PMA was included as a covariate in our statistical analyses.

Measurements

DNA collection and methylation (DNAm) analysis. Genomic DNA was extracted from buccal swab samples obtained near-term-equivalent age, using the Isohelix Buccal Swab system (Boca Scientific). DNA was quantified using the Qubit Fluorometer (Thermo Fisher, Waltham, MA, USA), then aliquoted into a standardized concentration for subsequent analyses, and samples were randomly plated across 96-well plates. The Emory University Integrated Genomics Core performed bisulfite modification using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA). Subsequent assessment of genome-wide DNAm was performed using the Illumina MethylationEPIC Beadarray (Illumina, San Diego, CA) following standardized methods that were based on the manufacturer’s protocol. Samples with >5% of probes yielding detection p values $>1.0E-5$, with a mismatch between reported and predicted sex, and samples with incomplete covariate data were excluded from the analysis. The array data were normalized by functional normalization and standardized across probe designs Type-I and Type-II with beta-mixture quantile normalization.¹⁵ Probes that measured methylation on X and Y chromosomes, probes with single-nucleotide polymorphisms (SNPs) within the binding region, probes that could cross-hybridize to other regions of the genome, and probes that had low variability (beta values range <0.05) were removed.¹⁶ After exclusions, 706,278 probes were available for analysis from 341 samples that also had acoustic cry data. The methylation data are publicly available through the Gene Expression Omnibus via accession series GSE128821.

Tissue heterogeneity estimates. Cellular heterogeneity is a well-recognized source of confounding in epigenome-wide association studies.¹⁷ As a result, the proportion of epithelial, fibroblast, and immune cells (including B cells, natural killers, CD4⁺ T cells, CD8⁺ T cells, monocytes, neutrophils, and eosinophils) were estimated in cheek swab samples using reference methylomes.¹⁸ For 95% of our buccal swab samples, 95.7% of the cells were epithelial cells and immune cells were the majority of the remaining cell types, which were inversely correlated with epithelial cell proportions.¹² Thus, cellular heterogeneity was adjusted for by including the proportions of epithelial cells as a covariate in statistical models.

Recording and extraction of cry episodes. Cries of the preterm infants were elicited and recorded during the administration of the NNS during the week of NICU discharge (± 3 days). Audio recordings were made using an Olympus direct to PCM digital voice recorder and saved in an uncompressed.wav audio format (recording parameters: 16 bit, 48 kHz). Recorders were attached to the side of the infant’s bassinet/pram at a standardized location and oriented towards the infant’s mouth. This allowed for a

Table 1. Maternal and infant characteristics for included vs. not included subjects.

	Included (n = 289)	Not included (n = 312)	P value
Maternal characteristics, N (%) or mean (SD)			
Minority race or ethnicity	166 (57.4)	181 (58.0)	0.89
Maternal education (<high school)	37 (12.8)	42 (13.6)	0.78
No prenatal care	10 (3.5)	6 (1.9)	0.25
No partner	77 (26.6)	75 (24.1)	0.48
Low SES, Hollingshead = 5	27 (9.3)	32 (10.3)	0.69
	Included (n = 341)	Not included (n = 363)	P value
Neonatal characteristics, N (%) or mean (SD)			
Gender (female)	147 (43.4)	162 (45.3)	0.62
Cesarean delivery	237 (70.1)	258 (72.1)	0.57
PMA at birth (weeks)	27.1 (1.9)	27.0 (1.9)	0.38
Birthweight (g)	966 (284)	931 (276)	0.10
PMA at discharge (weeks)	40.0 (4.7)	41.1 (6.0)	<0.01
Discharge weight (g)	2989 (812)	3037 (986)	0.49
Head circumference (cm)	24.5 (2.5)	24.4 (2.4)	0.64
Length of NICU stay (LOS days)	90.0 (39.3)	98.5 (48.0)	<0.01
Severe retinopathy of prematurity	18 (5.3)	23 (6.4)	0.53
Sepsis	36 (10.6)	56 (15.6)	0.05
Necrotizing enterocolitis	22 (6.5)	29 (8.1)	0.41
Chronic lung disease	160 (47.2)	197 (55.0)	0.04
Parenchymal echodensity	28 (8.3)	26 (7.3)	0.63
Parenchymal echolucency	22 (6.5)	23 (6.5)	0.98
Germinal matrix hemorrhage	71 (21.0)	83 (23.3)	0.46
Any intraventricular hemorrhage	59 (17.5)	65 (18.3)	0.78
Ventricular dilation	31 (9.2)	25 (7.0)	0.30

standard distance between the infant's mouth and the recording device during the infant examination. Episodes of cry vocalizations were identified from these audio recordings. Cry episodes suitable for acoustic analysis were identified based on the absence of background noises that would interfere with the analysis (i.e., adult talk, medical equipment noises, and other environmental noises). The first suitable cry episode from each exam was excerpted into a.wav file for subsequent acoustic analysis. The extracted recordings only contained infant cry and were stripped of any identifiable information. Of the infants with DNA methylation array data, 335 had extractable cries (infant cried during the audio recording, good sound quality, no background noise, extractable and analyzable files), which were then analyzed.

Cry acoustic analysis. The extracted cry episodes were then analyzed using a computer system that has been specifically designed and well-validated to perform cry analyses in infants.¹⁹ Analysis proceeded in two phases. The first phase applied a cepstral-based acoustic analysis used to extract acoustic parameters with a 12.5 ms frame advance. A second phase organized and summarized this information in cry utterances. Cry utterances were defined as a cry during the expiratory phase of respiration

lasting at least 0.5 s. The software then calculated summary variables within the cry utterances.

Data analysis

In order to analyze the cry acoustic data, principal component analysis (PCA) was performed with rotation (oblimin with Kaiser normalization). This PCA allowed for dimension reduction, as there were too many cry variables to examine individually. The factors that explained the greatest proportion of variation were used for subsequent analyses. These factors each had an eigenvalue >1 and explained >10% of the proportion of variation.²⁰ The statistical analyses for epigenomic data were performed in R version 3.5. The epigenome-wide association study (EWAS) analysis (n = 335 infants with complete cry data and epigenomic data) was performed with robust linear regression using the MASS package and robust standard errors were estimated using sandwich package for each CpG site that passed QC. While the majority of infants in this study were from independent families (275 of 335), we estimated cluster robust standard errors, specifying family ID as a clustering variable, to account for potential correlations in the errors between siblings. The models were adjusted for infant sex, PMA, recruitment site, and proportions of epithelial cells. Separate models were run for each factor; factor 2 was found to have a bimodal distribution, thus the variable was dichotomized.

RESULTS

Maternal infant characteristics

Mothers of infants included in the current study of cry acoustics (n = 289) when compared to those not included (n = 312, their infants did not have buccal and cry data) had no significant differences in maternal demographic characteristics. Included infants (infants with buccal and cry data, n = 341) were discharged ~1 week earlier, stayed 8.5 days less in the NICU, and were less likely to have chronic lung disease or sepsis compared to study infants not included in the current analyses (n = 363) (Table 1).

At enrollment, 57.4% of mothers self-identified as minority, 26.6% did not have a partner, 9.3% had low SES, and 12.8% did not complete high school (Table 1). The average PMA at birth was 27.1 weeks and average birthweight was 966 g. Infants were discharged at 40 weeks and weighed on average 2989 g. The average length of stay in the NICU was 90 days. Infant medical complications included necrotizing enterocolitis (6.5%), sepsis (10.6%), chronic lung disease (47.2%), and brain injuries, including parenchymal echodensity (8.3%), parenchymal echolucency (6.5%), ventricular dilation (9.2%), and intraventricular hemorrhage (17.5%) (Table 1).

Cry acoustics

Two cry factors explained the greatest proportion of variation. They were frequency/energy-related (factor 1) and pitch/hyperpitch-related (factor 2) (Table 2). Factor 1 included variables that described the overall spectral energy contained in cry utterances, including the amplitude of the cry signal at the fundamental frequency, overall loudness of the utterances, and signal amplitude within specific frequency bands. Factor 2 included measures of fundamental frequency (F₀, or "pitch"), as well as the portions of utterances where F₀ was >1 kHz ("hyperpitch"). Factor scores were then calculated for each individual for the two significant cry factors such that an increase in the factor score meant a higher degree of frequency/energy (factor 1) or higher pitch/hyperpitch (factor 2). These two factors were used for the EWAS, for those infants that had both infant cry data and epigenomic data (n = 335).

EWAS of cry acoustics

An EWAS between DNA methylation and the two cry factors were performed to explore whether differences in DNA methylation in

buccal cells was related to acoustical cry characteristics. The genomic distributions of these results are presented in Manhattan (Fig. 1) and the observed vs. expected distributions of *p* values are presented via quantile–quantile (QQ) plots, which do reveal some inflation and thus a potential for residual confounding (Supplementary Fig. 1). The results of these models can be found in Supplementary Material (Supplementary Tables S1 and S2), and all associations reported had a false discovery rate (FDR) <10% ($q < 0.1$). Two CpG sites (cg08643005 and cg06728103) were associated with factor 1 at genome-wide significance ($\alpha = 7.08E - 09$), and we include 651 CpGs with FDR $q < 0.1$ in Supplementary Table S1. The CpG site cg08643005 demonstrated reduced

methylation with an increase in factor 1 scores (frequency/energy) and it is linked to the gene *TCF3*. Site cg06728103 showed significantly increased methylation with an increase in the frequency/energy factor scores, and it is located on chromosome 12; however, it is intergenic. The results for these two significant CpG sites, as well as the identified gene’s roles, can be found in Table 3.

For factor 2 (pitch/hyperpitch), 16 CpGs attained genome-wide significance ($\alpha = 7.08E - 09$), and we include 220 CpGs with FDR $q < 0.1$ in Supplementary Table S2. Of the sites that were genome-wide significant, five exhibited lower levels of DNA methylation, associated with higher scores for the pitch/hyperpitch factor compared to the group with lower scores, while the other 11 exhibited higher methylation. The CpG site, cg16459276 (intergenic), exhibited the greatest increase in methylation related to this pitch/hyperpitch factor, while cg05429445, in the *PPP1CA* gene, showed a similarly decreased extent. The results for these 16 significantly differentially methylated CpG sites as well as the identified gene’s roles can be found in Table 4.

CpG annotation

We identified phenotypes or traits that have been associated with the genes that were annotated to CpGs associated with cry acoustics at genome-wide significance (Tables 5 and 6).

DISCUSSION

We studied the association between genome-wide DNA methylation variation and changes in cry acoustic characteristics, particularly frequency/energy and pitch/hyperpitch. Our EWAS highlighted 18 CpGs that were significantly associated with the acoustic cry factors. Two CpG sites, one that was intergenic and one that was linked to gene *TCF3* (important for B and T lymphocyte development), were associated with frequency/energy-related cry variables. Increased methylation of *TCF3* was an association with less energy in the cry across multiple frequency bands. We also found significant differences in DNA methylation at 16 CpG sites that were associated with pitch/hyperpitch in the cry, four of which were related to the regulation of the cell cycle and DNA repair. Of these, *PPP1CA* gene (involved in cellular processes such as cell division) showed the greatest

Table 2. Factor loadings for cry acoustics^a.

Measure	Factor 1	Factor 2
Percentage of voiced fricative frames	−0.066	0.420
Pitch max (Hz)	0.102	0.404
Pitch STD (Hz)	0.084	0.447
Pitch average-Amp (dB)	0.599	0.073
Hyperpitch frames	−0.051	0.814
Hyperpitch average (Hz)	−0.008	0.969
Hyperpitch max (Hz)	−0.015	0.968
Hyperpitch min (Hz)	0.007	0.956
Hyperpitch average-Amp (dB)	0.015	0.966
Energy average	0.950	0.041
Energy max	0.925	0.014
Energy min	0.744	0.030
0.5–10 kHz Energy average	0.915	0.047
0–0.5 kHz Energy average (dB)	0.659	−0.037
0.5–1 kHz Energy average (dB)	0.699	0.029
1–2.5 kHz Energy average (dB)	0.817	0.083
2.5–5 kHz Energy average (dB)	0.655	−0.024
5–10 kHz Energy average (dB)	0.559	−0.141

^aBased on the loadings, factor 1 was frequency/energy and factor 2 was hyperpitch; factor loadings >0.4 were retained.

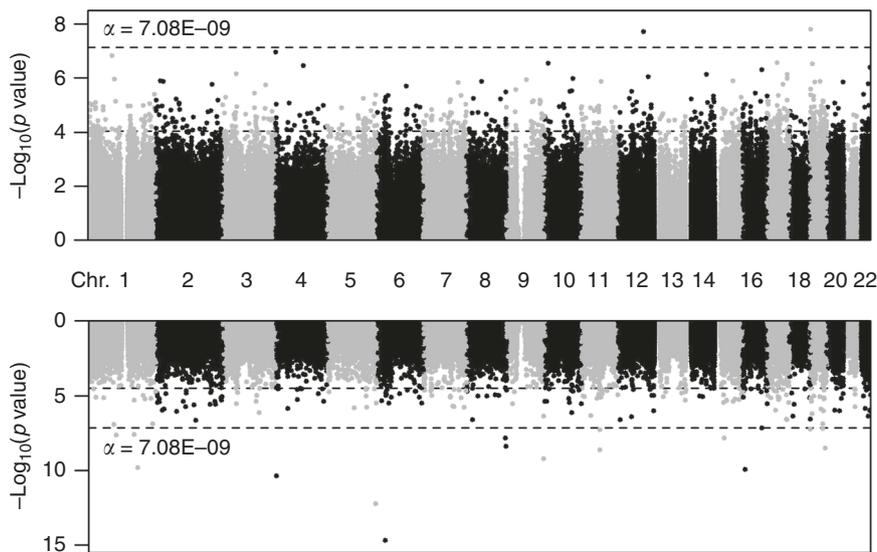


Fig. 1 Manhattan plots demonstrating the genomic distributions between DNA methylation levels and cry acoustics. The top panel demonstrates associations for cry acoustic factor 1 and the bottom panel demonstrates associations for cry acoustic factor 2, across 22 chromosomes (x-axis); alpha and the horizontal dotted lines represent the Bonferroni-threshold for statistical significance.

Table 3. Summary of the results for cry acoustics factor 1 (FDR < 10%) as well as biological roles of identified genes.

Location	CpG	β 1	P value	Gene	Biological role for gene
Chr12:93930699	cg06728103	0.0119	1.93E – 8	NA	
Chr19:1652627	cg08643005	–0.0025	1.55E – 8	<i>TCF3</i>	Encodes transcription factor 3 (TCF3), which activates transcription and is required for B and T lymphocyte development ²²

Table 4. Summary of the results for cry acoustics factor 2 (FDR < 10%) as well as biological roles of identified genes.

Location	CpG	β 1	P value	Gene	Biological role for gene
Chr4:1186469	cg01196336	–0.0127	4.21E – 11	<i>LOC 10013</i>	NA
Chr9:137263360	cg01601959	0.0530	5.84E – 10	<i>RXRA</i>	Encodes retinoid X receptors (RXRs) and retinoic acid receptors (RARs) that are involved in retinoic acid-mediated gene activation ²²
Chr11:67201998	cg03559915	–0.0110	5.42E – 8	<i>RPS6KB2</i>	Encodes kinase that phosphorylates the S6 ribosomal protein, which leads to an increase in protein synthesis and cell proliferation ²²
Chr1:180260316	cg03614159	0.0390	1.54E – 10	<i>ACBD6</i>	Encodes acyl CoA-binding domain protein 6, which is involved in the regulation of lipid metabolism ²²
Chr8:142517440	cg04744713	0.0155	1.43E – 8	<i>FLJ43860</i>	NA
Chr11:67169884	cg05429445	–0.1071	2.47E – 9	<i>PPP1CA</i>	Encodes protein phosphatase 1, which is involved in the regulation of cell processes such as cell division, protein synthesis, and glycogen metabolism ²²
Chr16: 8956240	cg06202930	0.0101	1.15E – 10	<i>CARHSP1</i>	Encodes calcium regulation heat-stable protein, which regulates target mRNA stability by binding to it ²²
Chr8:144902836	cg08924415	0.0523	4.03E – 9	<i>PUF60</i>	Encodes poly(U) binding splicing factor, which is important in nuclear processes such as pre-mRNA splicing, apoptosis, and transcription regulation ²²
Chr1:100889474	cg11357374	0.0554	2.23E – 8	<i>CDC14A</i>	Encodes dual-specificity phosphatase protein that is involved in cell mitosis and DNA replication, which suggests that it plays a role in cell cycle control ²²
Chr19:720793	cg14695566	0.0622	5.92E – 8	<i>PALM</i>	Encodes paralemmin, which is involved in plasma membrane dynamics in neurons in addition to other cell types. This protein is involved in D3 dopamine receptor binding ²²
Chr6:30433423	cg16459276	0.1034	2.14E – 15		NA
Chr19:45866931	cg17212420	0.0649	6.33E – 8	<i>ERCC2</i>	Encodes ERCC excision repair 2, involved in the nucleotide excision repair pathway ²²
Chr1:165831442	cg19880170	0.0900	2.66E – 8	<i>UCK2</i>	Encodes pyrimidine ribonucleoside kinase, which is involved in purine and pyrimidine metabolism pathway ²²
Chr15:35280520	cg19959497	–0.0087	1.55E – 8	<i>ZNF770</i>	Encodes zinc-finger protein 770, which is mainly involved in gene expression pathways ²²
Chr5:175301168	cg26303222	–0.0624	5.71E – 13	<i>CPLX2</i>	Encodes complexin 2, which is involved in synaptic vesicle exocytosis. Diseases associated with CPLX2 include schizophrenia and Huntington disease ²²
Chr19:54492713	cg26548871	0.0557	3.20E – 9	<i>CACN58</i>	NA

Table 5. Genes annotated to cry acoustics (factor 1)-associated CpGs that have been linked to traits from GWAS.

Gene	Trait types	Specific traits
<i>TCF3</i>	Neurological, immune	Lithium-responsive bipolar disorder, Hodgkins lymphoma

decrease in methylation, which was associated with increased hyperpitch in the cry.

Sixteen epigenetic loci were annotated in GWAS studies. Although these results likely include some false positives, a number of genes annotated to these CpGs have been linked to cellular growth, proliferation, and differentiation (*TCF3*, *CDC14A*, *ERCC2*, and *UCK2*), cognition (*RXRA* and *PALM*), and neurobehavioral or neurological disorders (*TCF3*, *RXRA*, *CARHSP1*, *PALM*, and *CPLX2*) in GWAS studies.²¹ We also found that the CpGs from this analysis were enriched for genes within pathways involving DNA binding and repair, cell cycle control, mRNA expression, immune response, and synaptic activity. Of particular note, four genes annotated to the significant CpGs were within the cell division Gene Ontology terms and pathways: protein phosphatase 1

catalytic subunit alpha (*PPP1CA*), ribosomal protein s6 kinase b2 (*RPS6KB2*), cell division cycle 14a (*CDC14A*), and uridine-cytidine kinase 2 (*UCK2*). The CpG site with the largest magnitude of inverse association between our statistically significant epigenetic loci, cg05429445, was within the body of the *PPP1CA*. This gene codes for one of the subunits of protein phosphatase 1, which is known to be involved in the regulation of cell processes, such as cell division, glycogen metabolism, and protein synthesis. Protein phosphatase 1 is involved in the regulation of ionic conductance and long-term memory.²² Experimental mouse models have suggested that protein phosphatase 1 functions as a suppressor of learning and memory.²³ Therefore, differential regulation of this gene is implicated in cognitive impairments. A number of these genes, such as *RPS6KB2*, *CDC14A*, *UCK2*, *TCF3*, and *RXRA*, have been associated with cellular growth, proliferation, and differentiation.^{24–31} These are primarily processes related to growth regulation and are highly active and dynamic during this early developmental period, and thus cry-associated differential methylation of these genes may be an indicator of the developmental state of these infants.

Of the genes associated with hyperpitch, two were related to neuronal dynamics: paralemmin-1 (*PALM*) and complexin 2 (*CPLX2*). The protein encoded by *PALM* is involved in neuronal

Table 6. Genes annotated to cry acoustics (factor 2)-associated CpGs that have been linked to traits from GWAS.

Gene	Trait types	Specific traits
<i>RXRA</i>	Cognitive, neurological, growth and metabolism, biological/visual	Intelligence, reaction time measurement, BMI-adjusted waist-hip ratio, eye morphology measurement
<i>RPS6KB2</i>	Hematopoietic	Mean corpuscular hemoglobin
<i>PPP1CA</i>	Biological, growth	Height
<i>CARHSP1</i>	Neurological	Brain aneurysm
<i>PUF60</i>	Cardiac	Diastolic blood pressure
<i>CDC14A</i>	Growth, cell division	Breast carcinoma
<i>PALM</i>	Neurobehavior, neurological, cognitive	Intelligence, anxiety, cognitive function measurement, worry measurement
<i>ERCC2</i>	Growth	Lung carcinoma
<i>UCK2</i>	Growth	Testicular carcinoma, testicular germ cell tumor
<i>CPLX2</i>	Neurobehavior, neurological, growth and metabolism, biological	ADHD, age-related nuclear cataract, facial morphology measurement, obesity-related traits

plasma membrane dynamics, particularly D3 dopamine binding.³² The protein encoded by *CPLX2* is involved in synaptic vesicle exocytosis, and it is associated with Huntington disease and schizophrenia. Overall, our findings suggest that our cry acoustic-associated epigenetic variations occurred at various genomic regions with recognized roles in cognitive impairments, growth regulation, and neurodevelopment and neurodegenerative disorders. In previous work with this sample, we found that epigenetic variation was associated with neurobehavioral profiles on the NNNS at key genes involved in neurodevelopment, including *PLA2G4E*, *TRIM9*, *GRIK3*, and *MACROD2*. Thus, acoustic cry characteristics and NNNS profiles were related to epigenetic variation in different sets of genes related to neurodevelopment, suggesting that separate molecular pathways could be involved.

The neonatal cry provides information about CNS function as it is produced by the coordination among several brain regions, including the brainstem, midbrain, and limbic system. The lower brain stem controls the muscles of the larynx, pharynx, chest, and upper neck through the vagal complex and the phrenic and thoracic nerves that are responsible for acoustic cry characteristics.^{33,34} These associations have also been found in work relating acoustic cry characteristics to developmental outcome in preterm infants.³⁵ Thus, the study of acoustic cry analysis and epigenetic variation could lead to a better understanding of the molecular underpinnings of acoustic cry characteristics in preterm and other at-risk populations. The combination of the acoustic cry analysis and DNA methylation measures could help in the identification of children who may be at greater risk for adverse developmental outcomes, which is the focus of some of our ongoing work in the NOVI cohort.

There were limitations to this investigation. First, both infant cry and adequate buccal swabs were only available for half the sample. Second, QQ plots revealed that we identify a larger number of small *p* values than would be expected, thus we focused our results and discussion only on those CpGs that were significant after Bonferroni adjustment. However, it is likely that some false positives are mixed within these results. Additional studies are required to examine genome-wide DNA methylation and cry characteristics to determine whether similar relationships are observed in independent populations or in larger samples sizes. Third, residual confounding could still exist; however, we robustly adjusted for many potential confounding variables throughout the study. Fourth, this investigation is associational and not experimental. Therefore, the associations were not hypothesized in advance and it is important to remain cautious in terms of interpretations as they relate to cellular and molecular mechanisms. Finally, there were differences in neonatal characteristics between those included and those not included. Interestingly, these differences were more favorable to the included group.

Despite these limitations, this is a novel study relating acoustic cry characteristics to variation in DNA methylation in very preterm infants. Many of these differentially methylated CpGs were within genes that are involved in multiple neurodevelopmental processes. Thus, we hypothesize that the observed differences in methylation may be correlated with neuro- and/or behavioral-developmental processes. This is the first large-scale genome-wide study of DNAm and infant cry acoustics, and we encourage future studies to examine the reproducibility of these findings in independent datasets, and/or to examine whether DNAm at these same CpGs is related to other biobehavioral characteristics.

ACKNOWLEDGEMENTS

This work was funded by National Institutes of Health (NIH)/Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grant R01HD072267 (Lester and O'Shea), R01HD084515 (Lester and Everson), and UH3OD023347 (B.M.L., C.M., and T.M.O'S.), National Center for Advancing Translational Sciences through UCLA CTSI Grant UL1TR001881 and by the National Institute of Environmental Health Sciences HERCULES Center (P30 ES019776).

AUTHOR CONTRIBUTIONS

G.A. designed the study, extracted and processed cry recordings, analyzed and interpreted data, drafted the article and revised critically for important intellectual content, and approved the final version as submitted. S.J.S. developed the acoustic cry system, reviewed and revised critically for important intellectual content, and approved the final version as submitted. C.J.M. performed the laboratory studies, interpreted methylation data, reviewed and revised critically for important intellectual content, and approved the final version as submitted. T.M.E. performed the laboratory studies, analyzed and interpreted methylation data, reviewed and revised critically for important intellectual content, and approved the final version as submitted. H.L. extracted and processed cry recordings, reviewed and revised critically for important intellectual content, and approved the final version as submitted. A.A.B. analyzed methylation data, reviewed and revised critically for important intellectual content, and approved the final version as submitted. B.S.C. reviewed and revised critically for important intellectual content and approved the final version as submitted. J.B.H. reviewed and revised critically for important intellectual content and approved the final version as submitted. J.A.H. reviewed and revised critically for important intellectual content and approved the final version as submitted. E.C.McG. reviewed and revised critically for important intellectual content and approved the final version as submitted. C.R.N. reviewed and revised critically for important intellectual content and approved the final version as submitted. T.M.O'S. conceptualized and designed the study, reviewed and revised critically for important intellectual content, and approved the final version as submitted. S.L.P. reviewed and revised critically for important intellectual content and approved the final version as submitted. L.M.S. reviewed and revised critically for important intellectual content and approved the final version as submitted. A.S. reviewed and revised critically for important intellectual content and approved the final version as submitted. L.M.D. analyzed and interpreted data, reviewed and revised critically for important

intellectual content, and approved the final version as submitted. S.A.D. coordinated data collection, reviewed and revised critically for important intellectual content, and approved the final version as submitted. Dr. J.F.P. coordinated laboratory analyses, reviewed and revised critically for important intellectual content, and approved the final version as submitted. B.M.L. conceptualized and designed the study, developed the acoustic cry system, interpreted data, drafted the article and revised critically for important intellectual content, and approved the final version as submitted.

ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41390-020-01172-0>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Statement of consent: The study was approved by Western Institutional Review Board, Puyallup, WA (HI site), the John F. Wolf Human Subjects Committee, Los Angeles, CA, Spectrum Health Systems, Inc. IRB, Grand Rapids, MI, Wake Forest University Health Sciences IRB, Winston-Salem, NC, Children's Mercy Hospital IRB, Kansas City, MO, Women & Infants Hospital IRB, Providence, RI.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Purisch, S. E. & Gyamfi-Bannerman, C. Epidemiology of preterm birth. *Semin. Perinatol.* **41**, 387–391 (2017).
- Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. in *Preterm Birth: Causes, Consequences, and Prevention*, The National Academies Collection: Reports funded by National Institutes of Health (eds Behrman, R. E. & Butler, A. S.) (National Academies Press, Washington, 2007).
- Aylward, G. P. Neurodevelopmental outcomes of infants born prematurely. *J. Dev. Behav. Pediatr.* **35**, 394–407 (2014).
- Amer, R. Neurodevelopmental outcomes of infants born at <29 weeks of gestation admitted to Canadian Neonatal Intensive Care Units based on location of birth. *J. Pediatr.* **196**, 31–37 (2018).
- Rogers, E. E. Early neurodevelopmental outcomes of extremely preterm infants. *Semin. Perinatol.* **40**, 497–509 (2017).
- LaGasse, L. L., Neal, A. R. & Lester, B. M. Assessment of infant cry: acoustic cry analysis and parental perception. *Mental Retard. Dev. Disab. Res. Rev.* **11**, 83–93 (2005).
- Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L. & Lester, B. M. Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Res.* **5**, 331–339 (2012).
- Provenzi, L. et al. Pain-related stress during the Neonatal Intensive Care Unit stay and SLC6A4 methylation in very preterm infants. *Front. Behav. Neurosci.* **9**, 99 (2015).
- Lester, B. M. et al. Neurobehavior related to epigenetic differences in preterm infants. *Epigenomics* **7**, 1123–1136 (2015).
- Tendl, K. A. et al. DNA methylation pattern of CALCA in preterm neonates with bacterial sepsis as a putative epigenetic biomarker. *Epigenetics* **8**, 1261–1267 (2013).
- Sheinkopf, S. J., Righi, G., Marsit, C. J. & Lester, B. M. Methylation of the glucocorticoid receptor (NR3C1) in placenta is associated with infant cry acoustics. *Front. Behav. Neurosci.* **10**, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4889592/> (2016).
- Everson, T. M. et al. Epigenome-wide analysis identifies genes and pathways linked to neurobehavioral variation in preterm infants. *Sci. Rep.* **9**, 6322 (2019).
- Walden, R. V. et al. Major congenital anomalies place extremely low birth weight infants at higher risk for poor growth and developmental outcomes. *Pediatrics* **120**, e1512–e1519 (2007).
- Child Mind Institute. *Hollingshead Four-Factor Index of Socioeconomic Status (SES-Child)*. Sample documentation (Nathan Kline Institute, Rockland, 2019).
- Teschendorff, A. E. et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
- Pidsley, R. et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 208 (2016).
- Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15**, R31 (2014).
- Zheng, S. C. et al. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics* **10**, 925–940 (2018).
- Reggiannini, B., Sheinkopf, S. J., Silverman, H. F., Li, X. & Lester, B. M. A flexible analysis tool for the quantitative acoustic assessment of infant cry. *J. Speech Lang. Hear. Res.* **56**, 1416–1428 (2013).
- Ruscio, J. & Roche, B. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychol. Assess.* **24**, 282–292 (2012).
- Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
- Stelzer, G. et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33 (2016).
- Koshibu, K. et al. Protein phosphatase 1 regulates the histone code for long-term memory. *J. Neurosci.* **29**, 13079–13089 (2009).
- Verdugo-Sivianes, E. M. et al. Coordinated downregulation of Spinophilin and the catalytic subunits of PP1, PPP1CA/B/C, contributes to a worse prognosis in lung cancer. *Oncotarget* **8**, 105196–105210 (2017).
- Nguyen, J. T. et al. Mammalian EAK-7 activates alternative mTOR signaling to regulate cell proliferation and migration. *Sci. Adv.* **4**, eaao5838 (2018).
- Paulsen, M. T. et al. The p53-targeting human phosphatase hCdc14A interacts with the Cdk1/cyclin B complex and is differentially expressed in human cancers. *Mol. Cancer* **5**, 25 (2006).
- Imtiaz, A. et al. CDC14A phosphatase is essential for hearing and male fertility in mouse and human. *Hum. Mol. Genet.* **27**, 780–798 (2018).
- Shen, G. et al. Overexpression of uridine-cytidine kinase 2 correlates with breast cancer progression and poor prognosis. *J. Breast Cancer* **20**, 132–141 (2017).
- Yu, S. et al. UCK2 upregulation might serve as an indicator of unfavorable prognosis of hepatocellular carcinoma. *IUBMB Life* **71**, 105–112 (2019).
- Somasundaram, R., Prasad, M., Ungerback, J. & Mikaek, S. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood* **126**, 144–152 (2015).
- Li, C., Cai, S., Wang, X. & Jiang, Z. Hypomethylation-associated up-regulation of TCF3 expression and recurrence in stage II and III colorectal cancer. *PLoS ONE* **9**, e112005 (2014).
- Basile, M. et al. Paralemmin interacts with D3 dopamine receptors: Implications for membrane localization and cAMP signaling. *Arch. Biochem. Biophys.* **446**, 60–68 (2006).
- Lester, B. M. & Boukydis, C. F. Z. No language but a cry. in *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (eds Papoušek, H., Jurgens, U. & Papoušek, M.) 41–69 (Cambridge Univ. Press, 1992).
- Lester, B. M. Developmental outcome prediction from acoustic cry analysis in term and preterm infants. *Pediatrics* **80**, 529–534 (1987).
- Lester, B. M. A biosocial model of infant crying. in *Advances in Infant Research* (eds Lipsitt, L. & Rovee-Collier, C.) 167–212 (Ablex, Norwood NJ, 1984).