Comparison of Frozen and Unfrozen Blood Spots for Gene Expression Studies

Changshuai Wei, BS[1], Qing Lu, PhD[1], Sok Kean Khoo, PhD[2,*], Madeleine Lenski, MSPH[1], Raina N. Fichorova, MD, PhD[3], Alan Leviton, MD, MS[4], and Nigel Paneth, MD, MPH[1,5]

We studied gene expression in 9 sets of paired newborn blood spots stored for 8-10 years in either the frozen state or the unfrozen state. Fewer genes were expressed in unfrozen spots, but the average correlation coefficient for overall gene expression comparing the frozen and unfrozen state was 0.771 (95% CI, 0.700-0.828). *(J Pediatr 2014;164:189-91).*

Newborn blood spots, archived after genetic screening in at least 10 states,[1] represent a unique resource for the study of diseases with roots in the perinatal period. In Michigan, newborn spots were archived unfrozen from 1987 to 2008, but since 2009 have been archived at −20°C.

We previously showed that gene transcripts[2] and gene markers[3] can be detected in archived unfrozen blood spots, and that gene expression profiles differ in children who later develop cerebral palsy.[4] In fresh blood spots, we found nearly identical gene expression in fresh umbilical and newborn blood.[5] Here we compared microarray gene expression in paired 8- to 10-year-old blood spots collected from 9 newborns, with 1 spot stored frozen and the other spot stored unfrozen.

Methods

This study was approved at the institutional review boards at Michigan State University, Sparrow Hospital, and Brigham and Women's Hospital. With parental permission, we obtained frozen (−70°C) blood spots stored for the Extremely Low Gestational Age Newborns (ELGAN) study at Sparrow Hospital, Lansing, Michigan from 9 newborns at <28 weeks gestational age at birth in 2002-2004. We also retrieved these newborns' state-archived blood spots, obtained around the same age of 1-3 days but stored unfrozen. The ELGAN spots were kept unfrozen for days to weeks before being sent in batches to the ELGAN laboratories. For this study, the 9 frozen spots were sent in dry ice to the Laboratory of Microarray Technology at Van Andel Research Institute, where they were stored at −70°C until RNA extraction. The state-archived blood spots were stored at ambient temperature until RNA extraction.

Total RNA was extracted, purified, and concentrated from three 3-mm punches using a glass-fiber filter and then subjected to deoxyribonuclease treatment. Average messenger RNA yield was higher from frozen spots compared with unfrozen spots (73.4 ng vs 22.8 ng), but the range was very wide (5.9-113.3 ng). RNA integrity was equivalent in the 2 spots (unfrozen, 2.3; frozen, 2.2).

The Ovation Pico RNA amplification system (NuGEN Technologies, San Carlos, California) generated single-stranded complementary DNA, which was purified and labeled with Alexa Fluor 3 fluorescent dye before hybridization onto the Agilent whole human genome gene expression 8 × 60K microarray (Agilent Technologies, Santa Clara, California), which contains 60 000 oligonucleotide probes (60-bp probe) covering 27 958 Entrez Gene RNAs and 7419 long intergenic noncoding RNAs. Arrays were hybridized (with the same amount of labeled complementary DNA from all samples) for 17 hours at 65°C and a 10-rpm rotation speed, washed for 2 min each with wash buffers 1 and 2, and scanned with an Agilent G3 high-resolution scanner. Probe features were extracted from the microarray scan data using Feature Extraction version 10.7.3.1 (Agilent Technologies). All paired specimens were run on the same array.

After filtering and quantile normalization,[6,7] we aggregated the expression signal at the gene level for each of the 18 specimens. We enumerated the genes expressed, and calculated the correlation coefficients within pairs (frozen vs unfrozen) both for the whole genome and for 3 gene sets associated with cerebral palsy, designated inflammatory (n of genes = 36), hypoxic (n = 127), and thyroidal (n = 140).[4] The summary correlation coefficients and their 95% CIs were estimated using a random-effects model.[8,9] For correlation coefficients, we obtained a *z*-score and variance for each pair using a Fisher transformation, summarized the *z*-scores using a random-effects model, and transformed these scores to the final point and interval estimates shown.

Results

On average, 15 864 genes were expressed in the frozen samples and 11 897 genes were expressed in the unfrozen samples,

ELGAN	Extremely Low Gestational Age Newborns

From the [1]Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI; [2]Laboratory of Microarray Technology, Van Andel Research Institute, Grand Rapids, MI; [3]Laboratory of Genital Tract Biology, Department of Obstetrics, Gynecology, and Reproductive Biology, Harvard Medical School and Brigham and Women's Hospital; [4]Neuroepidemiology Unit, Department of Neurology, Boston Children's Hospital, Harvard University, Boston, MA; and [5]Department of Pediatrics and Human Development, College of Human Medicine, Michigan State University, East Lansing, MI

*Present address: Department of Cell and Molecular Biology, Grand Valley State University, Grand Rapids, MI

Funded by the National Institutes of Health (R01 NS055101 to N.P.). The frozen blood spot specimens were collected under the auspices of the ELGAN study (U01NS040069 to A.L.). The unfrozen blood spot specimens were provided by the Michigan Department of Community Health. The authors declare no conflicts of interest.

 http://dx.doi.org/10.1016/j.jpeds.2013.09.025

representing 75% and 56%, respectively, of all 22 266 genes in the array (Table I). Of the genes expressed in the frozen samples, 69% (95% CI, 57%-81%) were detected in the unfrozen sample. A larger fraction of genes expressed in the unfrozen samples was expressed in the frozen samples (93%; 95% CI, 90%-95%). Variability in the number of genes expressed was small in the frozen spots (range, 14 589-16 744) and greater in the unfrozen spots (range, 7017-17 911). Two infants (nos. 6 and 8) had relatively few genes expressed in their unfrozen samples (7017 and 7454), but the number of genes expressed in their frozen samples (15 984 and 15 829) was nearly identical to the mean of all frozen samples (15 864).

The average pairwise correlation of all genes expressed in both the frozen and unfrozen samples was 0.771 (95% CI, 0.700-0.828) (Table II). In the 2 infants with low gene expression, correlations between the frozen and unfrozen samples were lower (0.674 and 0.635). The mean difference in expression level in the paired samples, expressed as log2-fold change (unfrozen minus frozen), was not significantly different from zero (−0.485; 95% CI, −1.175 to 0.205) (Table III; available at www.jpeds.com).

Only 4316 genes were expressed in all 18 specimens (available upon request from authors), indicating considerably heterogeneity in genes expressed in infants. The average pairwise correlation coefficient for these genes was 0.777 (95% CI, 0.712-0.828).

Findings in the 3 gene sets of interest paralleled those for the whole genome. Of the 301 genes in these sets, 263 (87%) were expressed in the frozen sample and 189 (62%) were expressed in the unfrozen sample. Of the genes expressed in frozen samples, 70% (95% CI, 69%-83%) were detectable in unfrozen samples. The within-pair correlation coefficients for expression of the gene sets of interest were nearly identical to those found for the whole genome (0.768; 95% CI, 0.686-0.830) (Table II). As in the whole genome analysis, the expression levels for the 3 gene sets within pairs (unfrozen minus frozen) did not differ significantly (−0.56; 95% CI, −1.222 to 0.097) (Table III). The 2 infants whose unfrozen spots showed fewer genes overall also had fewer genes expressed in the gene sets of interest, as well as lower correlations between the 2 specimen types (Table II).

Discussion

After 8-10 years of storage, 30% of genes detectable in blood spots at −70°C were not detectable in unfrozen spots prepared at the same time; however, the genes that were expressed were very highly correlated in frozen and unfrozen spots, indicating that unfrozen spots generally provide a reasonable approximation of the frozen state for both all genes expressed and selected gene sets of interest. In specimens with greater loss of gene signals, the degree of approximation was reduced. Based on our data, we cannot know the relative value of storage at −20°C compared with −70°C, or whether these findings extend beyond very premature infants, although morbidity- or gestational age–dependent regulation of messenger RNA decay or

Table I. Number and percent of genes expressed in frozen and unfrozen specimens

Pair number	Genes expressed in unfrozen sample, n	Proportion of all 21 266 genes in array, %	Genes expressed in frozen sample, n	Proportion of all 21 266 genes in array, %	Proportion of all genes expressed in frozen also expressed in unfrozen, %	Proportion of all genes expressed in unfrozen also expressed in frozen, %
1	13 009	61.2	15 786	74.2	76.3	92.6
2	10 277	48.3	15 588	73.3	62.0	94.1
3	12 929	60.8	14 589	68.6	81.9	92.4
4	11 640	54.7	17 372	81.7	64.9	96.9
5	11 550	54.3	15 285	71.9	68.8	91.0
6	7454	35.1	15 984	75.2	43.9	94.1
7	15 287	71.9	15 606	73.4	88.8	90.7
8	7017	33.0	15 829	74.4	42.7	96.4
9	17 911	84.2	16 744	78.7	94.0	87.9
Column mean	11 897	55.9	15 865	74.6	69.3	92.9