Mining the Archives: A Cross-Platform Analysis of Gene Expression Profiles in Archival Formalin-Fixed Paraffin-Embedded Tissues

A. Francina Webster*,†, Paul Zumbo‡, Jennifer Fostel§, Jorge Gandara‡, Susan D. Hester¶, Leslie Recio‖, Andrew Williams*, Charles E. Wood¶, Carole L. Yauk*,1,2, and Christopher E. Mason‡,‖‖,‖‖‖,1,2

*Environmental Health Science and Research Bureau, Health Canada, Ottawa K1A 0K9, Canada; †Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa K1S 5B6, Canada; ‡Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York 10065; §National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709; ¶Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, North Carolina 27709; ‖ILS, Inc., PO Box 13501, Research Triangle Park, North Carolina 27709; ‖‖The Feil Family Brain and Mind Research Institute (BMRI), 413 East 69th Street, New York, New York 10021; and ‖‖‖The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, 1305 York Avenue, New York, New York 10065

[1]To whom correspondence should be addressed at Carole L. Yauk, Environmental Health Science and Research Bureau, ERHSD, HECSB, Health Canada, Tunney's Pasture, Bldg. 8 (P/L 0803A), 50 Colombine Driveway, Ottawa, Ontario K1A 0K9, Canada. E-mail: Carole.Yauk@hc-sc.gc.ca; Christopher E. Mason, Weill Cornell Medical College, 1305 York Ave., 13th floor, Rm. Y13-04, Box 140, New York, NY 10021, USA. E-mail: chm2042@med.cornell.edu
[2]These authors contributed equally to this study.

ABSTRACT

Formalin-fixed paraffin-embedded (FFPE) tissue samples represent a potentially invaluable resource for transcriptomic research. However, use of FFPE samples in genomic studies has been limited by technical challenges resulting from nucleic acid degradation. Here we evaluated gene expression profiles derived from fresh-frozen (FRO) and FFPE mouse liver tissues preserved in formalin for different amounts of time using 2 DNA microarray protocols and 2 whole-transcriptome sequencing (RNA-seq) library preparation methodologies. The ribo-depletion protocol outperformed the other methods by having the highest correlations of differentially expressed genes (DEGs), and best overlap of pathways, between FRO and FFPE groups. The effect of sample time in formalin (18 h or 3 weeks) on gene expression profiles indicated that test article treatment, not preservation method, was the main driver of gene expression profiles. Meta- and pathway analyses indicated that biological responses were generally consistent for 18 h and 3 week FFPE samples compared with FRO samples. However, clear erosion of signal intensity with time in formalin was evident, and DEG numbers differed by platform and preservation method. Lastly, we investigated the effect of time in paraffin on genomic profiles. Ribo-depletion RNA-seq analysis of 8-, 19-, and 26-year-old control blocks resulted in comparable quality metrics, including expected distributions of mapped reads to exonic, untranslated region, intronic, and ribosomal fractions of the transcriptome. Overall, our results indicate that FFPE samples are appropriate for use in genomic studies in which frozen samples are not available, and that ribo-depletion RNA-seq is the preferred method for this type of analysis in archival and long-aged FFPE samples.

Tissue repositories have immense untapped potential in translational research. Applications of archival tissue resources range from biomarker discovery and toxicogenomic profiling to large-scale studies of genome-disease interactions. Curated sample banks house diverse tissue types from different models or study populations often linked with detailed clinical or pathologic outcomes. In many cases archives may contain unique samples from animal bioassays, clinical trials, or epidemiologic studies that would be difficult or impossible to repeat. Despite this promise, direct use of archival samples for transcriptomic profiling has been relatively limited to date. The majority of archival tissues are stored in formalin-fixed paraffin-embedded (FFPE) blocks, which preserve tissue architecture for histopathological analysis and allow for tissue storage at room temperature. However, formalin treatment degrades RNA through cross-linking and fragmentation, which significantly impairs molecular analyses (Bass *et al.*, 2014; Farragher *et al.*, 2008; Klopfleisch *et al.*, 2011). These RNA effects can result in inconsistent genomic data and present important technical and analytical challenges when working with FFPE samples. Nevertheless, studies have demonstrated the value of using microarrays to analyze FFPE samples for understanding the molecular basis of a variety of cancers (Budczies *et al.*, 2011; Jacobson *et al.*, 2011; Linton *et al.*, 2012; Sadi *et al.*, 2011). As technologies continue to improve, so too does the prospect of FFPE genomics.

Recently, whole-transcriptome RNA-sequencing (RNA-seq) methodologies have been developed as a more precise and less biased measurement of transcript levels that may overcome issues associated with highly fragmented or degraded RNA. When compared with conventional DNA microarrays, RNA-seq enriches for many additional fragments as it is not restricted to predefined probes and has (in principle) no limitations to dynamic range (Li *et al.*, 2014b; SEQC/MAQC-III Consortium, 2014). Previous work has shown concordance between gene expression profiles produced by microarrays and by RNA-seq for fresh-frozen (FRO) and FFPE samples (Malone and Oliver, 2011; Marioni *et al.*, 2008; Wang *et al.*, 2014). Recent studies investigating the use of RNA-seq in FFPE samples have also shown promising results (Auerbach *et al.*, 2014; Hedegaard *et al.*, 2014; Zhao *et al.*, 2014a). To extend these findings, the Health and Environmental Sciences Institute (HESI) Committee on Application of Genomics to Mechanism-Based Risk Assessment coordinated a working group to assess the utility of various gene expression profiling methods for FFPE tissues. The main goals of this group were to evaluate current RNA-seq and microarray methods in paired FRO and FFPE samples, and investigate different preanalytical and experimental factors influencing FFPE results.

The reference compound for this study was furan, which is a known liver carcinogen in rodents (IARC, 1995; Moser *et al.*, 2009; NTP, 1993). Typically, humans are exposed to furan by consuming heat-treated foods or inhaling combustion emissions (IARC, 1995; Moro *et al.*, 2012). Furan is metabolised by cytochrome P450 2E1 (CYP2E1) into an electrophilic and cytotoxic metabolite, cis-2-butene-1,4-dial (BDA) (Kedderis *et al.*, 1993; Kedderis and Held, 1996). The carcinogenicity of furan has been attributed to BDA-induced cytotoxicity followed by dysregulated regenerative proliferation in the liver (Fransson-Steen *et al.*, 1997; Moser *et al.*, 2009). We recently confirmed this mode of action using DNA microarray analysis of FRO livers taken from mice that were sub-chronically exposed to both carcinogenic and non-carcinogenic doses of furan alongside solvent-exposed controls (Jackson *et al.*, 2014). We proposed that Cyp2E1 protein stabilization and metabolism of furan would lead to chronic activation of the Nrf2 Oxidative Stress Response pathway, which is a primary driver in furan-induced carcinogenic transformation. Here we extend this study by investigating whether the same furan-dependent molecular changes can be detected in paired FFPE samples.

In this study, we compared gene expression profiles from paired FRO and FFPE-preserved liver tissues from female B6C3F1 mice that were exposed for 3 weeks to a carcinogenic dose of furan alongside treatment controls. We investigated the effect of tissue time in formalin on the quality of transcriptomic data given that different institutions have different time in formalin protocols. For this aim, we used paired samples that had been preserved in formalin for either 18 h or 3 weeks (Supplementary Fig. S1 for experimental design). For all of these samples, gene expression profiling was completed using 4 protocols: Illumina Hi-seq RNA-seq (ribo-depletion and poly-A-enrichment formats) and Agilent SurePrint G3 Mouse GE $8 \times 60\,K$ microarrays (1- and 2-color formats). Both 1- and 2-color microarrays were included in the analysis because they are conventional approaches used in many laboratories and serve as important points of reference. Finally, to better understand the effects of time in paraffin, we used the RNA-seq ribo-depletion protocol to evaluate rat liver samples that were obtained from the National Toxicology Program (NTP) archives and had been in storage for up to 26 years. Our results provide a comprehensive comparison of genomics methodologies for FFPE samples and recommendations for the most appropriate approach for the study of archival specimens.

MATERIALS AND METHODS

Frozen and FFPE Tissue Samples

Frozen and FFPE samples for the microarray/RNA-seq comparison and time in formalin analysis were obtained from a recent study of female mice treated with furan (CAS No. 110-00-9) (>99% pure) (Sigma-Aldrich Chemical Co., Milwaukee, Wisconsin). Doses of 0 or 8 mg/kg bw furan dissolved in Mazola corn oil were prepared as previously described (Jackson *et al.*, 2014; Recio *et al.*, 2013). Mice were housed, dosed, and sacrificed as previously described (Jackson *et al.*, 2014; Recio *et al.*, 2013). Briefly, 5–6 week-old specific-pathogen-free (SPF) B6C3F1 mice were purchased from Charles River Breeding Laboratories (Portage, Maine), and housed in a SPF and Association for Assessment and Accreditation of Laboratory Animal Care-accredited facility. All experiments were conducted in compliance with the Animal Welfare Act Regulations (9CFR1–4). Mice were handled and treated according to the guidelines provided in the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals (ILAR, 1996; http://dels.nas.edu/ilar/). Mice ($n = 4$ per group) were dosed daily by oral gavage for 21 days. Four hours after the final dose, animals were anesthetized by CO_2 inhalation, and euthanized by exsanguination. Livers were removed and trimmed, and portions were either flash-frozen or fixed in a standard 10% solution of neutral phosphate-buffered formalin. Frozen samples were stored at or below $-70°C$. Fixed samples were kept in formalin for 18 h or 3 weeks,

and then transferred to 100% ethanol. After approximately 72 h, tissues were processed and embedded in paraffin using standard histologic procedures. Please see Supplementary Figure S1, which summarizes the experimental design. FFPE tissue blocks were stored at room temperature for < 1 year before RNA was isolated for this study.

For the time in paraffin analysis, FFPE blocks of control liver tissue were obtained from the NTP archive collection. The corresponding slides from each block were re-evaluated on histopathology to confirm that the selected tissue was morphologically normal. Samples were obtained from 3 different studies in control F344 rats in which tissues had been stored as FFPE blocks for different lengths of time: (1) 8 years (13-week control rats for an experiment investigating beta-picoline exposure); (2) 19 years (2-year controls for oxazepam exposure); and (3) 26 years (13-week controls for trichlorfon exposure). The difference in study length affects both the age of the animal at sacrifice and also the allowable time in formalin. According to NTP study protocol, for a 13-week study the tissues may only stay in formalin for up to 6 weeks, while for a 2-year study fixed tissues are permitted to remain in formalin for up to 6 months. Recently collected rat liver was used as FRO control.

RNA Extraction and cDNA Synthesis

Total RNA was extracted from FRO tissues using the RNeasy Midi RNA Extraction kit (Qiagen, Mississauga, Ontario, Canada) as described previously (Jackson et al., 2014, 2013). Total RNA was extracted from slide-mounted furan and archival FFPE tissues using the FFPE RNeasy kit (Qiagen). One 10 μm section was used per FFPE extraction per animal for the 18 h in formalin group. Due to low yields of RNA, two 10 μm sections were used per FFPE extraction per animal for the 3 weeks in formalin group. RNA extracted from FFPE tissues was assessed for integrity using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc, Mississauga, Ontario, Canada), quantitated using a NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc, Wilmington, Delaware), and stored at −80°C (Supplementary Table S1). For cDNA synthesis, RNA (250 ng per sample) was reverse transcribed and amplified using the Whole Transcriptome Amplification kit 2 (WTA2; Sigma, St Louis, Missouri). cDNA was purified using the QiaQuick PCR clean-up kit (Qiagen).

Microarray Analysis

Microarray analysis for the FRO tissues was performed using 1- and 2-color protocols (all microarray experiments were conducted in the Yauk laboratory, Ottawa, Canada). The 1-color FRO protocol was carried out according to the One-Color Microarray-Based Gene Expression Analysis, Low Input Quick Amp Labeling Protocol, version 6.6 (Agilent Technologies, 2012). Briefly, using the RNA Spike-In Kit, One-Color (Agilent Technologies) and the Low Input Quick Amp Labeling Kit, One-Color (Agilent Technologies), 200 ng of RNA from each sample was used to synthesize, amplify, and Cy3-label cRNA. We labeled 4 samples each for the FRO, 18 h in formalin FFPE, and 3 weeks in formalin FFPE (control animals 2, 3, 4, 6, and treated animals 41, 42, 43, 44); RNA from animal 3 in the FFPE-3 weeks group did not label so the number of biological replicates for the FFPE-3 weeks group was reduced to 3 (we included an additional technical replicate of sample 2 in this group). Labeled cRNA was purified using the RNeasy Mini Kit (Qiagen) and quantified using a NanoDrop Spectrophotometer. Hybridization mixes were prepared using the Gene Expression Hybridization Kit (Agilent Technologies). Cy3-labeled cRNA (600 ng) was hybridized to

SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent Technologies) at 65°C for 17 h at 10 rpm.

The 2-color FRO protocol has been previously described (Jackson et al., 2014). Briefly, using a reference design, 200 ng of sample RNA and 200 ng of mouse universal reference RNA (Stratagene, Agilent Technologies Inc) were labeled with Cy5 and Cy3, respectively, using the Low Input Quick Amp Labeling Kit (Agilent Technologies). Labeled cRNA was purified and quantified as described earlier. cRNA (300 ng per sample) and reference cRNA were hybridized to SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent Technologies) at 65°C for 17 h at 10 rpm.

Microarray analysis for FFPE tissues was performed according to our published 1- and 2-color cDNA input protocols (Jackson et al., 2013). For the 1-color FFPE protocol, 1650 ng of cDNA was labeled with 2.48 μl Cy3 dye using the Genomic DNA ULS Labeling Kit (Agilent Technologies). For the 2-color FFPE protocol, 825 ng of cDNA was labeled with 1.44 μl Cy5 dye, and 825 ng of reference cDNA (made from a pool of all sample cDNA) was labeled with 1.24 μl Cy3 using the Genomic DNA ULS Labeling Kit (Agilent Technologies). In both protocols, excess dye was removed using KREApure columns (Agilent Technologies), labeled product was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific Inc), and degree of labeling was calculated using the following calculation: degree of labeling = $[(340 \times \text{pmol}/\mu l \ \text{dye})/(\text{ng}/\mu l \ \text{cDNA} \times 1000)] \times 100\%$ (with an acceptable range of 1.5–3.0%).

Hybridization mixtures were prepared using the Gene Expression Hybridization kit (Agilent Technologies) and CGHBlock (Agilent Technologies). For the 1-color FFPE protocol, 600 ng of Cy3-labeled cDNA was hybridized to SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent Technologies). For the 2-color FFPE protocol, 300 ng each of sample and reference cDNA was hybridized to SurePrint G3 Mouse GE 8 × 60 K microarrays (Agilent Technologies). Hybridization for both protocols occurred in a randomized block design, at 65°C for 17 h at 20 rpm in the dark in an Agilent SureHyb hybridization chamber.

All slides were washed with Wash Buffers 1 and 2 (Agilent Technologies) and scanned at 5 μm resolution on an Agilent G2505B scanner. Feature extraction was accomplished using Agilent feature extraction software version 11. The complete microarray dataset is available through the Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo/) accession number: GSE62843.

Microarray Statistical Analysis

One-Color. Non-background subtracted median signal intensities were cyclic-LOWESS normalized (Bolstad et al., 2003) in R (R-Core-Development-Team, 2012) using the Affy library (Gautier et al., 2004). Differential gene expression analysis between the control and treated samples was applied using the MAANOVA library (Wu et al., 2003) in R. The Fs statistic (Cui et al., 2005) was used to test for differential expression. The P-values from these tests were estimated using the permutation method with residual shuffling, and adjusted for multiple comparisons by using the false discovery rate (FDR) approach (Benjamini and Hochberg, 1995). Least square means (Goodnight and Harvey, 1978; Searle et al., 1980) were used to estimate the fold change for each pairwise comparison tested.

Two-Color. A reference design (Kerr, 2003) was employed for the 2-color analysis. Non-background subtracted median signal intensities were LOWESS-normalized (Yang et al., 2002) in R

(http://www.R-project.org/) using the MAANOVA library (Wu et al., 2003). The MAANOVA library was used for the DEG analysis using the log2 of the relative intensities. The P-values were estimated using the permutation method with residual shuffling, and FDR-adjusted P-values were computed using the FDR approach. Fold change estimates were computed as in the 1-color analysis.

Correlation analysis. Correlations in Figure 1 were determined using the regression function in the Microsoft Excel 'Data Analysis' tool pack.

Microarray Bioinformatic Analysis

Cluster analysis. The universal reference used in the 2-color FRO data was different than the pooled sample reference used in the FFPE 3-color analyses. For the cluster analysis, the data from the 2- and 1-color experiments were normalized to the median of the time-matched controls. By applying this normalization, the FRO and FFPE data could then be analyzed jointly. Heatmaps were constructed using a similarity metric based on 1-correlation. The correlations were estimated using the Spearman approach. Genes were considered significant with a 2-fold change and an FDR $P < 0.05$. The row dendrogram for the heatmaps was estimated by pooling the 1- and 2-color data in order to have the same row order for the 2 plots to help facilitate comparison of the 2 heatmaps.

Pathway analysis. Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com/) was used to identify enriched molecular pathways. Gene expression data were analyzed using an IPA Core Analysis with a gene expression threshold of fold change $\geq \pm 1.5$ and FDR $P \leq 0.05$. Enrichment of IPA canonical pathways was determined based on the number of DEGs in the dataset that also appear in that pathway. Pathway significance thresholds were determined in IPA using a right-tailed Fisher's exact test. Only mapped DEGs were considered for Figure 2 (unmapped DEGs in Supplementary Fig. S2).

NextBio. To identify chemical effectors and disease states that produce changes to gene expression that are similar to those produced following exposure to furan, gene expression profiles from all microarray datasets were mined against large genomic database repositories using NextBio (www.nextbio.com). NextBio applies proprietary algorithms that use pairwise gene signature correlations and rank-based enrichment statistics to produce scores, where the chemical or disease with the highest similarity is given a score of 100 and the rest are normalized accordingly (Kupershmidt et al., 2010). The meta-analyses included up to the top 250 DEGs (mapped) from each group with an FDR $P < 0.05$.

RNA-Seq Methods

RNA-seq was performed on the FRO and FFPE tissues using 2 different protocols: ribosomal RNA-depleted-sequencing (rmRNA-seq; conducted in the Mason laboratory, NYC, USA) and poly-A-selected RNA-seq (mRNA-seq; conducted at Genome Quebec, Montreal, Canada). The rmRNA-seq and mRNA-seq protocols were carried out according to Illumina's TruSeq Total RNA Sample Prep Kit and Illumina's TruSeq RNA Sample Preparation Kit v2, respectively. For rmRNA-seq library construction, ribosomal RNA was removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads (Human/Mouse/Rat). For mRNA-seq library construction, poly-A-containing mRNA was purified using oligo-dT-attached

magnetic beads. Following purification, rRNA-depleted or poly-A-enriched RNA was chemically fragmented into small pieces using divalent cations under elevated temperature. For all FFPE samples, we used double the input of FRO samples. The RNA fragments were copied into first strand cDNA using reverse transcriptase and random hexamer primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. Adapters were subsequently ligated to the cDNA fragments, and then purified and enriched for with PCR to create a final cDNA library. The final cDNA libraries were single-end sequenced on a HiSeq2000. Ribo-depletion experiments had a read length of 42 bases and a read depth of 10–60 M reads. Poly-A-enrichment experiments had a read length of 50 bases and a read depth of 3.8–11 M reads. It has previously been reported that the required sequencing depth to be microarray-equivalent is 5 M (Black et al. 2014). The complete dataset is available through the Gene Expression Omnibus http://www.ncbi.nlm.nih.gov/geo/) accession number: GSE62843.

RNA-seq Bioinformatic Analysis

The RNA-seq data were processed using Illumina RTA software and converted into FASTQ files using the Illumina CASAVA pipeline. The FASTQ files were aligned to either the UCSC mm9 mouse genome (for the furan study) or the rn5 rat genome (for the time in paraffin study) with STAR (Dobin et al., 2013), a universal RNA-seq aligner, using default parameters. Sequences that mapped to more than 1 locus were excluded from further analysis. Gene expression values were calculated based on composite gene models of NCBI's Reference Sequence (RefSeq) gene annotation models (retrieved via UCSC's Table Browser May 7, 2014) using featureCounts (Liao et al., 2014), a program for assigning sequence reads to genomic features. Composite gene models for each gene consisted of the union of the exons from the set of all transcript isoforms from that gene.

Expression values were also calculated for a subset of RefSeq genes that corresponded to the intersection between the RefSeq gene models and the Agilent SurePrint G3 Mouse Microarray targets. The intersection between the RefSeq gene models and the Agilent SurePrint G3 Mouse Microarray targets was determined as follows. Microarray probe sequences corresponding to Agilent SurePrint G3 Mouse Microarrays were downloaded from NCBI (accession no. GPL13912). The probe sequences were aligned to the mouse genome using STAR. Probe sequences that mapped to more than 1 locus (5.65%) were excluded from downstream consideration. Probe sequences that mapped to just 1 locus were intersected with composite gene models from NCBI's RefSeq database using featureCounts. Probe sequences that unambiguously overlapped with no more than 1 RefSeq composite gene model (53.9%) were paired with that gene model; the remaining probe sequences were discarded. The pairs were collated into a unique, non-redundant list of RefSeq genes, which was used in subsequent differential gene expression analyses.

All sets of gene counts were normalized using voom (Law et al., 2014), which performs a LOWESS regression to estimate the mean-variance relation and transforms the gene counts into the appropriate log form for linear modeling. Differential gene expression analysis was then performed for each set of transformed gene counts using limma (Ritchie et al., 2015) for the following set of contrasts: FF Control versus FF High Dose, FFPE Control 18 h versus FFPE High Dose 18 h, and FFPE Control 3 weeks versus FFPE High Dose 3 weeks.

Online Data Access

The complete dataset is available through the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/). The superseries accession number is GSE62843.

RESULTS

Gene Expression Profiles

Our initial aim was to evaluate transcriptomic profiles obtained using different protocols. The mean RNA integrity numbers (RIN) for the FRO, FFPE-18 h, and FFPE-3-week samples were 9.3 (range: 8.9–9.5), 2.1 (range: 1.9–2.3), and 2.2 (range: 1.3–2.5), respectively. RNA integrity values (RIN) are reported in Supplementary Table S1. In order to make fair comparisons between RNA-seq and microarray datasets, RNA-seq reads were only considered if they could be mapped back to a transcript corresponding to a probe on the Agilent SurePrint G3 Mouse GE 8 × 60 K microarray. Mapping produced a list of 20 167 RefSeq genes. This list was also used to retrospectively filter the list of microarray-derived differentially expressed genes (DEGs). The numbers of DEGs in the full and mapped gene lists are listed in Table 1. With respect to the FRO groups, more DEGs were detected in the FFPE-18 h groups, and fewer were detected in

the FFPE-3-week groups in all experiments apart from the poly-A-enrichment RNA-seq. One-color DNA microarrays consistently detected about twice the number of DEGs as 2-color arrays across all sample types. Detailed discussion of differences in protocol and results between the 1- and 2-color Agilent array platforms can be found in our previously published work (Jackson et al., 2013). When compared with microarray protocols, RNA-seq DEG numbers were intermediate (FRO and FFPE-3 weeks) or modestly lower (FFPE-18 h) for the ribo-depletion protocol, and either higher (FRO) or lower (FFPE-18 h and FFPE-3 weeks) for the poly-A protocol. Applying a 2-fold cut-off reduced the number of DEGs by at least 75% across FRO and FFPE groups, limiting the ability to perform pathway analysis in some groups.

Next, we evaluated gene expression correlations by fold-change and sample type (Fig. 1a). The R^2 correlation coefficients between paired FRO and FFPE samples ranged from 0.29 to 0.93 depending on protocol; all correlations were statistically significant (regression $P < 0.05$). The correlation between FRO and FFPE-18 h samples was markedly higher for both ribo-depletion and poly-A RNA-seq (but not microarray) protocols compared with correlations between FRO and FFPE-3-week samples, indicating an erosion of signal with time in formalin at the read
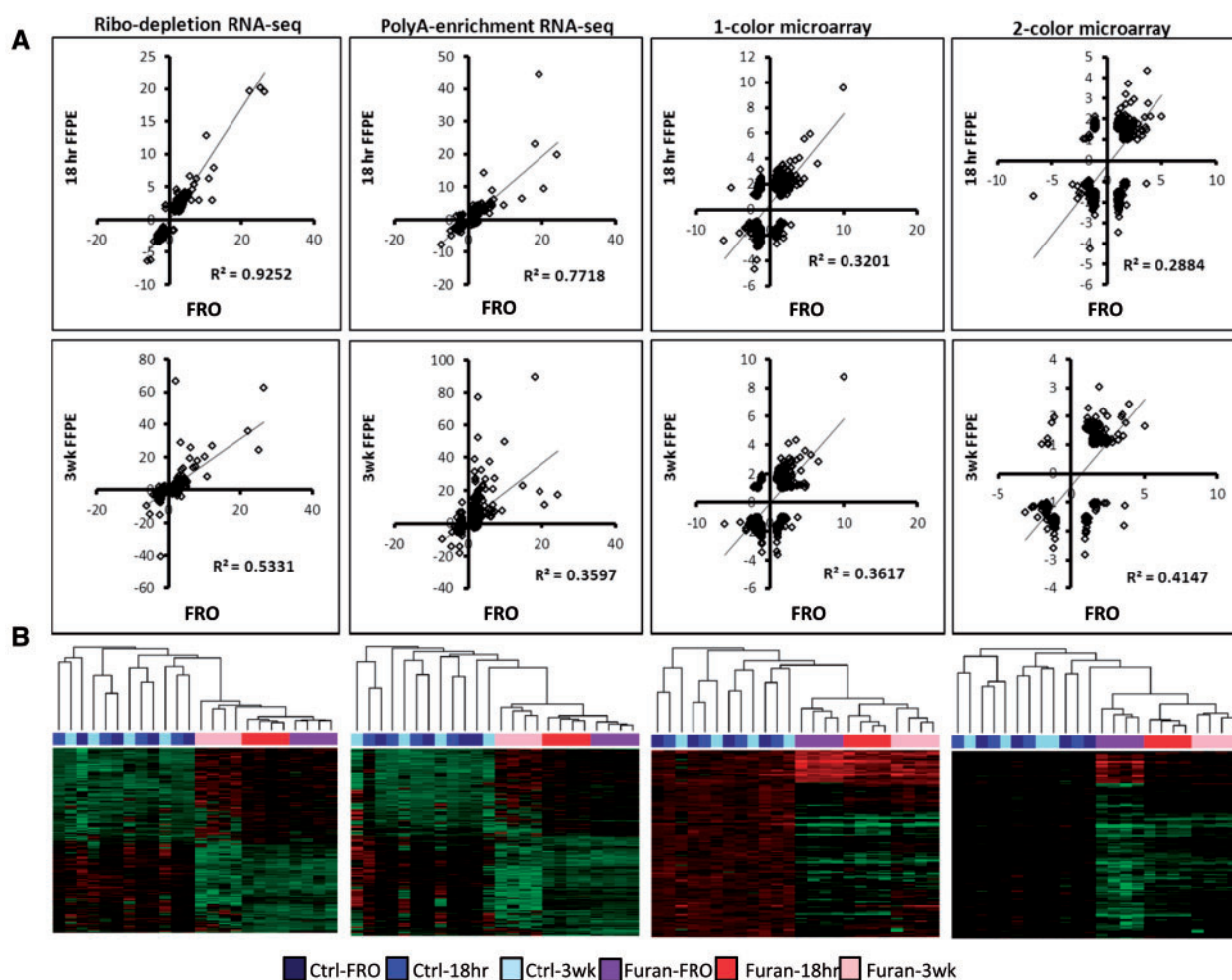
FIG. 1. Differential gene expression. Expression profiles following exposure to furan (8 mg/kg/day) in FRO, 18 h in formalin (FFPE-18 h), or 3 weeks in formalin (FFPE-3 weeks) liver samples evaluated using different DNA microarray and RNA-seq protocols. (A) Correlation analysis of DEG fold changes for FRO versus FFPE-18 h (top row) and FRO versus FFPE-3 weeks (bottom row) ($P < 0.0001$ for all linear regressions). Genes were significantly changed in at least one list (FDR $P < 0.05$, fold change $> \pm 1.5$ compared with control). (B) Hierarchical clustering of all DEGs (FDR $P < 0.05$, fold change $> \pm 2$).

depths used. Overall, the ribo-depletion method clearly provided the highest correlation between FRO and FFPE samples (at both 18 h and 3 weeks).

Hierarchical clustering of the datasets demonstrated that furan treatment was the main driver of gene expression profiles (Fig. 1b). A decrease in this signal with time in formalin is evident from the microarray cluster diagrams, which show a decrease in signal intensity in the furan groups from FRO to FFPE-18 h to FFPE-3 weeks (the increased amount of black on the heatmaps). Only 5 genes were up-regulated across all 12 datasets (*Osgin1, Tnfsfr12a, Srxn1, Ephx1,* and *Chka*). Genes up-regulated in 11/12 datasets were: *Cdkn1a, Jun, Dusp6, Hmox1, Brd2, Ugt2b35, Gclc,* and *Trp53inp1*; and in 10/12 datasets: *Tubb2a, Phlda3, Gdf15, Gsta2, Zmat3, Ugdh, Rcan1, Rbp1, Apoc2,* and *Ces2e*. For the 4 FRO datasets (ribo-, poly-A+, 1- and 2-color), the overlap of shared DEGs was: 78 (4/4 FRO datasets), 101 (3/4), and 206 (2/4). For the FFPE-18 h groups, these number were: 54 (4/4), 63 (3/4), and 807 (2/4). For the FFPE-3 weeks groups these numbers were: 7 (4/4), 16 (3/4), and 112 (2/4), demonstrating an erosion of quality with time in formalin. The DEG overlap for the 3 preservation methods (FRO, FFPE-18 h, and FFPE-3 weeks) for ribo-depletion was: 86 (3/3) and 165 (2/3); for poly-A-enrichment: 43 (3/3) and 233 (2/3); 1-color: 45 (3/3) and 167 (2/3); and 2-color: 18 (3/3) and 76 (2/3). Therefore, once again, the ribo-depletion protocol was the most consistent between groups.

Pathway Analysis

Pathway analysis of each DEG list resulted in a large number of enriched pathways (Table 1), which were ranked for significance in IPA using the $-\log(P\text{-value})$ of a Fisher's Exact test. To simplify inter-group comparisons, we first considered the top 15 pathways from each group by sample type (FRO, FFPE-18 h, FFPE-3 weeks). Using FRO as a reference group, the overlap was: 10/15 (ribo-depletion), 5/15 (poly-A-enrichment), 4/15 (1-color microarray), 2/15 (2-color microarray) (Fig. 2, bar charts). Once again, the greatest consistency among sample types was found with the ribo-depletion method. We then compared all enriched pathways between FRO and FFPE for each platform. The percent of FRO pathways that also occurred in FFPE-18 h were: 49% (ribo-depletion), 53% (poly-A-enrichment), 59% (1-color), and 52% (2-color) shared pathways. The percent of FRO pathways that also occurred in FFPE-3 weeks were: 41% (ribo-depletion), 18% (poly-A-enrichment), 24% (1-color), and 23% (2-color) shared pathways (Fig. 2, Venn diagrams). Overall, ribo-depletion and poly-A-enrichment were about equal at 18 h; however, ribo-depletion was better at 3 weeks. 1-color out-performed 2-color microarrays at both time-points.

Pathways relevant to the mode of action for furan (Jackson et al., 2014) were consistently enriched in FFPE samples (eg, *Nrf2-mediated Oxidative Stress Response, Xenobiotic Metabolism,* and *p53 Signaling*). Other relevant pathways related to cell cycle
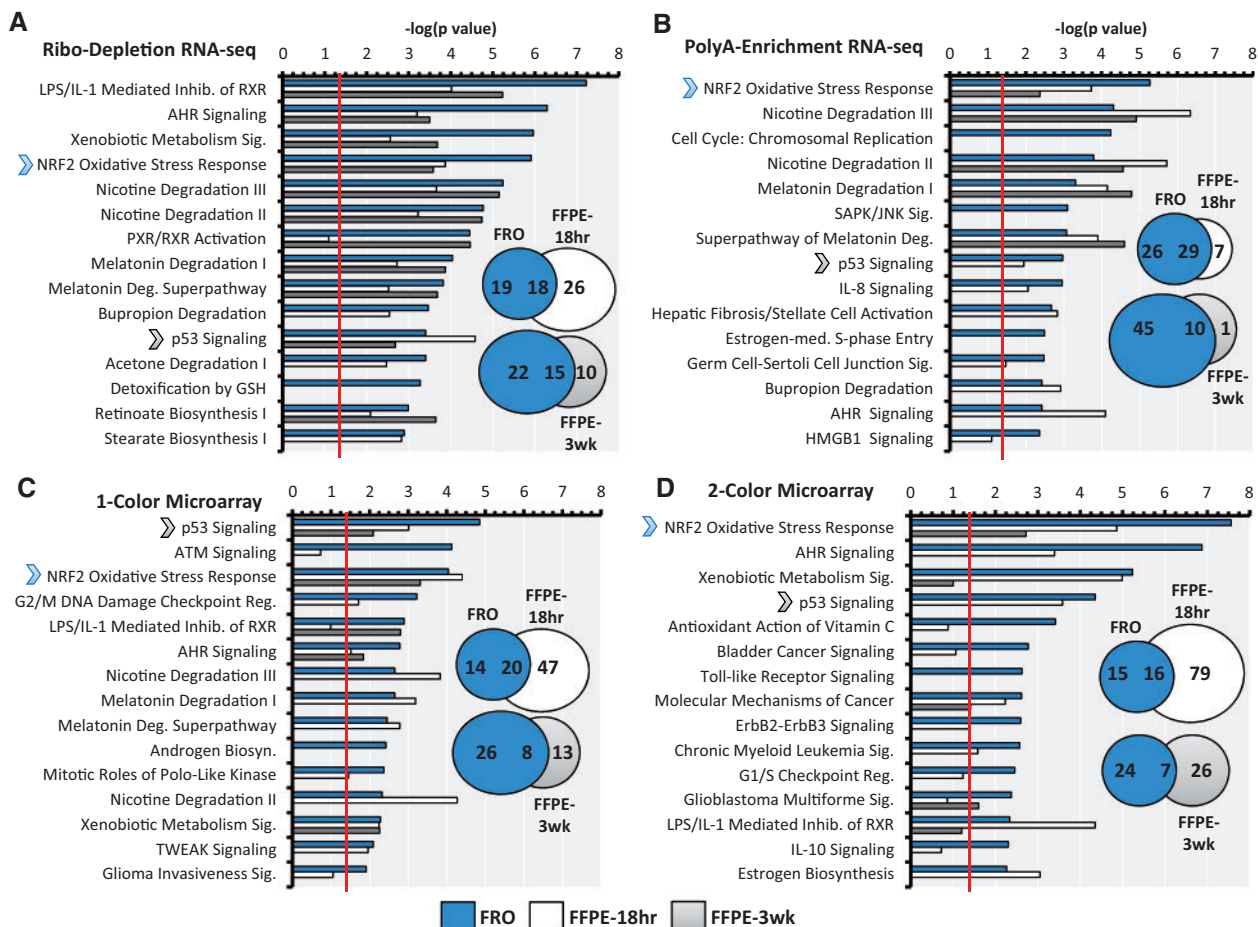
FIG. 2. Pathway analysis for (A) ribo-depletion RNA-seq, (B) poly-A-enrichment RNA-seq, (C) 1-color microarrays, and (D) 2-color microarrays. The top 15 pathways for the FRO groups (blue) are listed vertically. The level of enrichment of the corresponding FFPE groups (white-FFPE-18 h; grey-FFPE-3 weeks) is plotted. Pathway significance was calculated in IPA using a Fisher Exact test; $-\log(P\text{-value}) = 1.3$ corresponds to a $P = 0.05$ (indicated in red). Venn diagrams indicate the overlap of each full list. For all analyses, only mapped DEGs with FDR $P < 0.05$ and fold change $> \pm 1.5$ were used and only pathways with at least 4 DEGs were considered. Nrf2 Oxidative Stress Response was the only pathway that was enriched in all 12 groups (blue chevrons); p53 was enriched in 11/12 groups (grey chevrons). Full color version available online.

TABLE 1. Number of DEGs (by unique gene symbol) in each microarray and RNA-seq experiment using different filtering thresholds

	Fold change cut-off	Fresh Frozen				FFPE							
						18 h				3 weeks			
		±1.5		±2		±1.5		±2		±1.5		±2	
	P ≤ 0.05	FDR	Unadj.	FDR	Unadj.	FDR	Unadj.	FDR	Unadj.	FDR	Unadj.	FDR	Unadj.
Mapped RefSeq	RNA-seq												
	Ribo−	307 (37)	768	129	207	576 (44)	814	189	248	204 (25)	928	77	451
	PolyA+	609 (55)	675	189	199	438 (36)	741	160	249	128 (11)	1447	51	812
	Microarray												
	1 color	443 (35)	489	103	111	1656 (66)	1847	302	323	362 (20)	739	72	118
	2 color	197 (35)	234	47	53	798 (90)	798	77	77	149 (29)	152	12	12
Mapped RefSeq	RNA-seq												
	Ribo−	316 (44)	836	133	226	619 (46)	883	208	265	212 (25)	986	76	483
	PolyA+	659 (58)	728	206	217	455 (37)	783	165	257	136 (11)	1519	58	870
	Microarray												
	1 color	667 (28)	811	157	173	3543 (133)	3753	626	640	491 (40)	887	69	105
	2 color	339 (32)	405	84	93	1854 (196)	1853	192	188	266 (48)	272	25	24

The corresponding number of enriched IPA pathways (with P < 0.05 and at least 4 molecules) is indicated in brackets. Lists that were used to derive enriched pathways are indicated by italics text.

(G2/M DNA damage checkpoint), cell death (TWEAK signaling), inflammatory response (LPS/IL-1 inhibition of RXR function), and various cancer pathways were less consistently present across groups. Only the Nrf2 Oxidative Stress Response and p53 Signaling pathways were shared across all 4 experimental set-ups, and only the former was enriched in all 12 datasets. The pathway analysis using the unmapped gene lists for each group can be seen in Supplementary Figure S2.

Meta-analysis of Chemicals and Diseases Produces Similar Gene Expression Profiles

We performed a meta-analysis using NextBio to identify chemicals that produce similar changes in gene expression to furan in order to determine if these similarities were detectable in the FFPE furan samples. This meta-analysis was conducted using the data generated by the best-performing protocols from each platform: ribo-depletion RNA-seq and 1-color microarray protocols. We performed 2 chemical meta-analyses. In the first, expression profiles of the top 250 furan-dependent DEGs were compared with those produced in all mouse chemical studies (across the entire NextBio database). Furan was most similar to thioacetamide (on both platforms and across samples), followed by 1–5-napthalenediamine, dimethylnitrosamine, methapyrilene, and diethylnitrosamine (Table 2). In the second, we compared our datasets with the changes produced following a 3-week exposure to 26 different chemicals in female B6C3F1 mice by Thomas et al. (2011); this latter study was preselected as a more targeted study subset with the same exposure time, mouse strain, and sex as the current furan study. When compared with the Thomas et al. study, furan produced gene expression changes most similar to malathion, benzofuran, methylene chloride, and 1,5-napthalenediamine. There was a remarkable degree of overlap between FRO and FFPE samples, including identical top 5 rankings for FRO and FFPE-18 h groups across all NextBio datasets. Overall, the ribo-depletion study produced the most reproducible results across preservation techniques.

A third meta-analysis was performed to identify liver and biliary disease states that produce similar gene expression profiles to furan exposure (Table 3). The most correlated studies involved differential gene expression following partial hepatectomy and exposure to carbon tetrachloride (a Cyp2E1 substrate). Both of these are highly consistent with the known mode of action for furan, which includes xenobiotic metabolism by Cyp2E1, cytotoxicity, and regenerative proliferation of the liver. The FRO, FFPE-18 h, and FFPE-3-week groups shared 9/10 (ribo-depletion RNA-seq) or 10/10 (1-color microarray) of the most correlated studies, indicating consistency across the 2 platforms and 3 preservation methods for the highest ranking DEGs.

RNA-Seq Sample Profiles

To examine the impact of library preparation on FRO and FFPE transcriptomic profiles, we prepared libraries using either poly-A-enrichment or ribo-depletion protocols, and sequenced the libraries on an Illumina HiSeq2000 instrument. In order to evaluate the different libraries, we examined different aspects of the RNA-seq data using quality metrics that have been previously developed (Wang et al., 2012), including gene-body coverage uniformity, guanine-cytosine (GC) content distribution, and the distribution of the reads across genic structures (untranslated regions (UTRs), exons, introns, or intergenic regions).

We observed that the gene body coverage (calculated as a percentage of reads that covered each nucleotide position of all RefSeq genes scaled to 100 bins) was dependent on both the library preparation method and the sample type (Fig. 3a). In particular, because poly-A-enrichment is 3′ biased, and because FFPE samples are more degraded than FRO counterparts, the uniformity of gene coverage was biased toward the 3′ end of the gene for poly-A-enriched FFPE libraries. On the other hand, ribo-depleted RNA samples consistently showed more uniform gene coverage than did poly-A-enriched samples; however, after 3 weeks in fixation, even ribo-depleted samples began to show 3′ bias, indicating that 3 weeks in fixation is too long if one is concerned about detecting full-length RNA transcripts. Additionally, we observed that GC content distributions from each sample followed a bell curve, regardless of library preparation protocol (Fig. 3b). However, several differences were observed, namely in the widths and heights of those bell curves, likely indicating a difference in the population that was being sampled. It is noteworthy that a higher read depth is required

TABLE 2. Chemical signature analysis

	Ribo-depletion RNA-seq			1-color microarray		
		Time in formalin			Time in formalin	
	FRO	18 h*	3 weeks*	FRO	18 h	3 weeks
All NextBio						
Thioacetamide	1	1	2	1	1	1
1,5-Naphthalenediamine	2	2	1	6	31	3
Dimethylnitrosamine	3	3	5	3	35	25
Methapyrilene	4	4	3	2	2	2
Diethylnitrosamine	5	5	13	5	26	10
Anastrozole	6	14	14	4	17	6
Tunicamycin	7	19	6	7	10	31
Bleomycin	8	9	91	—	—	—
Benzo(a)pyrene	9	4	11	—	—	—
Bromobenzene	10	18	8	8	15	7
GSE18858	FRO	18 h*	3 weeks	FRO	18 h*	3 weeks*
Malathion (14,800 ppm)	1	1	1	1	1	1
Benzofuran (240 mg/kg)	2	2	2	3	2	2
Methylene chloride (4000 ppm)	3	6	14	5	14	14
1,5-Naphthalenediamine (2000 ppm)	4	3	3	4	3	3
Iodoform (93 mg/kg)	5	–	4	2	4	4
Methylene chloride (3000 ppm)	6	5	10	6	10	10
Methylene chloride (2000 ppm)	7	12	11	9	11	11
1,4-Dichlorobenzene (600 mg/kg)	8	4	5	7	5	5
1,2,3-Trichloropropane (60 mg/kg)	9	11	13	10	13	13
1,4-Dichlorobenzene (500 mg/kg)	10	10	8	8	8	8

Kendall's rank correlation for all NextBio (where * indicates a significant correlation with respect to FRO): Ribo- RNA-seq: FRO versus 18 h (tau = 0.674; **P = 0.0071**); FRO versus 3 weeks (tau = 0.511; **P = 0.0466**); 1 C microarray: FRO versus 18 h (tau = 0.214; P = 0.5484); FRO versus 3 weeks (tau = 0.429; P = 0.1789).

Kendall's rank correlation for GSE18858: Ribo- RNA-seq: FRO versus 18 h (tau = 0.556; **P = 0.0446**); FRO versus 3 weeks (tau = 0.467; P = 0.0726); 1C microarray: FRO versus 18 h (tau = 0.6; **P = 0.0167**); FRO versus 3 weeks (tau = 0.6; **P = 0.0167**). The expression values of the top 250 mapped DEGs (FC > 1.5, FDR P < 0.05 for furan-treated mice versus respective controls) stratified by sample group (FRO, FFPE-18-h, FFPE-3-week) were compared with liver gene expression signatures from all NextBio mouse studies (top) and GSE18858 (Thomas et al., 2011; bottom). In the latter, female B6C3F1 mice were exposed to one of 26 chemicals for 13 weeks. The top 5 most correlated datasets for each group are indicated by italics text (where 1 indicates the most correlated study, 2 the second most correlated, etc.); only positively correlated chemicals with at least 3 supporting studies were included (all correlations are P < 0.05, which was calculated in NextBio using a Fisher's exact test).

for ribo-depletion experiments (compared with poly-A-enrichment) in order to detect equivalent numbers of protein coding genes; therefore, the sequencing depth of the ribo-depletion experiment (10–60 M reads) was greater than that used in the poly-A-enrichment experiment (3.8–11 M reads). The distribution of reads across genic structure (calculated as the fraction of the reads that mapped to either exons, UTRs, introns, or intergenic regions) revealed differences between the 2 sample types (Fig. 3c). The proportion of exonic reads was largest in the FRO samples and decreased (with a concomitant increase of intronic reads) from FRO, to 18-h-FFPE, to 3-week-FFPE. The proportion of 3′ UTR reads also appeared to be slightly higher in the FRO samples compared with the FFPE samples. The distribution of reads across genic structures was very consistent between the 2 library preparation protocols.

Long-term Archival Samples
To examine whether these techniques would be applicable to a wide range of FFPE samples, we tested an additional set of FFPE samples. These archival FFPE samples were obtained from the NTP and had been banked for 8–26 years. It was not possible to obtained archival furan-exposed samples, so the decades-old samples that were tested here represented the control tissue blocks from toxicity studies for beta-picoline (8 years), oxazepam (19 years), and trichlorfon (26 years). These samples were derived from rat; however, it is unlikely that formalin would

interact with the genome in a species-specific manner and thus we would expect to see the similar results for mouse. We sequenced each set of samples using the ribo-depletion protocol and compared the expression profiles to standard metrics established by the FDA's SEQC RNA-seq Consortium (Li et al., 2014a).

We observed that the ribo-depletion protocol was very successful in creating expected distributions of mapped reads to the exonic, UTR, intronic, and ribosomal fractions of the transcriptome (Fig. 3d). The only aspect of the mapped reads that was high was the percent of mapped reads that were intergenic (violet), which could indicate degradation from the long-banked samples, although we found this trend across all samples. Moreover, the percentage of GC from each sample followed a normal distribution (Fig. 3e), and did not show any significant shift as a function of treatment or time. Overall, these results demonstrate the potential applicability of these techniques for very old FFPE samples.

DISCUSSION

Archival FFPE samples provide a unique opportunity to directly link molecular profiles and histopathologic or clinical outcomes without additional studies. Resulting profiles provide detailed mapping of target pathways, identification of biomarkers for both environmental chemicals and pharmaceutical agents, and insight for understanding of modes of action. However, better

TABLE 3. Disease Signature Analysis

Treatment	Time of tissue collection	Mouse details	GEO	FRO	Time in formalin	
					18 h*	3 weeks*
Ribo-depletion RNA-seq						
Partial hepatectomy	48 h post-PH	5–6 m/CB6F1	GSE20427	1	1	2
Partial hepatectomy	48 h post-PH	25–27 m/CB6F1	GSE20427	2	3	4
Partial hepatectomy	18 h post-PH	Adult/wildtype C57BL/6	GSE51801	3	2	5
CCl4-exposure	2 d post-exposure	BALB/c (IL4 IL13 double knockout)	GSE45002	4	4	1
CCl4-exposure	2 d post-exposure	BALB/c (wildtype)	GSE45002	5	6	6
Partial hepatectomy	2 h post-PH	25–27 m/CB6F1	GSE20427	6	5	3
Partial hepatectomy	2 h post-PH	5–6 m/CB6F1	GSE20427	7	9	-
Partial hepatectomy	24 h post-PH	Tob1-null	GSE21836	8	8	7
Partial hepatectomy	38 h post-PH	5–6 m/CB6F1	GSE20427	9	7	9
Partial hepatectomy	24 h post-PH	5–6 m/CB6F1	GSE20427	10	10	8
1-color microarray				FRO	18 h	3 weeks*
Partial hepatectomy	48 h post-PH	5–6 m/CB6F1	GSE20427	1	3	2
Partial hepatectomy	48 h post-PH	25–27 m/CB6F1	GSE20427	2	4	4
CCl4-exposure	2 d post CCl4-exposure	BALB/c (wildtype mice)	GSE45002	3	9	5
CCl4-exposure	2 d post CCl4-exposure	BALB/c (IL4/IL13 double knockout mice)	GSE45002	4	2	3
Partial hepatectomy	18 h post-PH	Adult/wildtype C57BL/6	GSE51801	5	1	1
Partial hepatectomy	38 h post-PH	5–6 m/CB6F1	GSE20427	6	7	6
Partial hepatectomy	24 h post-PH	Tob1-null (Tob1 = a repressor of liver regeneration)	GSE21836	7	5	8
Ehrlichia chaffeensis induced Hepatitis	15 d Wakulla strain	Severe combined immunodeficient mice	GSE8966	8	8	9
Partial hepatectomy	24 h post-PH	5–6 m/CB6F1	GSE20427	9	6	7
Partial hepatectomy	2 h post-PH	25–27 m/CB6F1	GSE20427	10	10	10

PH, partial hepatectomy; CCl4, carbon tetrachloride; m, month.
All samples are: mouse; liver; treated versus control.
Kendall's rank correlation (where *indicates a significant correlation with respect to FRO): Ribo- RNA-seq: FRO versus 18 h (tau = 0.778; $P = 0.0009$); FRO versus 3 weeks (tau = 0.611; $P = 0.0247$); 1 C microarray: FRO versus 18 h (tau = 0.378; $P = 0.1557$); FRO versus 3 weeks (tau = 0.644; $P = 0.0091$). The expression values of the top 250 mapped DEGs (FC > 1.5, FDR $P < 0.05$) in each group (FRO, FFPE-18-h, FFPE-3-week) determined by RNA-seq or microarray analysis were compared with gene expression signatures derived from publically available mouse liver and biliary disease datasets using NextBio. The top 5 most correlated datasets for each group are indicated by italics text (where 1 indicates the most correlated study, 2 the second most correlated, etc.); only positively correlated studies were included. All correlations have a $P < 0.05$, which was calculated in NextBio using a Fisher's Exact Test.

methods for transcriptomic analysis of FFPE tissues are needed. Several recent proof-of-concept studies have evaluated RNA-seq as a novel way to derive gene expression profiles from FFPE or degraded RNA samples and reported promising results (Auerbach *et al.*, 2014; Hedegaard *et al.*, 2014; Linton *et al.*, 2012; Spencer *et al.*, 2013; Zhao *et al.*, 2014b). Our goal was to expand on these studies by examining specific questions related to the use of RNA-seq in FFPE tissues, including characterizing the performance versus microarray platforms, discerning the optimal library preparation method, and examining the effects of time in formalin and time in paraffin. As expected, RNA from FFPE preserved tissues was highly degraded, with RIN values declining from approximately 9 in FRO tissues to 2 in FFPE tissues. Although signal declined with RNA degradation, all of the genomics tools were able to identify perturbations that are consistent with the known mode of action for furan. Thus, our results strongly indicate that valuable information can also be obtained from samples with low RINs. These FFPE results support recent work which has shown that degraded RNA samples with low RIN values can provide comparable sets of DEGs to intact, fresh RNA (Li *et al.*, 2014b). We further hypothesized that the use of short-read sequencing (where shearing to produce shorter fragments is an essential step) would provide the best method to meaningfully query such degraded RNA. We found that the ribo-depletion RNA-seq method with a relatively brief (18 h) fixation time provided the strongest correlation between FRO and FFPE tissues at a global transcriptomic level. Methods

were also applied to extremely old FFPE samples, showing no clear effect of time in paraffin on RNA-seq quality metrics. For highly altered DEGs concordance was also observed across FRO and FFPE samples for both RNA-seq and microarray platforms. These findings indicate that increased time in formalin and paraffin may result in loss of signal intensity and other indicators of degradation, but not to the extent that key molecular signatures are lost.

Our first aim was to compare both platforms (2 protocols each) in order identify the best method for future FFPE experiments. Across platforms, ribo-depletion RNA-seq provided the strongest gene fold change correlations between FRO and FFPE (Fig. 1) and the highest overlap of top pathways (Fig. 2). Within platforms, ribo-depletion outperformed poly-A-enrichment RNA-seq, and 1-color outperformed 2-color microarrays. When compared with the poly-A-enrichment protocol, the ribo-depletion protocol produced a more even coverage across all genes and also revealed more low-expressing transcripts. Meta-analyses of the ribo-depletion datasets also showed high concordance in top chemical (Table 2) and disease (Table 3) signatures. The main advantage of the ribo-depletion approach over the poly-A-approach for FFPE samples is that it does not rely on the poly-A tails of the mRNA molecules, which are known to be highly modified and degraded by formalin (Farragher *et al.*, 2008; Klopfleisch *et al.*, 2011). The strength of the RNA-seq platform over microarrays is that it has (essentially) an unlimited dynamic range, and reads are not confined
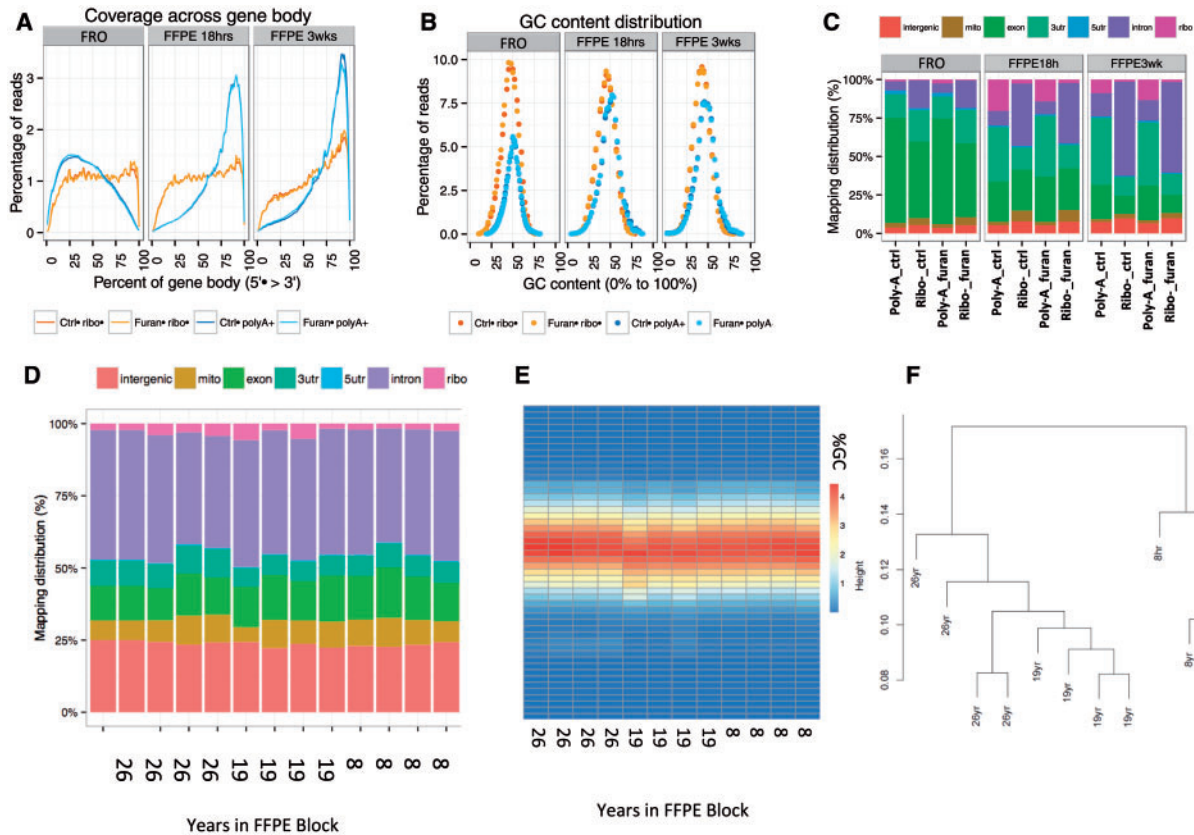
FIG. 3. Examination of RNA-seq data quality metrics. **(A–C)** Mapping distributions and GC content distributions for the furan-treated sampled. (A) The percentage of reads that covers each nucleotide position of all genes scaled to 100 bins, from 5′ UTR to 3′ UTR for FRO and FFPE samples. The ribo-depleted RNA samples consistently showed more uniform gene coverage than did their poly-A–selected counterparts. (B) GC content distributions for FRO and FFPE samples. (C) The percentage of FRO and FFPE reads that map to various genic regions. Overall, we detected more intronic reads from ribo-depleted RNA samples than from poly-A–enriched libraries; this trend increased as a function of time in formalin. (D,E) Mapping distributions and GC content distributions, expressed as a percentage, for 8-, 19-, and 26-year-old control FFPE tissues. (F) Inter-group Kendell's tau comparison.

by predetermined probe sequences. However, because microarrays have been in use for much longer, the data normalization, processing, and analysis associated with them is more established and straightforward. Collectively, these findings support the ribo-depletion approach for future FFPE genomics experiments.

Our second aim was to examine how formalin fixation impacts RNA quality and the detection of a chemical signature. Formalin generates crosslinks between cellular macromolecules and introduces various adducts and other covalent modifications. Sample fixation time varies (from hours to weeks) across laboratories; therefore, an important consideration in planning a retrospective FFPE genomics study is to understand how the initial fixation in formalin will affect the quality of the final gene expression data. Our findings indicate that longer fixation times decrease RNA yields from paraffin-embedded tissues (herein 3 weeks in formalin resulted in half the RNA yield/section compared with 18 h) and increase RNA input needed during library preparation (see methods). This latter effect is likely due to 'non-functional' RNA (eg, adducts or damaged bases) that could be potentially mitigated by treatments like PreCR or other enzymatic and chemical methods. Extended time in formalin (3 weeks) led to lower correlation with FRO samples (RNA-seq) (Fig. 1a), detection of fewer DEGs (RNA-seq and microarray) (Table 1), and erosion of signal intensity (Fig. 1b) (microarray). It is unclear whether this erosion in signal occurs early, or if signal continues to decline beyond 3 weeks. Given that time in

formalin varies, future studies should investigate samples held in formalin for longer durations.

Samples that stayed in formalin for only 18 h produced a larger number of DEGs than their FRO counterparts, especially for the microarray protocols. Reasons for this effect are not clear, but the same phenomenon has been observed previously (eg, Hedegaar *et al.*, 2014), suggesting that FFPE samples may have an increased propensity for producing false positive genes. To account for this, future experiments may consider including technical replicates and only treating DEGs that appear in both replicates as true positives. Also, RNA-seq is more variable at lower read depths; thus, deeper sequencing is another way to increase signal intensity and reduce potential false-positives. Moreover, our results suggest that longer incubations in formalin may reduce DEG numbers due to lower signal intensity. These data indicate that FFPE genomics studies should, whenever possible, favor samples that have spent limited time in formalin. Our findings also highlight the fact that FFPE tissue blocks may vary widely in RNA quality (despite similar RIN profiles), pointing to a need for improved quality metrics specific to FFPE samples.

An important future application of FFPE genomics is the identification of key molecular features in a chemical or drug mode of action. Despite inter-platform and inter-preservation differences in DEG lists, the higher-level biology (pathways and meta-analyses) detected in the FFPE groups in our study (particularly the FFPE-18-h) corresponded well with paired FRO groups

and the expected furan-induced molecular changes. Concordance was strongest for the ribo-depletion RNA-seq protocol. The test article in this study, furan, is a well-characterized hepatocarcinogen with an extensive toxicological literature. We recently characterized early molecular key events in the mode of action of furan in female mice using FRO livers and identified a key role for the Nrf2 Oxidative Stress Response pathway (Jackson *et al.*, 2014; Recio *et al.*, 2013). In this study, this pathway was consistently enriched across all groups and platforms. Pathways enriched in the FRO groups were typically well represented in the corresponding FFPE-18 hr groups, while the FFPE-3-week groups showed fewer response pathways. These findings indicate that samples fixed in formalin for shorter periods of time provide more consistent pathway-level results.

In addition to pathway analysis, we performed 2 meta-analyses using NextBio software. This approach showed higher inter-group consistency than the pathway analysis, likely because it does not group gene expression profiles into specific events (such as pathways). Instead, comparisons are made between entire biosets generated from different chemical treatments or disease states, allowing comparisons between data-poor and data-rich compounds. Here, furan gene expression changes were most similar to those induced by thioacetamide and carbon tetrachloride, which are also metabolized in the liver by Cyp2E1 (Kang *et al.*, 2008; Manibusan *et al.*, 2007). Moreover, the most similar disease state was liver regeneration following partial hepatectomy, consistent with previous studies showing that the carcinogenic mode of action for furan includes cytotoxicity followed by dysregulated regenerative proliferation (Fransson-Steen *et al.*, 1997; Moser *et al.*, 2009). These meta-analyses indicate that FFPE samples can provide valuable expression-based information for prioritization of chemicals and next-tier studies.

Lastly, we investigated effects of time in paraffin block on RNA quality metrics. No clear effects of time in paraffin were observed, supporting the idea that older FFPE samples could be used for transcriptome profiling (Fig. 3). Other researchers have reported RNA-seq for 9- and 20-year-old samples (Meng *et al.*, 2013; Hedegaard *et al.*, 2014), but to our knowledge these are the oldest samples tested. Although the data from these samples did not have the matching FRO samples for comparison, the data suggest that changes in RNA quality parameters due to time in paraffin may be limited and that decades-old samples may be a viable option for FFPE genomics. These data show initial proof-of-concept for the feasibility of using archival FFPE samples from both mouse and rat samples using ribo-depletion RNA-seq. Future work remains to directly compare ancient/archival FFPE and paired FRO samples, to define optimal pre-analytical and bioinformatics approaches, and to examine performance metrics following treatment for archival tissues.

Although there are differences between the genomic profiles of FRO and FFPE samples, our study demonstrates that: (1) RNA-seq with ribo-depletion library preparation is the preferred method for transcriptomic analysis of FFPE tissues; (2) FFPE samples can be used to obtain reliable transcriptomic information related to toxicant mode of action; and (3) the RNA profile of a compound is affected by the amount of time in formalin (although this effect becomes less pronounced with respect to higher-level analyses). Time in paraffin did not have a large effect on distributions of mapped reads across control FFPE samples. These findings provide evidence that FFPE samples can be used in genomics studies in absence of existing FRO samples (or when new animal studies are not feasible). Additional studies could also evaluate whether higher read depth may compensate in part for the loss of signal observed in FFPE samples formalin-fixed for longer periods of time, which could impact work from archival samples in quantitative applications such as transcriptomic benchmark dose estimation. Based on our experiment, several recommendations can be made. Fresh frozen tissues should be collected from experiments where it is envisioned that toxicogenomics may be used, as is standard practice. However, if flash freezing is not possible, the results clearly indicate that time in formalin is an important consideration. We strongly recommend that all studies in the future use the 18–24 h time in formalin window traditionally used for studies with immunohistochemical endpoints, and also keep the time in formalin similar between samples, to optimize any future genomic work. Time in paraffin appears to be less of a factor, and thus for retrospective studies the first consideration should be whether information is available on time in formalin. RIN values are not apparently effective in determining whether samples are amenable to genomics analyses at this stage. We recommend that effort be invested in developing a more appropriate metric for assessing whether the sample quality is high enough to proceed with genomic analyses. In addition, there are other variations of RNA-seq protocols in use for low-input and varying size RNAs, and these remain to be tested, as well as very deep sequencing of these samples to estimate the ability to detect splicing isoforms and splice junctions. Finally, ribo-depletion RNA-seq yields the most comparable data to frozen tissues, and we advise that this methodology be the first considered for analysis of archival tissues.

Overall, our work demonstrates that analysis of FFPE tissues represents an important opportunity for performing retrospective, phenotypically anchored research that characterizes chemical- and disease-dependent genomic changes. These chemical/disease signatures can be applied and reapplied (via meta-analysis) to the many data-poor environmental chemicals in need of assessment. This information would also provide a critical bridge between emerging *in vitro* data and health effects established in prior toxicologic and epidemiologic studies, accelerating the input and value of higher-throughput models in safety assessment, biomarker development, and drug discovery.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

REFERENCES

Auerbach, S. S., Phadke, D. P., Mav, D., Holmgren, S., Gao, Y., Xie, B., Shin, J. H., Shah, R. R., Merrick, B. A., and Tice, R. R. (2014). RNA-Seq-based toxicogenomic assessment of fresh frozen and formalin-fixed tissues yields similar mechanistic insights. *J. Appl. Toxicol.*

Bass, B. P., Engel, K. B., Greytak, S. R., and Moore, H. M. (2014). A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: How well do you know your FFPE specimen? *Arch. Pathol. Lab. Med.* **138**, 1520–1530.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* **57**, 289–300.

Black, M., Parks, B. B., Pluta, L., Chu, T. M., Allen, B. C., Wolfinger, R. D., and Thomas, R. S. (2014). Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. *Toxicol Sci.* **137**, 385–403.

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193.

Budczies, J., Weichert, W., Noske, A., Muller, B. M., Weller, C., Wittenberger, T., Hofmann, H. P., Dietel, M., Denkert, C., and Gekeler, V. (2011). Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *J. Histochem. Cytochem.* **59**, 146–157.

Cui, X., Hwang, J. T., Qiu, J., Blades, N. J., and Churchill, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6**, 59–75.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Farragher, S. M., Tanney, A., Kennedy, R. D., and Paul Harkin, D. (2008). RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochem. Cell Biol.* **130**, 435–445.

Fransson-Steen, R., Goldsworthy, T. L., Kedderis, G. L., and Maronpot, R. R. (1997). Furan-induced liver cell proliferation and apoptosis in female B6C3F1 mice. *Toxicology* **118**, 195–204.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315.

Goodnight, J. H., and Harvey, W. R. (1978). Least-squares means in the fixed-effects general linear models. *SAS Technical Report R-103*, Cary, NC: SAS Institute.

Hedegaard, J., Thorsen, K., Lund, M. K., Hein, A. M., Hamilton-Dutoit, S. J., Vang, S., Nordentoft, I., Birkenkamp-Demtröder, K., Kruhøffer, M., Hager, H., *et al.* (2014). Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One* **9**, e98187.

IARC. (1995). Dry cleaning: Some chlorinated solvents and other industrial chemicals. *IARC Monogr. Eval. Carcinog. Risks Hum.* **63**, 33–475.

Jackson, A. F., Williams, A., Moffat, I., Phillips, S. L., Recio, L., Waters, M. D., Lambert, I. B., and Yauk, C. L. (2013). Preparation of archival formalin-fixed paraffin-embedded mouse liver samples for use with the Agilent gene expression microarray platform. *J. Pharmacol. Toxicol. Methods* **68**, 260–268.

Jackson, A. F., Williams, A., Recio, L., Waters, M. D., Lambert, I. B., and Yauk, C. L. (2014). Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicol. Appl. Pharmacol.* **274**, 63–77.

Jacobson, T. A., Lundahl, J., Mellstedt, H., and Moshfegh, A. (2011). Gene expression analysis using long-term preserved formalin-fixed and paraffin-embedded tissue of non-small cell lung cancer. *Int. J. Oncol.* **38**, 1075–1081.

Kang, J. S., Wanibuchi, H., Morimura, K., Wongpoomchai, R., Chusiri, Y., Gonzalez, F. J., and Fukushima, S. (2008). Role of CYP2E1 in thioacetamide-induced mouse hepatotoxicity. *Toxicol. Appl. Pharmacol.* **228**, 295–300.

Kedderis, G. L., Carfagna, M. A., Held, S. D., Batra, R., Murphy, J. E., and Gargas, M. L. (1993). Kinetic analysis of furan biotransformation by F-344 rats in vivo and in vitro. *Toxicol. Appl. Pharmacol.* **123**, 274–282.

Kedderis, G. L., and Held, S. D. (1996). Prediction of furan pharmacokinetics from hepatocyte studies: Comparison of bioactivation and hepatic dosimetry in rats, mice, and humans. *Toxicol. Appl Pharmacol.* **140**, 124–130.

Kerr, M. K. (2003). Design considerations for efficient and effective microarray studies. *Biometrics* **59**, 822–828.

Klopfleisch, R., Weiss, A. T., and Gruber, A. D. (2011). Excavation of a buried treasure–DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histol. Histopathol.* **26**, 797–810.

Kupershmidt, I., Su, Q. J., Grewal, A., Sundaresh, S., Halperin, I., Flynn, J., Shekar, M., Wang, H., Park, J., Cui, W., *et al.* (2010). Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS One* **5**, pii:e13066.

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29.

Li, S., Labaj, P. P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., Phan, J., Wu, P. Y., Wang, M., Wang, C., *et al.* (2014a). Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat. Biotechnol.* **32**, 888–895.

Li, S., Tighe, S. W., Nicolet, C. M., Grove, D., Levy, S., Farmerie, W., Viale, A., Wright, C., Schweitzer, P. A., Gao, Y., *et al.* (2014b). Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat. Biotechnol.* **32**, 915–925.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

Linton, K., Howarth, C., Wappett, M., Newton, G., Lachel, C., Iqbal, J., Pepper, S., Byers, R., Chan, W. J., and Radford, J. (2012). Microarray gene expression analysis of fixed archival tissue permits molecular classification and identification of potential therapeutic targets in diffuse large B-cell lymphoma. *J. Mol. Diagn.* **14**, 223–232.

Malone, J. H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 34.

Manibusan, M. K., Odin, M., and Eastmond, D. A. (2007). Postulated carbon tetrachloride mode of action: A review. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* **25**, 185–209.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517.

Meng, W., McElroy, J. P., Volinia, S., Palatini, J., Warner, S., Ayers, L. W., Palanichamy, K., Chakravarti, A., and Lautenschlaeger, T. (2013). Comparison of microRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. *PLoS One.* **8**, e64393.

Moro, S., Chipman, J. K., Wegener, J. W., Hamberger, C., Dekant, W., and Mally, A. (2012). Furan in heat-treated foods: Formation, exposure, toxicity, and aspects of risk assessment. *Mol. Nutr. Food Res.* **56**, 1197–1211.

Moser, G. J., Foley, J., Burnett, M., Goldsworthy, T. L., and Maronpot, R. (2009). Furan-induced dose–response relationships for liver cytotoxicity, cell proliferation, and tumorigenicity (furan-induced liver tumorigenicity). *Exp. Toxicol. Pathol.* **61**, 101–111.

NTP. (1993). NTP technical report on the toxicology and carcinogenesis studies of Furan in F344/N Rats and B6C3F1 Mice (Gavage Studies). *Tech. Rep. Ser.* **402**, 1–286.

R-Core-Development-Team. (2012). R: A language and environment for statistical computing. Version 2.11.1.

Recio, L., Phillips, S. L., Maynor, T., Waters, M., Jackson, A. F., and Yauk, C. L. (2013). Differential expression of long non-coding RNAs in the livers of female B6C3F1 mice exposed to the carcinogen furan. *Toxicol. Sci.* **135**, 369–379.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.

Sadi, A. M., Wang, D. Y., Youngson, B. J., Miller, N., Boerner, S., Done, S. J., and Leong, W. L. (2011). Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens. *BMC Cancer* **11**, 253:1–13.

Searle, S. R., Speed, F. M., and Milliken, G. A. (1980). The population marginal means in the linear model: An alternative to least squares means. *Am. Stat.* **34**, 216–221.

SEQC/MAQC-III Consortium. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914.

Spencer, D. H., Sehn, J. K., Abel, H. J., Watson, M. A., Pfeifer, J. D., and Duncavage, E. J. (2013). Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J. Mol. Diagn.* **15**, 623–633.

Thomas, R. S., Clewell H. J., III, Allen, B. C., Wesselkamper, S. C., Wang, N. C. Y., Lambert, J. C., Hess-Wilson, J. K., Zhao, Q. J., and Andersen, M. E. (2011). Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol. Sci.* **120**, 194–205.

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* **32**, 926–932.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics.* **28**, 2184–2185.

Wu, H., Kerr, K. K., Cui, X., and Churchill, G. A. (2003). *MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments*. Springer, London.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15.

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D. N., and Perou, C. M. (2014a). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics.* **15**, 419.

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D. N., and Perou, C. M. (2014b). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419.