Molecular Genetics and Metabolism 119 (2016) 207–213

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Effect of storage time on gene expression data acquired from unfrozen archived newborn blood spots

Nhan T. Ho [a,1], Julia V. Busik [b], James H. Resau [c], Nigel Paneth [a,d], Sok Kean Khoo [e,*]

[a] Department of Epidemiology & Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI, USA
[b] Department of Physiology, Michigan State University, East Lansing, MI, USA
[c] Graduate School, Van Andel Research Institute, Grand Rapids, MI, USA
[d] Department of Pediatrics & Human Development, College of Human Medicine, Michigan State University, East Lansing, MI, USA
[e] Department of Cell and Molecular Biology, Grand Valley State University, Allendale, MI, USA

ARTICLE INFO

Article history:
Received 10 August 2016
Accepted 13 August 2016
Available online 18 August 2016

Keywords:
mRNA gene expression
quantitative real-time PCR
microarray
newborn blood spots
storage time
ambient temperature

ABSTRACT

Unfrozen archived newborn blood spots (NBS) have been shown to retain sufficient messenger RNA (mRNA) for gene expression profiling. However, the effect of storage time at ambient temperature for NBS samples in relation to the quality of gene expression data is relatively unknown. Here, we evaluated mRNA expression from quantitative real-time PCR (qRT-PCR) and microarray data obtained from NBS samples stored at ambient temperature to determine the effect of storage time on the quality of gene expression. These data were generated in a previous case-control study examining NBS in 53 children with cerebral palsy (CP) and 53 matched controls. NBS sample storage period ranged from 3 to 16 years at ambient temperature. We found persistently low RNA integrity numbers (RIN = 2.3 ± 0.71) and 28S/18S rRNA ratios (~0) across NBS samples for all storage periods. In both qRT-PCR and microarray data, the expression of three common housekeeping genes—beta cytoskeletal actin (*ACTB*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), and peptidylprolyl isomerase A (*PPIA*)—decreased with increased storage time. Median values of each microarray probe intensity at $\log_2$ scale also decreased over time. After eight years of storage, probe intensity values were largely reduced to background intensity levels. Of 21,500 genes tested, 89% significantly decreased in signal intensity, with 13,551, 10,730, and 9925 genes detected within 5 years, > 5 to <10 years, and >10 years of storage, respectively. We also examined the expression of two gender-specific genes (X inactivation-specific transcript, *XIST* and lysine-specific demethylase 5D, *KDM5D*) and seven gene sets representing the inflammatory, hypoxic, coagulative, and thyroidal pathways hypothesized to be related to CP risk to determine the effect of storage time on the detection of these biologically relevant genes. We found the gender-specific genes and CP-related gene sets detectable in all storage periods, but exhibited differential expression (between male vs. female or CP vs. control) only within the first six years of storage. We concluded that gene expression data quality deteriorates in unfrozen archived NBS over time and that differential gene expression profiling and analysis is recommended for those NBS samples collected and stored within six years at ambient temperature.

1. Introduction

Newborn blood spotted on filter paper is used throughout the United States and in most industrialized countries for newborn genetic screening. Archives of leftover newborn blood spot (NBS) samples are available in many states [1]. Even after many years of unfrozen storage, mRNA can be extracted from NBS [2–4] and analysis of mRNA phenotype can distinguish individuals with different thalassemia genotypes using archived NBS samples [5]. In our previous studies, we showed that mRNA extracted from unfrozen archived NBS can be applied on microarrays to perform genome-wide gene expression profiling, even though the yield of RNA declined over storage time of NBS samples [6]. We also showed that differential expression could be detected between genders [7] and between cerebral palsy (CP) cases and matched controls [8]. Thus, archived NBS samples can serve as an essential research resource to examine gene expression profiles in the newborn period and possible relation to later child health.

NBS collection differs from most human blood collection methods in that no collecting tube is used, but blood is spotted directly, usually from a heel prick, onto filter paper and dries in minutes. The absence of a liquid environment appears to reduce the activity of ribonucleases and micro RNAs that degrade mRNA, and most mRNA species within the cell are quickly degraded after being transcribed. Some loss of mRNA

* Corresponding author at: Department of Cell and Molecular Biology, Grand Valley State University, 1 Campus Drive, Allendale, MI 49401, USA.
E-mail address: khoos@gvsu.edu (S.K. Khoo).
[1] Present address: Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam.

http://dx.doi.org/10.1016/j.ymgme.2016.08.001
1096-7192/

is likely, especially in older blood spot samples, and this may affect the quality of data such as gene expression profiles obtained from these samples. mRNA degradation is especially possible if blood spots are stored at ambient conditions [9].

A commonly used method for assessing RNA quality is the RNA integrity number (RIN), which ranges from 1 (indicating severe mRNA degradation) to 10 (indicating high quality, largely intact mRNA) [10]. Another common approach for RNA quality assessment is examining the 28S/18S ribosomal RNA (rRNA) ratio. This approach assumes that rRNA quality and quantity reflect that of the underlying mRNA population and a ratio of 2:1 is considered the benchmark for intact RNA [11, 12].

Although RIN and 28S/18S rRNA ratios are useful proxies for assessing RNA quality, they are indirect measures of mRNA and do not directly describe the quality and quantity of mRNA. Low scores on these measurements do not necessarily preclude the detection of biologically-relevant, differentially-expressed genes between experimental and control groups.

In this study, we addressed two topics related to assessing the quality of mRNA microarray data obtained from unfrozen archived NBS samples. First, we evaluated whether mRNA microarray data quality reflects the yield of mRNA or level of mRNA degradation with respect to NBS storage period. Second, we assessed the detectability of differentially-expressed genes between genders and CP vs. controls from mRNA microarray data, also based on NBS storage period. Our study will provide the optimal cutoff for a recommended ambient storage period of NBS samples to maximize the benefit of using these samples for microarray gene expression profiling studies.

2. Materials and methods

2.1. Study subjects

Our mRNA microarray dataset was acquired from unfrozen archived NBS of 53 singleton CP cases and 53 matched controls (similar year of birth, sex and gestational age) included in our previous case-control study to investigate the molecular etiology of CP [8]. Among the 106 study subjects, 31 were females and 75 were males, with an NBS storage period ranging from 2.9–16 years. In 2009, Michigan Department of Community Health (MDCH) began storing all newly collected NBS at −20 °C. However, our study subjects were born within 1994–2007 and NBS samples were retrieved from the archive in 2008 (stored at ambient temperature without climate control). These NBS samples were retrieved from MDCH upon receipt of written informed consent from parents or guardians. This project was approved by the institutional review boards of all participating institutions.

2.2. RNA isolation, microarray and qRT-PCR assays

Total RNA was extracted from three 3 mm punches of NBS using the illustra RNAspin Mini RNA kit (GE Healthcare). Due to large sample size, all spots were not extracted at the same time but in batches of 24. Extracted mRNA was treated the same way for all spots prior to analysis. To minimize variables, all assays were performed by a single technician and extracted mRNA was immediately stored at −80 °C until the next procedure. The Agilent Whole Human Genome Gene Expression 8 × 60 K Microarray platform was used for microarray assays. Quantitative real-time PCR (qRT-PCR) was performed using Applied Biosystems 7500 Fast Real-Time PCR System. Details regarding sample processing, RNA isolation, microarray and qRT-PCR techniques were published in our previous study [8].

2.3. Assessment of mRNA microarray data quality

We examined the following four features to assess mRNA microarray data:

(1) Distribution of RIN and 28S/18S rRNA ratios across storage periods of NBS samples.
(2) Overall distribution of raw fluorescent probe intensity data across storage periods of NBS samples, illustrated by brightness of raw digital images as well as density and scatter plots of $\log_2$ signal intensity.
(3) Linear regression slopes of $\log_2$ signal intensity across storage periods of NBS using raw aggregated gene expression data for all genes available in the microarrays.
(4) Number of aggregated genes being filtered and number of remaining genes after filtering, in correspondence with storage period of NBS samples.

We employed the commonly used approach for filtering unqualified spots of microarray data by Paterson *et al.* in which probe intensity was removed when the gProcessed signal was less than two times the gProcessed signal error [13]. We applied quantile normalization on filtered microarray data with a slight modification, which was stratifying the storage period of NBS samples. Probe intensity data was then aggregated to a gene level using the average values of $\log_2$ signal intensity data for each gene with multiple probes. R limma package [14] was used for microarray data processing.

2.4. Assessment of housekeeping gene expression based on NBS storage period

To further evaluate the pattern of mRNA expression signals over storage period of NBS samples, we examined the signal intensities of three commonly used housekeeping genes—peptidylprolyl isomerase A (*PPIA*), beta cytoskeletal actin (*ACTB*), and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*)—in microarray and qRT-PCR data.

2.5. Evaluation on the power of detecting differentially-expressed genes in NBS samples stored across 3–16 years at ambient temperature

We also assessed the impact of storage period on the power of detecting differentially-expressed genes in NBS samples. First, we examined the effect of storage period on detecting two sex-linked genes—X inactivation-specific transcript, *XIST* and lysine-specific demethylase 5D, *KDM5D*—that are known to be differentially-regulated between male and female samples [7]. A linear model for microarray data implemented in the R limma package [14] was used to analyze differential expression of these two genes. Second, we examined the effect of storage period on detecting seven gene sets—four empirical and three canonical gene sets representing the inflammatory, hypoxic, coagulative, and thyroidal pathways—hypothesized to be differentially-expressed in newborns who later developed CP or without CP. GAGE (generally applicable gene set enrichment for pathway analysis) [15] was applied for gene set statistical analysis, as described in our previous study [8]. GAGE t-statistics of each matched case-control pair were standardized by converting to the equivalent z-statistics for empirical inflammatory, hypoxic, and thyroidal gene sets, using Stouffer's method [16], which showed significant differential expression between CP and controls in our previous study. The absolute values of these converted z-statistics were then plotted against storage period of NBS samples for each case-control pair. All analyses were performed using R version 2.13.2.

3. Results

3.1. mRNA microarray and qRT-PCR data quality in relation with storage period of NBS samples

3.1.1. Variation in RIN and 28S/18S rRNA ratio

Average RIN in 106 NBS samples was 2.3 ± 0.71 and RIN values did not vary significantly over storage period of 3–16 years at ambient