Computational Gene Expression Modeling Identifies Salivary Biomarker Analysis that Predict Oral Feeding Readiness in the Newborn

Jill L. Maron, MD, MPH[1], Jooyeon S. Hwang[1], Subash Pathak, MS[2,3], Robin Ruthazer, MPH[2,3], Ruby L. Russell[4], and Gil Alterovitz, PhD[4]

Objective To combine mathematical modeling of salivary gene expression microarray data and systems biology annotation with reverse-transcription quantitative polymerase chain reaction amplification to identify (phase I) and validate (phase II) salivary biomarker analysis for the prediction of oral feeding readiness in preterm infants.

Study design Comparative whole-transcriptome microarray analysis from 12 preterm newborns pre- and post-oral feeding success was used for computational modeling and systems biology analysis to identify potential salivary transcripts associated with oral feeding success (phase I). Selected gene expression biomarkers (15 from computational modeling; 6 evidence-based; and 3 reference) were evaluated by reverse-transcription quantitative polymerase chain reaction amplification on 400 salivary samples from successful (n = 200) and unsuccessful (n = 200) oral feeders (phase II). Genes, alone and in combination, were evaluated by a multivariate analysis controlling for sex and postconceptional age (PCA) to determine the probability that newborns achieved successful oral feeding.

Results Advancing PCA ($P < .001$) and female sex ($P = .05$) positively predicted an infant's ability to feed orally. A combination of 5 genes, *neuropeptide Y2 receptor* (hunger signaling), *adneosine-monophosphate-activated protein kinase* (energy homeostasis), *plexin A1* (olfactory neurogenesis), *nephronophthisis 4* (visual behavior), and *wingless-type MMTV integration site family, member 3* (facial development), in addition to PCA and sex, demonstrated good accuracy for determining feeding success (area under the receiver operator characteristic curve = 0.78).

Conclusions We have identified objective and biologically relevant salivary biomarkers that noninvasively assess a newborn's developing brain, sensory, and facial development as they relate to oral feeding success. Understanding the mechanisms that underlie the development of oral feeding readiness through translational and computational methods may improve clinical decision making while decreasing morbidities and health care costs. *(J Pediatr 2015;166:282-8)*.

Preterm births affect an estimated 11.5% of all pregnancies in the US, resulting in medical costs exceeding $26 billion annually.[1] Before being discharged, each infant must demonstrate mature oral feeding skills in accordance to the American Academy of Pediatrics' guidelines.[2] The determination of oral feeding readiness in the preterm newborn remains a significant clinical challenge.[3] Oral feeding is a complex developmental task that requires maturation and integration of the nervous, gastrointestinal, sensory, skeletal muscular, and hypothalamic systems.[4] Disruption or delayed maturation in one or several of these developmental systems may result in choking, feeding aversion, and poor growth.[5] Furthermore, infants either born at term gestation or who correct to term postconceptional age (PCA) who cannot successfully orally feed are at increased risk for developmental disabilities.[6,7]

Because of the biological complexities of oral feeding, caregivers have been limited to subjective feeding assessment tools or "best-guess" clinical assessments to determine the feeding readiness of preterm newborns.[8-10] This, in turn, has resulted in significant feeding-associated morbidities, prolonged length of stay, and millions of dollars in health care expenditures. A recent Cochrane Review assessing the benefits of neonatal feeding assessment tools concluded that "there is no evidence to inform clinical practice," highlighting

AMPK	*Adneosine-monophosphate-activated protein kinase*	NICU	Neonatal intensive care unit
AUROC	Area under the receiver operator characteristic curve	*NPHP4*	*Nephronophthisis 4*
Ct	Cycle threshold	*NPY2R*	*Neuropeptide Y2 receptor*
CV	Coefficient of variation	PCA	Postconceptional age
DAVID	Database for Annotation, Visualization and Integrated Discovery	*PLXNA1*	*Plexin A1*
GEE	Generalized estimating equation	RT-qPCR	Reverse-transcription quantitative polymerase chain reaction
		WNT3	*Wingless-type MMTV integration site family, member 3*

From the [1]Division of Newborn Medicine, Mother Infant Research Institute at Tufts Medical Center, Floating Hospital for Children at Tufts Medical Center; [2]Research Design Center/Biostatistics Research Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center; [3]Tufts Clinical and Translational Science Institute, Tufts University; and [4]Center for Biomedical Informatics, Harvard Medical School, Boston, MA

Funded by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (K08 HD 059819-05), the Gerber Foundation, and the Richard B. Saltonstall Charitable Fund (xx [to J.M.] and NIH UL1TR001064 [to the Tufts Clinical and Translational Science Institute]). The authors declare no conflicts of interest.

http://dx.doi.org/10.1016/j.jpeds.2014.10.065

the strong need for novel approaches to assess oral feeding readiness in the preterm newborn.[11]

Transcriptomic analysis of neonatal salivary samples represents an innovative and noninvasive strategy to monitor, in real-time, the gene expression patterns of the multiple biological and developmental systems required for oral feeding readiness.[12] In this study, we combined computational modeling of gene expression microarray data and a priori systems biology knowledge with high-throughput, reverse-transcription quantitative polymerase chain reaction (RT-qPCR) amplification to identify and validate objective and biologically relevant salivary biomarkers predictive of neonatal oral feeding readiness.

Methods

This study was approved by the Tufts Medical Center Institutional Review Board, with parental consent. Both preterm and term neonates (gestational age ≥37 weeks) were recruited for this study. For the majority of enrolled subjects, PCA was based upon dating by first trimester ultrasound. In the rare instant when a first trimester assessment was not available, second trimester imaging was used to determine the age of the infant. Feeding status of infants was determined with the use of a cue-based feeding assessment tool.[13] Infants ≥32 weeks' PCA were allowed to feed if they maintained a stable cardiorespiratory status, demonstrated appropriate feeding cues, and tolerated enteral nutrition. Percent oral feeding success was calculated by dividing the volume of enteral nutrition taken orally by the total volume of enteral nutrition provided in the day. Successful oral feeders took 100% of their feeds by mouth; unsuccessful oral feeders took <100% of feeds orally. A χ^2 test was performed between successful and unsuccessful oral feeders to assess the possibility that human derived breast milk was impacting gene expression.

Salivary samples were collected with techniques developed in our laboratory and previously described.[14] Saliva was sampled approximately 1 hour after a feed to limit contamination with breast milk or formula. Samples were only collected during the day to reduce potential effects of circadian rhythms on gene expression. Saliva was immediately stabilized with 500 μL of RNAProtect saliva (QIAGEN; Venlo, Limburg, The Netherlands), vortexed, and placed at 4°C for a minimum of 48 hours before total RNA extraction with QIAGEN's RNeasy Mini Kit per manufacturer's instructions. On column DNase digestion occurred for each sample. Extracted total RNA (14 μL) was stored at −80°C awaiting downstream analysis.

Phase I: Biomarker Discovery

Two salivary samples were collected from preterm infants: one pre- and one postoral feeding success. Total RNA was amplified with the NuGEN Ovation Pico WTA system (San Carlos, California) per manufacturer's instructions. Five micrograms of amplified RNA was fragmented and biotinylated with the NuGEN Encore Biotin module before hybridization onto Affymetrix HG U133a Plus 2.0 arrays (Santa Clara, California). Each array was washed and stained in the GeneChip Fluidics Station 400, scanned with the GeneArray Scanner, and analyzed using the GeneChip Microarray Suite 5.0 (Affymetrix).

Identification of Biomarkers. Gene expression data were normalized with GenePattern before statistical analyses. Robust multiarray average and quantile normalization were applied to microarray data. Statistical analyses used MultiExperiment Viewer (www.tm4.org/mev.html). Statistical significance was set at $P < .01$. Genes that met statistical criteria and were differentially expressed pre- and postoral feeding success were enriched via Database for Annotation, Visualization and Integrated Discovery (DAVID) to further explore their biological function. DAVID is a publically available database that provides researchers with comprehensive functional annotation tools to better understand the biological relationships of lists of genes (http://david.abcc.ncifcrf.gov/home.jsp).[15,16] Three genes were cross-listed in the Bayes network: *chromosome 4 open reading frame 34* (*C4orf34*), *olfactory receptor, family 5, subfamily AK, member 4 pseudogene* (*OR5AK4P*), and *myosin regulatory light chain interacting protein* (*MYLIP*). Using 10-fold cross validation, we calculated an area under the receiver operator characteristic curve (AUROC) of 0.90 for the Bayesian network with 200 best features.

Phase II: Biomarker Validation

Development of the RT-qPCR Platform. All efforts were made to adhere to the minimum information of publication of quantitative real-time PCR experiments guidelines.[17] Genes were only considered for the platform if they were identified within the Bayesian Network or were shown previously to be associated with feeding success.[12,18,19] Each potential gene candidate had a systems biology review with Ingenuity Pathway Analysis and PubMed.org to determine its potential role in oral feeding. Candidate genes were incorporated onto the platform if they were involved in one or more of the following developmental systems: sensory integration, hypothalamic regulation of feeding, facial development, neurodevelopment, and/or gastrointestinal development. Because of budget limitations, the platform was restricted to 24 genes, inclusive of 3 reference genes. Pertinent information regarding all genes on the platform is listed in **Table I** (available at www.jpeds.com).

Extracted RNA was converted to complimentary DNA with the SuperScript VILO cDNA Synthesis kit (Life Technologies) per manufacturer's instructions. Next, we performed a selective preamplification of all gene targets on the platform with a pooled custom assay mix prepared by Life Technologies. Amplified samples were diluted 1:5 with RNase-free water before qPCR. A positive control was run on all plates. Median cycle threshold (Ct) values and CIs for the positive control reference genes were calculated to assess plate-to-plate variability. A negative control was run once to ensure that no genes amplified in the absence of