RESEARCH ARTICLE

Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue

Christy L. Trejo◯[ID]◉, **Miloš Babić**[ID]◉, **Elliot Imler**◉, **Migdalia Gonzalez, Sergei I. Bibikov, Peter J. Shepard, Harper C. VanSteenhouse, Joanne M. Yeakley**[ID], **Bruce E. Seligmann**[ID]*

BioSpyder Technologies, Inc., Carlsbad, California, United States of America

◉ These authors contributed equally to this work.
* bruceseligmann@biospyder.com

Abstract

We describe the use of a ligation-based targeted whole transcriptome expression profiling assay, TempO-Seq, to profile formalin-fixed paraffin-embedded (FFPE) tissue, including H&E stained FFPE tissue, by directly lysing tissue scraped from slides without extracting RNA or converting the RNA to cDNA. The correlation of measured gene expression changes in unfixed and fixed samples using blocks prepared from a pellet of a single cell type was $R^2 = 0.97$, demonstrating that no significant artifacts were introduced by fixation. Fixed and fresh samples prepared in an equivalent manner produced comparable sequencing depth results (+/- 20%), with similar %CV (11.5 and 12.7%, respectively), indicating no significant loss of measurable RNA due to fixation. The sensitivity of the TempO-Seq assay was the same whether the tissue section was fixed or not. The assay performance was equivalent for human, mouse, or rat whole transcriptome. The results from 10 mm^2 and 2 mm^2 areas of tissue obtained from 5 µm thick sections were equivalent, thus demonstrating high sensitivity and ability to profile focal areas of histology within a section. Replicate reproducibility of separate areas of tissue ranged from $R^2 = 0.83$ (lung) to 0.96 (liver) depending on the tissue type, with an average correlation of $R^2 = 0.90$ across nine tissue types. The average %CVs were 16.8% for genes expressed at greater than 200 counts, and 20.3% for genes greater than 50 counts. Tissue specific differences in gene expression were identified and agreed with the literature. There was negligible impact on assay performance using FFPE tissues that had been archived for up to 30 years. Similarly, there was negligible impact of H&E staining, facilitating accurate visualization for scraping and assay of small focal areas of specific histology within a section.

Introduction

Gene expression profiling of tissue is vitally important for understanding both normal and disease processes. Tissue can be prepared as snap frozen blocks or prepared as formalin fixed paraffin embedded (FFPE) tissue blocks, then sectioned and assayed. Frozen tissue blocks are

amenable to gene expression assays, but not without significant problems. The samples are difficult to handle and transport, as they must be kept frozen from the moment of collection onwards. FFPE samples preserve tissue morphology, and can be stained and cut easily for use in diagnostics (especially for grading and staging cancers). Thus, the general clinical pathological practice is to collect FFPE blocks rather than freeze [1, 2].

Assays of gene expression in formalin-fixed paraffin-embedded tissues have historically been complex and problematic, and have often provided subpar data [1–5]. Extraction of RNA from FFPE is typically low yield compared to fresh or frozen tissue, and the resulting RNA is highly fragmented and degraded, leading to poor performance in the usual methodologies which rely on reverse-transcription of RNA. The problem is particularly significant for archival FFPE samples, where RNA degradation is generally more pronounced. Vast collections of such samples are currently present in various hospitals and research centers around the world, and these samples are often matched with detailed clinical and outcome data. Similar archives exist for animal tissues from experimental studies performed over many decades. Yet, the treasure trove of gene expression data available in such archives has largely remained out of reach.

Additional problems are caused by tissue heterogeneity, as samples usually contain many different cell types and associated histology within a section, where the cell type or histology of interest may represent a small percentage of the total. Current FFPE RNA extraction methods typically require multiple complete sections of tissue to be processed together to recover sufficient material for transcription assays [6]. Therefore, a method that does not require the extraction of RNA, is not sensitive to fragmentation, and which can be used to profile small focal histologically and distinct areas of archived (not just fresh) FFPE has tremendous potential to advance science.

Two other commercial platforms enable investigators to profile FFPE without extraction of RNA, but both require dedicated hardware and neither permit the whole transcriptome to be profiled. nCounter (NanoString, Inc.) is limited to profiling the expression of up to 700 genes. EdgeSeq (HTG Molecular, Inc.) is limited to a few thousand genes. Thus, to use either of these methods the investigator must already know the set of genes to monitor in a focused assay, which means they have to first carry out a method such as RNA-Seq (or microarray) which necessitates extracting RNA from FFPE, or using fresh or frozen tissue. Translating such data to a different platform is typically problematic. Hence, a method that enables profiling of the whole protein-coding transcriptome, ~20,000 genes, from FFPE without RNA extraction, without use of dedicated hardware, and then (as desired) selecting genes to formulate a focused assay on the same platform, would be a significant advance.

In this study, we extend the use of the targeted, ligation-based Templated Oligo Sequencing (TempO-Seq) whole transcriptome assay to FFPE samples [7]. The experiments described were carried out using human, mouse, and rat whole transcriptome panels. Since it relies on probe hybridization rather than reverse transcription, TempO-Seq chemistry is highly resistant to RNA fragmentation and degradation, making it perfectly suited for fixed tissue samples. We show that FFPE samples can produce gene expression data on par with fresh tissues and cell cultures, that decades-old archival samples can be successfully processed without laborious extraction and purification methods, and that such tissues can even be H&E stained so that small focal areas of interest can be profiled independently of the surrounding tissue.

Materials

TempO-Seq FFPE assay reagents are commercially available from BioSpyder Technologies, Inc. The components of the kit are proprietary and consist of 2X Lysis buffer, FFPE Protease, FFPE nuclease, species-specific detector oligo pools designed to recognize the whole

transcriptome, and buffers necessary for annealing, nuclease clean up, ligation, amplification, and library generation. For library purification, we used the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel cat # 740609.50). Molecular biology grade light mineral oil was sourced from Sigma. Phosphate Buffered Saline (PBS), Ca^{2+} and Mg^{2+} free, was purchased from Thomas Scientific. Molecular biology grade water and TE were purchased from Invitrogen. Ethanol was sourced from Decon Laboratories. Neutral buffered formalin was purchased from VWR (16004–128). All reagents, tips, plates, and reservoirs were RNase free.

Methods

Tissue sources

All human tissue was sourced from the University of Arizona Cancer Center Biorepository. Prostate samples were sourced from the UACC Prostate Biorepository. Human samples were consented for research use after clinical testing and de-identified before receipt. Human samples were exempt from IRB approval as per the grant funding source (NCI 5R33CA183688-03 & NIEHS 1R43ES024107-01). The TempO-Seq assay does not sequence patient DNA or RNA, and instead only produces predetermined probe sequence data; therefore, it is not possible to use TempO-Seq sequencing readouts to identify patients. Mouse tissue was provided as a gift from Kathleen Scully and Pamela Itkin-Ansari of the Sanford Burnham Prebys Medical Discovery Institute (La Jolla, Ca.). Rat tissues were obtained from Tissue Acquisition and Cellular/Molecular Analysis Shared Resource at University of Arizona Cancer Center. Rats were euthanized with CO2 asphyxiation in accordance with American Veterinary Medical Association (AVMA) guidelines. The collection of animal samples was approved by IACUC (A3248-01).

Hematoxylin and eosin staining

Slides were deparaffinized using the Leica Bond Dewax solution by soaking in a Coplin staining jar for three minutes. Slides were washed 3 times in 100% ethanol, then either air dried, or continued through H&E staining with the following protocol: rehydrated in distilled water for three minutes; immersed in hematoxylin solution (Leica hematoxylin 560 diluted 1:6 in distilled water) for three minutes; washed three times in distilled water; immersed briefly in 0.1x PBS; washed three more times in distilled water; soaked in 70% Ethanol for two minutes; immersed in Alcoholic Eosin Y with Phloxine (Sigma HT110332) for three minutes; washed in 100% ethanol three times then air dried.

Cell lysates and FFPE pellets

MCF7 and MDA-MB-231 cells were obtained from ATCC and grown in RPMI supplemented with 10% FBS. Fresh lysates were prepared by washing cells with 1X Ca^{2+} and Mg^{2+} free PBS, then lysing in 1X TempO-Seq lysis buffer in PBS at 2,000 cells per μL. Lysates were incubated for 10 minutes at room temperature, followed by storage at -80˚C. For FFPE cell pellets, live cells were washed twice with 1X PBS, and then fixed with 1% formaldehyde in 1X PBS for 30 minutes at room temperature. Cells were then embedded by the TACMASR (University of Arizona Tissue Acquisition and Cellular/Molecular Analysis Shared Resource).

FFPE tissues

All human tissue was sourced from the University of Arizona Cancer Center Biorepository. Prostate samples were sourced from the UACC Prostate Biorepository. Human samples were consented for research use after clinical testing and deidentified before receipt. Archival

samples were stored as FFPE blocks from the year indicated until 2016, at which point a pathologist identified homogenous tumor regions. 5 μm thick sections were then cut and mounted, and slides were stored until 2018. For these experiments, 25 mm^2 areas were cut from serial sections to represent biological replicates.

Mouse tissue was provided as a gift from Kathleen Scully and Pamela Itkin-Ansari of the Sanford Burnham Prebys Medical Discovery Institute (La Jolla, Ca.). All mice were wildtype adults of the C57BL/6 genetic strain. Tissues were fixed in 10% neutral buffered formalin at 4˚C for 24 hours, then moved to 70% EtOH for 24 hours before embedding. 5 μm thick tissue sections were cut and mounted on Superfrost Plus Micro slides (VWR). Slides were dried overnight at room temperature before processing for FFPE TempO-Seq.

Rat tissues were obtained from Tissue Acquisition and Cellular/Molecular Analysis Shared Resource at University of Arizona Cancer Center. Rats were euthanized with CO_2 asphyxiation in accordance with American Veterinary Medical Association (AVMA) guidelines. Tissues were fixed in 10% neutral buffered formalin for 24 hours before being transferred to 70% Ethanol prior to embedding. For fixation time studies, samples were kept in 10% NBF for the specified time before moving to ethanol prior to embedding.

TempO-Seq assay

The TempO-Seq assay for FFPE samples relies on the standard TempO-Seq chemistry [7–11]. The assay (Fig 1) was modified for FFPE samples, and was carried out without modification following the protocol from the User Manual provided with the kits, as described in the following text. An area of interest on a slide-mounted FFPE section (Fig 1A) is scraped from the slide and deposited directly into BioSpyder 1X FFPE lysis buffer (Fig 1B). The sample is then overlaid with molecular biology grade mineral oil and incubated at 95˚C for five minutes to dissolve the paraffin. This separates the paraffin from the lysate without the need for harsh chemicals. The FFPE lysate is further processed by addition of FFPE Protease reagent and incubation at 37˚C for 30 minutes. After a quick homogenization step (trituration by a pipette, or vortexing), the lysates can be frozen or used immediately in the remaining steps of the TempO-Seq FFPE assay [7–11].

As depicted in Fig 1B, a 2 μL aliquot of the processed lysate is then added to a microplate well containing a mix of annealing buffer and Detector Oligos (DOs) to measure each targeted gene. DO panels included in this study were designed against the whole transcriptome for human, mouse, and rat (commercially available assays from BioSpyder, Inc.). This mixture was then exposed to a ramp in temperature from 70˚C to 45˚C, followed by overnight incubation at 45˚C. This facilitates complete annealing of DOs to their target RNAs. The hybridization process is highly resistant to RNA fragmentation (as the DOs anneal to RNA sequences of <100 nt), which facilitates the assay efficiency in the context of FFPE.

A nuclease mix is then added, which degrades unbound and incorrectly bound DOs. Finally, addition of a ligase mixture allows for ligation of correctly bound DOs into full-length probes. The enzymes are then inactivated by a 15 minute incubation at 80˚C, and the resulting ligated probes are amplified in a PCR step. The PCR primers allow indexing of individual samples, so that hundreds or thousands of samples can be multiplexed within the same sequencing library (Fig 1C).

The assays used were the human whole transcriptome assay [7] which measures 19,283 genes (21,111 probes); the mouse whole transcriptome assay which measures 23,580 genes (30,147 probes); and the rat whole transcriptome assay which measures 21,119 genes (22,253 probes). Each gene is measured by one or more probes formed by ligation of a DO pair, as previously described [7]. TempO-Seq probes for the whole transcriptome assays were designed to
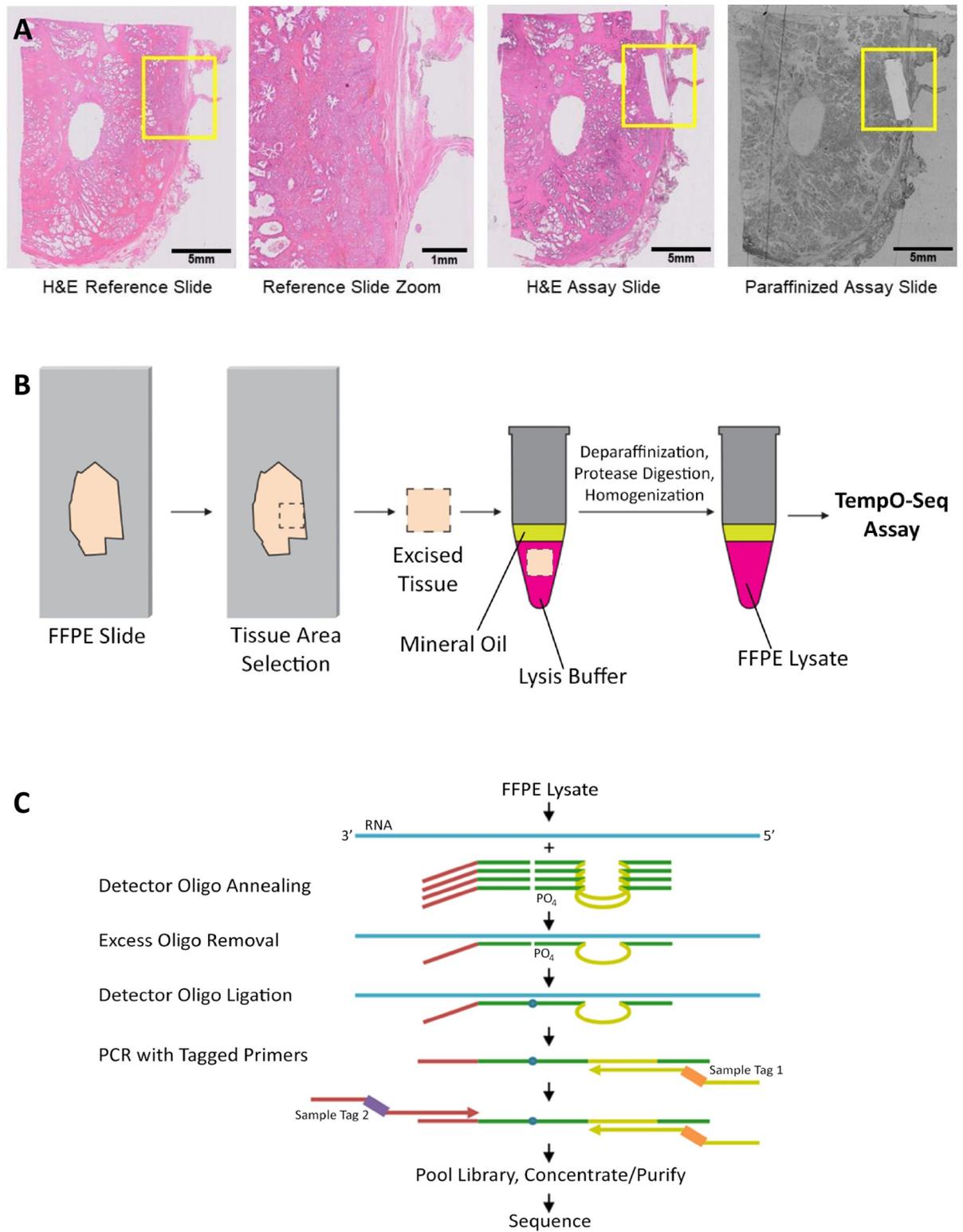
Fig 1. Processing FFPE samples for the TempO-Seq assay. (A) Examples of slides processed through the TempO-Seq FFPE assay. Left panel shows an H&E stained section, with the yellow box indicating the area of interest. Center left shows an expanded image of this area,

demonstrating the mixed histology that would affect the data if the entire area were to be scraped and profiled. Areas identified from a stained tissue section can be scraped from an unstained, paraffinized adjacent section (right), or (if RNase-free reagents are used for staining) directly from a stained section (center right). The scraped areas in this case were ~1 x 5 mm, aligned with the focal histology of interest, and sufficient for gene expression profiling. (B) An area of interest is manually scraped from mounted FFPE sections. The tissue is added directly into 1X FFPE lysis buffer, overlaid with mineral oil, and then heated at 95°C for 5 minutes. FFPE Protease is added and the sample is incubated and manually homogenized. The processed lysate is then ready for input directly into the annealing step of the TempO-Seq assay. (C) Schematic of the TempO-Seq detector oligo annealing and ligation process.

target only protein-coding genes (with few exceptions), so noncoding RNAs are not visible, nor do they take any sequencing reads in this approach. Custom TempO-Seq assays that measure noncoding RNA, splice junctions and variants, fusion genes, etc., have been designed and used in other work, but were not used for the experiments described in this manuscript.

Sequencing and data analysis

Purified libraries were run on the Illumina NextSeq 550 sequencing platform. All data analysis was done using the TempO-SeqR data analysis platform (BioSpyder., Inc.) as follows. After sample demultiplexing using the default Illumina sequencer and bcl2fastq settings, mapped reads were generated by TempO-SeqR alignment of demultiplexed FASTQ files from the sequencer to the ligated DO gene sequences using Bowtie, allowing for up to 2 mismatches in the 50-nucleotide target sequence.

For correlation analysis, genes with 20 or more raw counts were log2 transformed and plotted to derive coefficient of determination (R^2) values. Differential expression was assessed by the TempO-SeqR software which used the DESeq2 method for differential analysis of count data [12]. The count data are first normalized using the DESEq2 function estimateSizeFactor, which establishes size factors using the "median ratio method" described in [13]. DESEq2 then computes the probability of differential expression by comparing the relative count level for each condition and the dispersion of the respective counts using a negative binomial model. A user selected adjusted p-value of <0.05 and baseMean depth >20 were used as the thresholds of significance for differential expression.

For PCA and correlation plots related to colon matched normal and cancer, a pathologist identified regions of cancerous and normal tissue on sections of colorectal cancer. Two within-donor biological replicates were taken from each tissue type from each patient, and lysed. The lysate was used to perform three technical replicate experiments, the results of which were then averaged. Plots represent direct comparisons of all data, without cutoffs, normalized for total read depth.

Raw sequencing data in form of FASTQ files, along with aligned gene counts for all samples used in this study are available through GEO (accession number GSE119630).

Results

Reproducibility and sample types

To verify assay robustness and precision, we tested tissues from a variety of species and a broad range of tissue types. Data shown here includes FFPE samples of human colorectal adenocarcinoma, prostate adenocarcinoma, and pancreatic cancer; rat brain, kidney and liver; and mouse breast, lung, and hindlimb muscle. We chose pancreas due to its relative abundance of endogenous RNases, breast and lung for their low cellularity, and muscle for potential difficulties in digestion by the lysis protocol. For each sample type, 10mm^2 areas from 5μm thick slides were lysed, and 10% of the lysate (equivalent to 1 mm^2 of tissue) was used as input in the whole transcriptome FFPE TempO-Seq assay with species-specific DOs.

To gauge assay reproducibility, the same areas of adjacent 5 μm thick sections were independently processed, and gene expression patterns between replicates compared. It is worth noting that these replicates are biologically different: they represent different subsections of tissue from the same organ, with potentially different tissue composition (microvasculature, innervation, etc.), and are processed completely independently. These are within-donor biological replicates, in which some variance in expression is expected due to the heterogeneity of cellular composition of FFPE. This measures the repeatability of the assay independent of the variability between donors. Technical replicates, such as replicates of the same FFPE tissue lysate, can be used as another measure of assay repeatability, but do not address the variability of the lysis step or resulting from FFPE tissue heterogeneity. For these reasons, it is important to differentiate such "within-donor biological replicates" from "between-donor biological replicates," where samples from different individuals are measured, and where the range of biological variability within a disease can be seen.

Gene expression correlation among within-donor biological replicates for all sample types had R^2 values greater than 0.8, regardless of species. Of the human samples, pancreatic tissue had the highest reproducibility across biological replicates (R^2 = 0.916), (Fig 2). Human colorectal and prostate cancers had R^2 values of 0.872 and 0.885, respectively. Mouse breast, lung, and muscle had R^2 values that exceeded 0.8 (0.891, 0.833 and 0.895, respectively). Rat brain, kidney, and liver all had R^2 values that exceeded 0.9 (0.926, 0.949, and 0.959), (Fig 2). Across all nine tissue types the average R^2 was 0.903. Average %CVs for genes with minimum of 10, 50, or 200 counts were 26.7%, 20.3%, and 16.8%, respectively (Table 1). Larger variance was observed in samples from tissues known to contain large amounts of RNAses (pancreas), and in tissues with low cellularity and thus low RNA amounts (lung, breast).

Detected gene expression profiles match expectations from literature

While no cross-platform comparison can be expected to produce perfect agreement, results should broadly conform to the main parameters–e.g. genes that are highly expressed should be recognized as such independent of platform, and genes that are tissue-specific should not be detected in the wrong tissues. To verify this, we compared expression levels detected by FFPE TempO-Seq in pancreatic tissue with those reported in the Genotype-Tissue Expression (GTEx) database [14]. As shown in Table 2, while gene rankings were not exactly the same, genes recognized as highest expressers in TempO-Seq were also highly ranked in GTEx. Top ranking genes in TempO-Seq are shown along with their ranking in GTEx. Genes recognized as highly expressed in TempO-Seq are also near the top of the rank table in GTEx. The transcript ranking 10 in TempO-Seq data was a mitochondrial target not ranked in GTEx.

Simultaneously, FFPE TempO-Seq recognized expression of pancreas-specific aquaporin AQP12 [15, 16] in pancreatic tissues (Table 3), while the counts were zero in prostate or colon. The same observation can be made at the lower end of the expression range: pancreatic polypeptide PPY, which is annotated in GTEx as a very low expressing pancreas-specific gene, was correctly detected by FFPE TempO-Seq as a low abundance transcript and only detected in pancreas. Similar agreement with established annotation in GTEx can be seen in multiple mid to low-expressing genes, including CLDN10, CFC1, KLB, NPHS1, and TEX11 –all of which are annotated as expressed in pancreas [14], but not in prostate or colon, and which were found to be tissue restricted in FFPE TempO-Seq as well (Table 3).

Expression variability in matched cancer vs. normal tissue samples

The reproducibility shown in Fig 2 validates the precision and reproducibility of the assay when used on separate samples from the same source (same patient or animal, within-donor
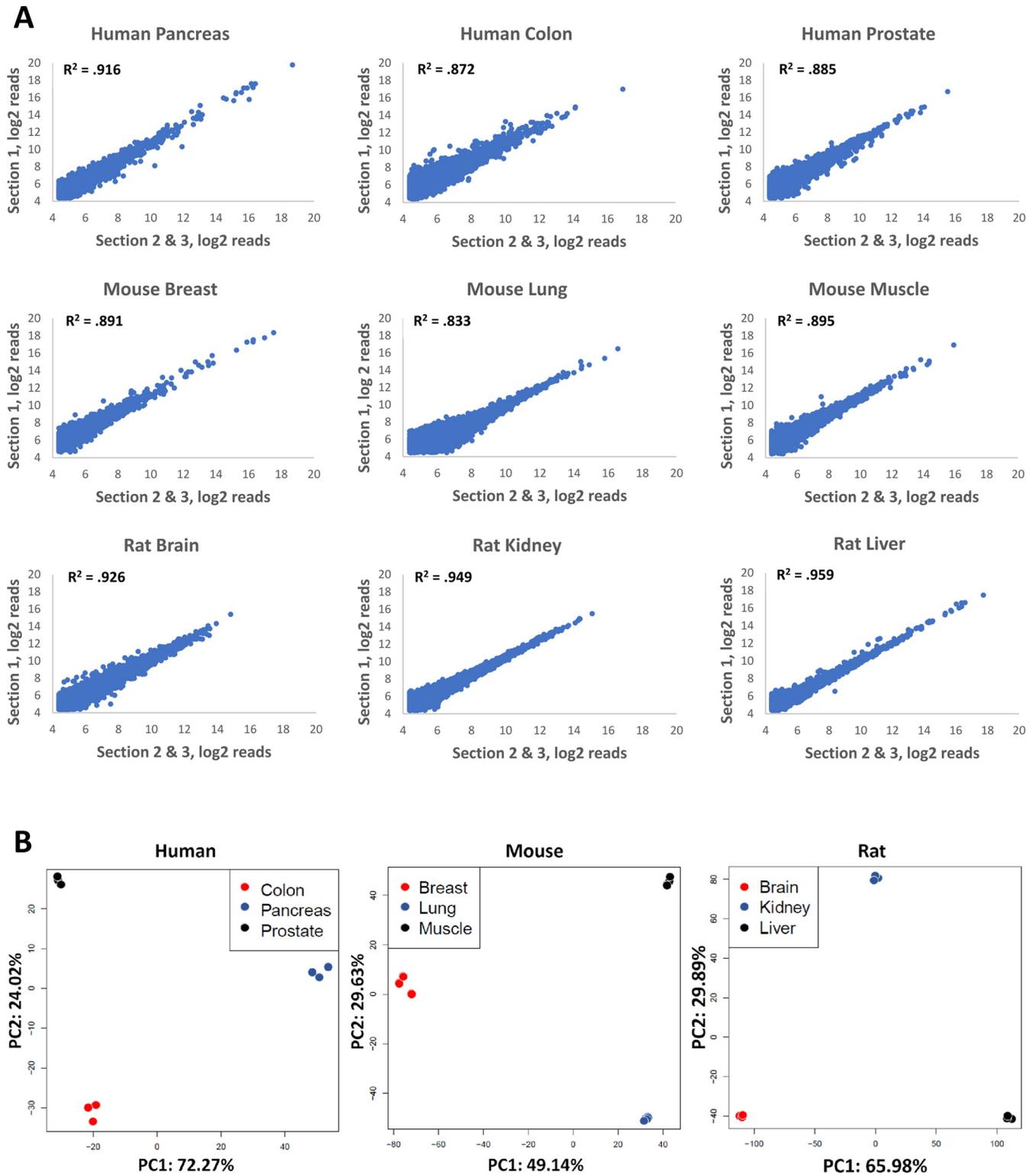
Fig 2. Correlation of gene expression from biological replicates of multiple tissue types. (A) FFPE samples of three different tissue types from human, mouse, and rat were used as input into the TempO-Seq FFPE assay. Replicates were obtained by scraping adjacent areas from three different serial sections. R^2 values were calculated by comparing gene expression of one section to the average of the remaining two. (B) Principal component analysis (PCA) of different tissue types of human (colon, pancreas, and prostate), mouse (breast, lung, and muscle), and rat (brain, kidney, and liver) was conducted. The first two principal components account for the majority of variance in the samples and clearly distinguish the different tissue types.

Table 1. Coefficients of variation observed for genes expressing at a minimum level of 10, 50, or 200 counts.

	>10	>50	>200
Human Pancreas	33.44	29.80	28.88
Human Colon	27.96	22.35	19.23
Human Prostate	27.57	19.96	15.33
Mouse Breast	26.45	19.67	17.01
Mouse Lung	37.72	27.58	20.74
Mouse Muscle	26.06	18.79	15.18
Rat Brain	21.33	15.96	12.98
Rat Kidney	20.62	15.24	11.52
Rat Liver	19.24	13.72	10.73
Average	**26.71**	**20.34**	**16.84**

https://doi.org/10.1371/journal.pone.0212031.t001

biological reproducibility). However, this result does not address the expected variation among multiple individuals (what we define as "between-donor biological replicates"). To address this, we performed the assay on matched samples of human normal and cancerous colon tissue from five different patients.

Fig 3 shows results from principal component analysis (PCA) of gene expression counts. Two within-donor biological replicates of each tissue from each of the five patients were assayed. Each patient is shown in one color, with normal samples shown as circles and cancer samples as triangles. As can be clearly seen, despite the expected diversity in gene expression, normal tissues from all five patients cluster together, with the small spread between them representing between-donor biological variability. The tight clustering of measurements reflects within-donor repeatability and technical repeatability of the assay. In contrast, cancers tend to follow their own trajectories, reflecting genomic instability, stages of progression, and varying pathways of oncogenesis and transcriptional phenotype, or subtypes of cancer. The PCA analysis suggests four subgroups of cancer, with two patients within one of the subgroups, that are clearly distinguished from normal tissues.

To further illustrate and quantify assay reproducibility, correlation plots of within-donor biological samples of normal tissue and are shown (Fig 4A and 4B) for two patients (Patient 3, brown in the Fig 3 PCA plot; and Patient 5, purple in the Fig 3 PCA plot). These two patients exhibited the greatest level of within-donor biological difference in their normal colon tissue expression profiles. Fig 4C and 4D depict the biological repeatability within cancer tissue from the same patients, demonstrating these measurements are also highly repeatable. Fig 4E shows

Table 2. Agreement between FFPE TempO-Seq human whole transcriptome assay data with the GTEx database rankings.

Gene symbol	Gene name	TempO-Seq Rank	GTEx Rank
CTRB1	chymotrypsinogen B1	1	8
PRSS1	serine protease 1	2	1
CPA1	carboxypeptidase A1	3	3
PNLIP	pancreatic lipase	4	5
AMY2A	amylase, alpha 2A (pancreatic)	5	23
CELA3A	chymotrypsin like elastase family member 3A	6	6
CPB1	carboxypeptidase B1	7	10
CEL	carboxyl ester lipase	8	14
CTRB2	chymotrypsinogen B2	9	7
CELA3B	chymotrypsin like elastase family member 3B	11	17

https://doi.org/10.1371/journal.pone.0212031.t002

Table 3. FFPE TempO-Seq counts for genes recognized in GTEx database as pancreas-specific or pancreas-expressing/prostate and colon non-expressing.

Gene	Pancreas average	Colon average	Prostate average
AQP12A	81	0	0
AQP12B	44	0	0
CLDN10	46	0	0
CFC1	45	0	0
NPHS1	32	0	0
PPY	28	0	0
TEX11	26	0	0

https://doi.org/10.1371/journal.pone.0212031.t003

as exemplar data the difference between patient 5 normal and cancer tissue, showing the differentially expressed genes underlying the PCA results depicted in Fig 3, readily identified because of the high repeatability of within-patient measurements. For comparison the largest difference between normal among the patients (between patients 3 and 5) is depicted in Fig 4F. Finally, Fig 4G and 4H show the difference between the cancer profiles for patients 3 and 5, and patients 2 and 4, respectively; demonstrating how different these cancers are, consistent
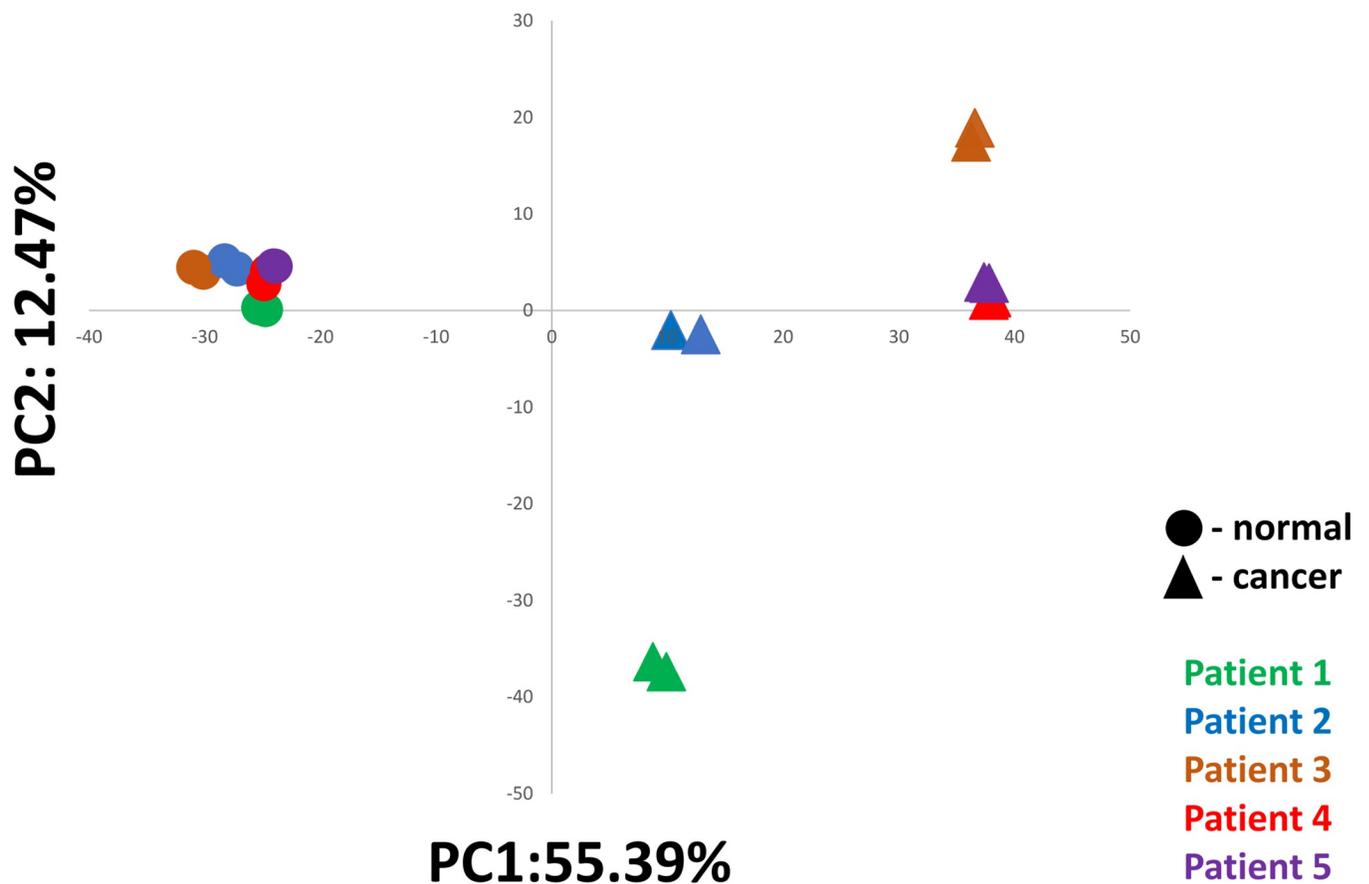

Fig 3. PCA analysis of within and between patient variability. Duplicate biological samples of matched normal and cancerous colon tissue from five anonymized patients were processed through the TempO-Seq FFPE assay. Fig 3 shows principal component analysis of all replicates, with normal tissue samples shown as circles, and cancer samples shown as triangles. Each patient is represented by one color. Despite normal biological diversity of expression, all normal samples cluster together, while cancer samples diverge widely.
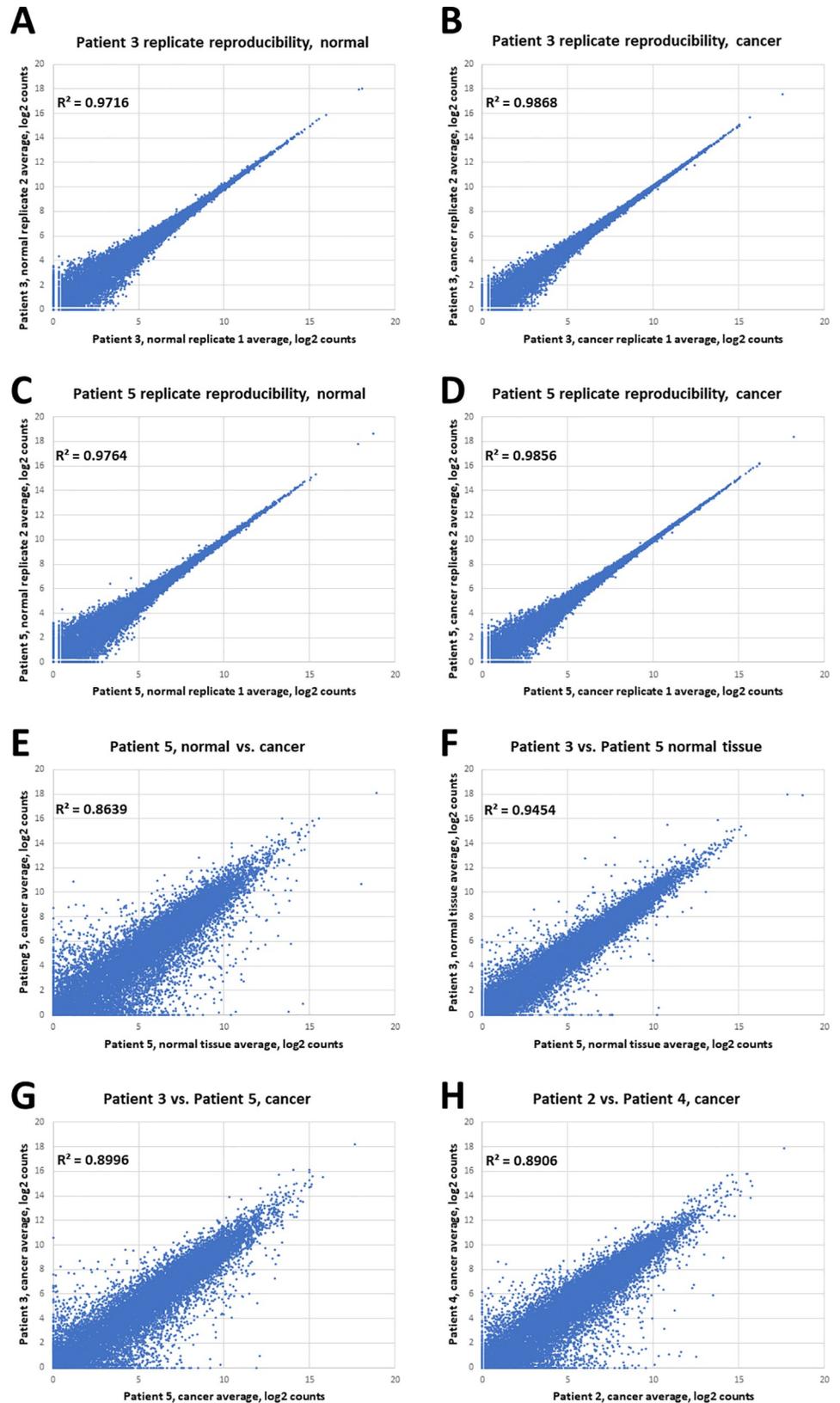
https://doi.org/10.1371/journal.pone.0212031.g003

Fig 4. Expression variability in matched cancer vs. normal tissue. Reproducibility of within-donor biological replicates (average of three technical replicates each) of (A) normal FFPE samples, patient 3 (brown symbols in Fig 3 PCA plot), (B) cancer FFPE samples, patient 3, (C) normal FFPE samples, patient 5 (purple symbols in the Fig 3 PCA plot), and (D) cancer FFPE samples, patient 5. (E) Comparison of normal and cancer samples from one patient (average of two biological replicates each). (F) Comparison of normal tissue between patients 3 and 5 (largest difference in normal tissue expression profiles), showing that differences due to normal gene diversity are far lower than those between cancer and normal. (G) Comparison between cancers of different pairs of patients (Patient 3 vs Patient 5) showing the range of differences. (H) Comparison between cancers of different pairs of patients (Patient 2 vs Patient 4), showing the range of differences that, again, exceeds those seen in normal-to-normal comparison range.

with the PCA analysis. These plots demonstrate the high repeatability within donors that permits robust measurement of differential expression, and also show that the between-donor variability in normal tissue measurements is much less than the variability between cancers from different patients, consistent with expected biological differences and the PCA analysis.

Focal input and sensitivity

Common gene profiling assays generally require RNA purification from large FFPE tissue samples (entire slides or multiple slides). TempO-Seq does not require use of extracted RNA, rather direct sample lysates can be used. Thus, while significant amounts of FFPE are required for RNA extraction, much smaller amounts of FFPE can be assayed as a lysate. The ability to assay very small tissue amounts would spare the use of rare and precious archival FFPE samples, enable profiling of small focal areas with specific pathologies, reduce input for tissues with very low cellularity, and allow profiling of small FFPE samples such as tissue from biopsies or prepared as tissue microarrays.

To evaluate the sensitivity and amount of tissue required for TempO-Seq, we tested areas as small as 2 mm^2 from 5 µm thick tissue sections. The sensitivity measures are given in the format of tissue area/thickness because extraction of RNA from such small samples proved to be extremely technically difficult (with purification losses being prohibitive to analysis). Extractions performed on much larger tissue amounts (whole sections and higher) could not be extrapolated downwards in a meaningful manner to give a valid comparison to using lysate from a 2 mm^2 area, because the percentage of RNA lost from smaller amounts of FFPE are much greater than from larger amounts of FFPE. The lysis buffer volume was scaled accordingly, so for this input, the amount of tissue in the 2 µL volume that is transferred into the assay was the same as for larger tissue excisions.

We excised both 2 mm^2 and 10 mm^2 from 5µm thick mouse liver sections. The correlation of gene expression across biological replicates of the same area excision were similar for 2 mm^2 and 10 mm^2 areas, with the R^2 = 0.969 and 0.95, respectively (Fig 5A and 5B). The correlation between 10 mm^2 and 2 mm^2 inputs was also very good, with R^2 of 0.969 (Fig 5C, average of three samples of each input). These data indicate that the TempO-Seq FFPE assay is highly sensitive and can handle very low input amounts.

Archival tissue

Fixation and paraffin-embedding of tissue allows for long term preservation and storage of samples while retaining useful morphological information. The process of fixation, embedding, and extraction can damage RNA, and long-term storage of such samples can make the damage progressively worse, making gene expression analysis difficult [4, 14]. However, due to the nature of DO hybridization and ligation chemistry and the short length of
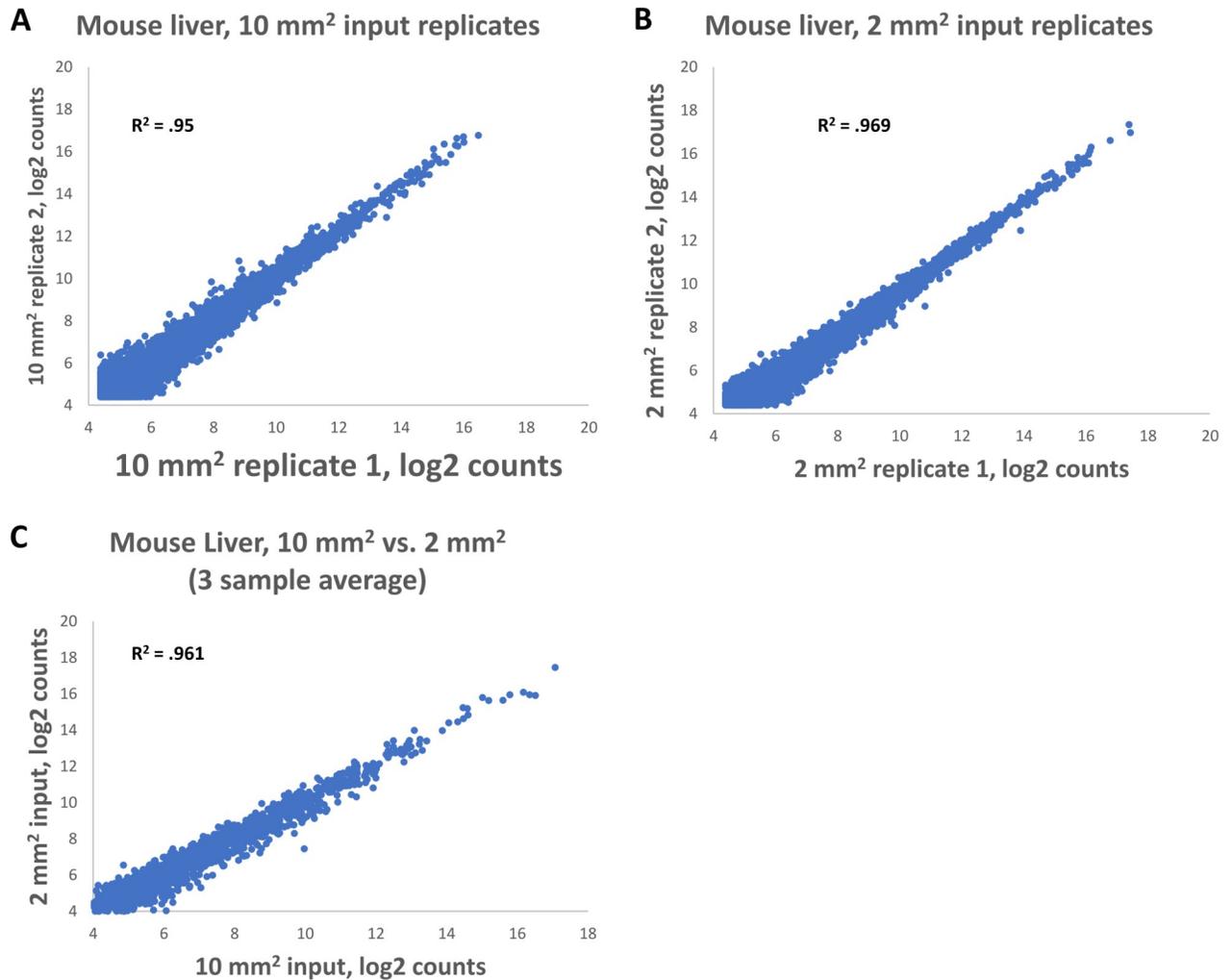
A Mouse liver, 10 mm² input replicates

B Mouse liver, 2 mm² input replicates

C Mouse Liver, 10 mm² vs. 2 mm² (3 sample average)

Fig 5. Replicability of 10 mm² and 2 mm² inputs in the TempO-Seq FFPE assay. Both (A) 10mm² and (B) 2mm² areas were scraped from rat liver. Biological replicates were generated by scraping the same area from adjacent tissue and gene expression correlation was calculated. (C) Gene expression correlation was calculated between 10 mm² and 2 mm² tissue sections from rat liver.

https://doi.org/10.1371/journal.pone.0212031.g005

RNA-Sequence that is targeted by each DO pair, TempO-Seq is highly resistant to this type of fragmentation, as well as to the presence of crosslinking.

To determine if storage time of FFPE blocks had a significant effect on performance of the TempO-Seq FFPE assay, we obtained archival human tumor FFPE samples from the University of Arizona Cancer Center Biorepository. Archival tissues, with their indicated year of harvest, were as follows: colorectal cancer (1986), hepatocellular carcinoma (1993), and two separate cases of kidney cancer (1994 and 1988). Blocks had been stored at room temperature, and in early 2016, were cut into 5μm thick sections. Slides were stored for two years before 25 mm² areas were scraped for TempO-Seq FFPE processing using the human whole transcriptome panel. The same area was cut from serial sections to produce biological replicates. On average, each sample generated 2.1 M mapped reads, which is sufficient for meaningful data analysis (only a 50 base pair region is sequenced and counted for each gene using TempO-Seq, compared to RNA-Seq in which identification of each gene requires sequencing and counting multiple fragments). We compared gene expression data between biological replicates from
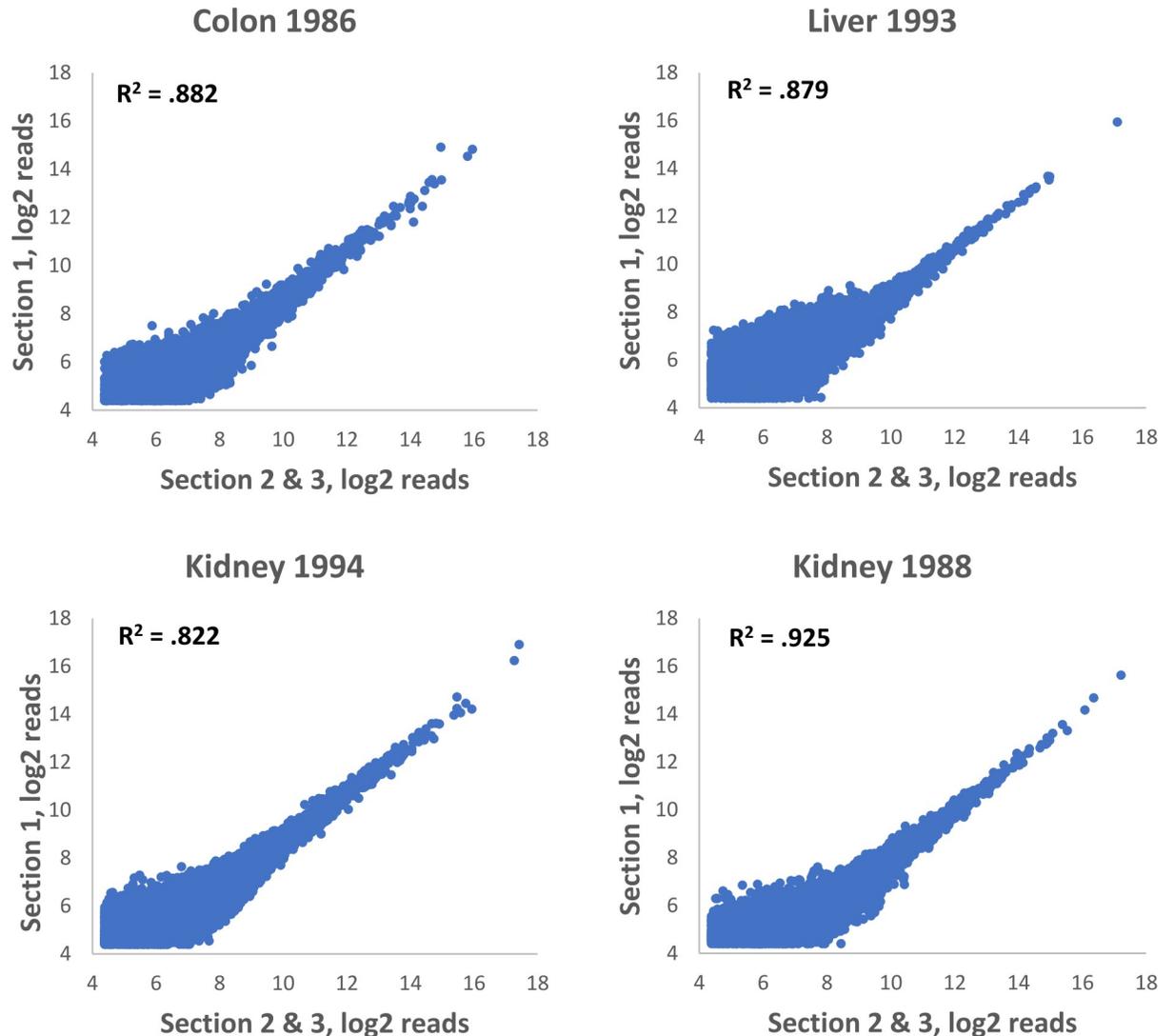
Fig 6. Gene expression correlation in biological replicates of archived human FFPE tissue. Human tissues were harvested in the year indicated and stored as FFPE blocks. Tissue sections were scraped and input into the TempO-Seq FFPE Human Whole Transcriptome assay. Replicates were obtained by scraping adjacent areas from three different serial sections. R^2 values were calculated by comparing one section to the average of the other two.

https://doi.org/10.1371/journal.pone.0212031.g006

the same block as a read out of assay reproducibility. Each of the archival samples had R^2 values of greater than 0.8, with the kidney harvested in 1994 having a within-donor biological replicate R^2 of 0.925 (Fig 6). Although a controlled study demonstrating consistent gene expression profiles in the same samples over decades of storage is not feasible, these results demonstrate that the TempO-Seq FFPE assay can produce robust data from FFPE samples that are more than 30 years old.

Fixation time

The amount of time tissue is exposed to fixative correlates with tissue autolysis and damage caused by endogenous endonucleases. Furthermore, total time of fixation affects RNA integrity, which directly impacts cDNA synthesis from RNA derived from fixed tissues [3, 4]. This

factor can significantly confound gene expression analysis which relies on methods dependent on reverse transcription such as microarrays or RT-PCR. Furthermore, additional fixation time can lead to overfixation, affecting accessibility of RNA [4, 17, 18, 19]. We tested whether fixation time had a notable impact on our assay by harvesting rat liver tissue and incubating in 10% neutral buffered formalin at 4˚C for 24, 96, 192, and 384 hours before embedding. 10mm^2 of tissue was scraped from 5μm thick sections and used as input for the TempO-Seq FFPE assay using rat whole transcriptome DOs.

There was no negative effect on sequencing quality with additional fixation time beyond 24 hours. Gene expression between biological replicates was high: R^2 = 0.96 for 24 hours; 0.93 for 96 hours; 0.95 for 192 hours, and 0.98 for 384 hours (Fig 7). For all fixation times, the observed expression pattern clearly matched that expected for hepatocytes. These data collectively demonstrate that the TempO-Seq assay performs robustly even on samples that have been fixed for extended periods of time.

FFPE samples vs. fresh samples

Since fixation denatures RNA-binding proteins and disrupts secondary structure, TempO-Seq probes may interact with RNA in the context of fixed tissue differently than in fresh tissue, which could affect sensitivity and conclusions drawn from the samples. We compared the TempO-Seq FFPE assay to the standard assay designed for fresh lysates or purified RNA to determine whether processing of FFPE samples may impact biological conclusions. FFPE cell pellets were made from MCF-7 and MDA-MB-231 breast cancer cell lines, derived from luminal A and claudin low subtypes, respectively. 10mm^2 areas were excised from 5 μm thick sections and used as input into the FFPE assay. Cells from the same plate were lysed fresh in lysis buffer and used as input into the standard TempO-Seq assay [7], to minimize variables other than fixation.

We conducted differential gene expression analysis using DESeq2 between the two cell types for both FFPE and fresh assays. Differentially expressed genes were defined as genes with raw counts > 20, and by $p_{adj} < 0.05$. A total of 4,461 genes were detected as differentially expressed in FFPE samples using these cutoffs, compared to 3,015 in fresh lysates. Thus, sensitivity was not reduced by fixation, as the FFPE assay actually detected more genes with lower levels of noise than the fresh lysate assay.

The log2 fold change between the two sample types showed a strong correlation (R^2 = 0.970), (Fig 8A). Literature and previous gene expression data for these two cell types agree with genes detected by TempO-Seq as differentially expressed in both sample types [7]. This shows that fixation does not significantly distort the underlying biological data.

Comparison to RNA-Seq

While these data show that the TempO-Seq direct lysis assay of FFPE is reproducible, precise, and sensitive, the question of accuracy remains: how well do these measures reflect biological reality? RNA-Seq is a method which depends on purification and reverse-transcription of RNA (both of which can introduce artifacts), and thus is not a perfect measure of biological reality. However, it has become the gold standard for measuring gene expression changes, and thus represents a sufficiently valid baseline measure for comparison.

We compared MCF-7 vs. MDA-MB-231 FFPE cell pellet results with previously published RNA-Seq data [7] which measured log2FoldChange differences between RNA purified from the same cell types (Fig 8B). These cell lines are well characterized, and the gene expression differences between them is expected and well understood. The data comparison was performed for genes with >20 counts, and whose expression was determined to be significantly different
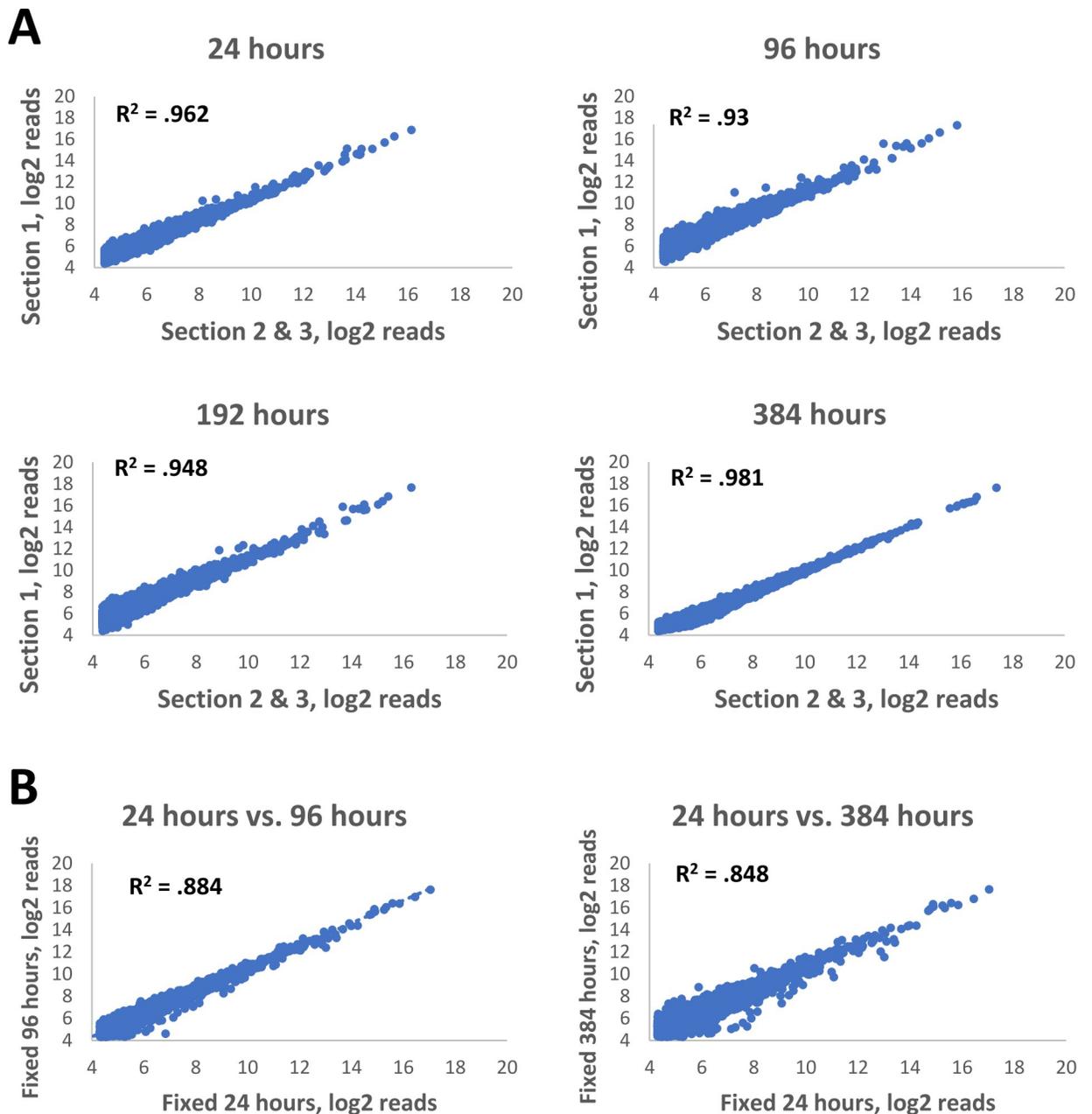
A

B

Fig 7. Gene expression correlation in biological replicates of rat FFPE tissues fixed for variable amounts of time. (A) Rat liver tissue was harvested and fixed in 10% neutral buffered formalin for 24, 96, 192, and 384 hours. Samples were then moved into 70% ethanol and processed for FFPE and used as input into the TempO-Seq FFPE rat whole transcriptome assay. Replicates were obtained by scraping adjacent areas from three different serial sections. R^2 values were calculated by comparing one section to the average of the remaining two. (B) Comparison between 24-hour fixation and 96 or 384 hour fixation. R^2 values were calculated between averages of three biological replicates for each condition.

https://doi.org/10.1371/journal.pone.0212031.g007

by DESeq2 (p_{adj} <0.05). The agreement of log2FoldChange measures is excellent (Fig 8B) for a cross-platform comparison (R^2 = 0.84), especially when considering sample differences (whole cell FFPE lysate vs. purified, reverse transcribed RNA). By comparison, the SEQC study reported a Pearson correlation between RNA-Seq and Affymetrix microarrays of 0.89, which is an R^2 of 0.79 [20]. A correlation of R^2 = 0.849 was reported between RNA-Seq and Illumina
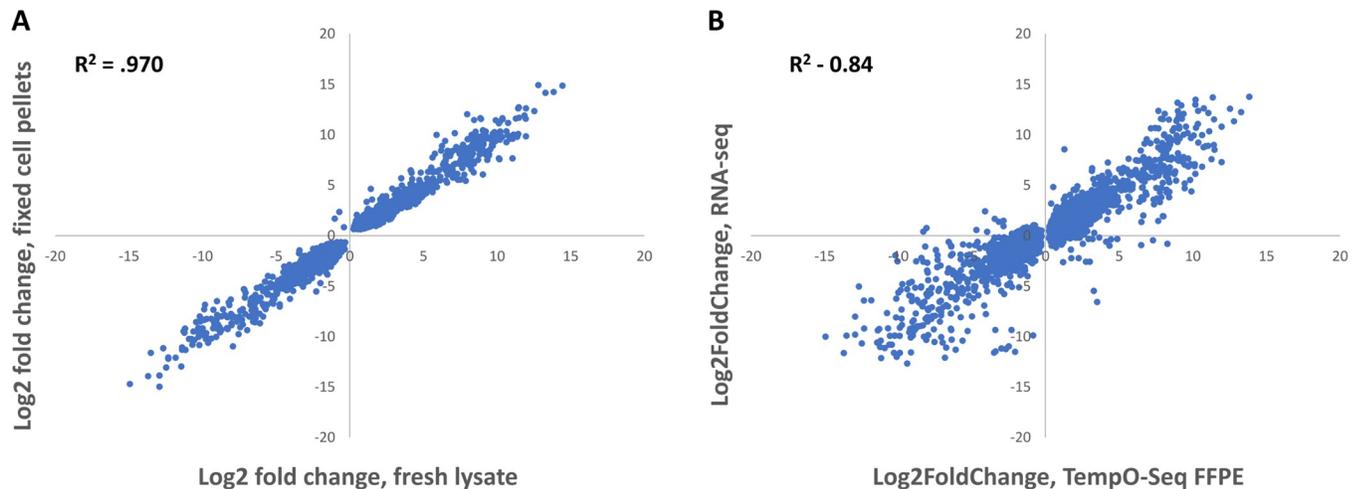
Fig 8. Correlation of gene log2 fold changes between fresh and fixed MCF-7 and MDA-MDA-MB-231 cells, and correlation of TempO-Seq with RNA-Seq. (A) Both MCF-7 and MDA-MB-231 cells from the same plate were split in half. One half was lysed directly and processed through the standard TempO-Seq human whole transcriptome assay (fresh lysate), and the other half was pelleted and fixed for FFPE embedding and sectioning to be processed through the TempO-Seq FFPE human whole transcriptome assay (fixed cell pellet). For each assay, differential expression was calculated between the two cell types. Log2 fold changes were then plotted for each assay to determine correlation of differential expression. (B) RNA-Seq was performed on RNA purified from fresh MCF-7 and MDA-MB-231 cells. TempO-Seq FFPE assay was performed on FFPE cell pellets produced from the same cell types. DESeq2 was used to determine statistically significant ($p_{adj} < 0.05$) log2 fold changes between the different cell types and plotted to derive correlation of differential gene expression between the two methods.

microarrays [21] using reference RNA samples, while significantly lower values of 0.757–0.774 were reported for cell line samples [22].

Stained tissue

Hematoxylin and eosin (H&E) staining of FFPE sections is a practice commonly used for histopathological interpretation of tissue samples and can be used to identify a wide variety of diagnostically relevant features including cellular organization, nuclear morphology, and lymphocytic invasion [23]. The H&E staining process requires deparaffinization and rehydration of tissue slides before staining. Samples that are rehydrated risk hydrolysis of RNA molecules and exposure to RNases, in addition to RNA degradation that can occur due to relatively high acidity of the staining process. Therefore, current practice is to prepare an H&E stained section and then process an adjacent unstained section using the H&E section as a guide. This is fine so long as there is sufficient material and the whole slide is being processed. However, if only a focal area is of interest because of its histology within the section, then marking slides accurately based on a serial H&E stained section can be problematic, particularly if the area is very small. Therefore, we pursued the possibility of profiling the H&E stained slide itself, so that the area scraped could be directly visualized and documented.

To test whether H&E staining would interfere with TempO-Seq, we used a set of 5 μm thick serial human prostate cancer slides. These slides were either processed directly, deparaffinized and processed, or deparaffinized then H&E stained and processed (Fig 9). RNase-free reagents were used for deparaffinization and staining. The same 5 mm^2 area of homogenous tumor was scraped from each serial section and lysed, with 2 mm^2 equivalent of the resulting lysate used as input into the human whole transcriptome TempO-Seq FFPE assay. Gene expression profiles between paraffinized and deparaffinized sections had a high correlation ($R^2 = 0.902$), indicating that the method of paraffin removal had little effect on assay performance. The R^2 value between deparaffinized and H&E stained was also high ($R^2 = 0.855$), demonstrating that the
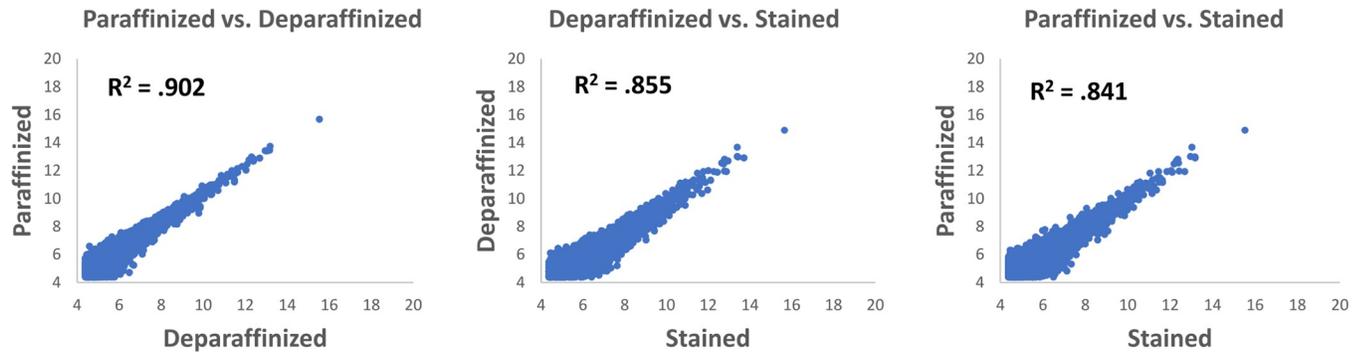
Fig 9. Gene expression correlation of human prostate FFPE samples that were unstained, deparaffinized, or deparaffinized and H&E stained. Serial sections were scraped and used as input into the TempO-Seq FFPE assay for human whole transcriptome. Gene expression was compared among paraffinized (sectioned and mounted, but not otherwise treated), deparaffinized, and deparaffinized then stained conditions.

https://doi.org/10.1371/journal.pone.0212031.g009

assay still worked well with RNA exposed to the H&E chemistry. Gene expression signatures from H&E stained tissue also correlated well with unstained sections ($R^2 = 0.841$). Overall, these data demonstrate that H&E stained FFPE tissue can serve as input for the TempO-Seq assay, with the note that we were careful to use RNase free reagents in the H&E staining process. This also further validates the ability of the assay to detect significantly degraded RNA within samples (Fig 9).

Discussion

The processing required for fixation and paraffin embedding of tissues tends to fragment nucleic acids, which presents significant obstacles to molecular analysis. This is particularly the case for RNA and measurement of gene expression. However, FFPE samples also preserve tissue morphology well over long periods of time, and are easy to handle and section, factors that make them extremely useful for pathology and long-term storage. Additionally, after the initial damage induced by the process itself, fixation and embedding provide significant protection from further damage and hydrolysis, without the need for expensive or cumbersome measures such as snap-freezing and keeping samples constantly frozen for years or decades. These advantages have led to accumulation of vast archives of annotated FFPE tissues which have until now been difficult or impossible to profile at the molecular level.

TempO-Seq [4, 7–11] is a targeted, ligation-based assay designed to minimize complexities usually associated with gene expression measurements. The targeted approach provides several advantages–since it processes and counts only specific pre-determined probe sequences, it avoids cumbersome bioinformatics (the output of the assay is a simple table of counts for each gene in each sample) and reduces sequencing costs significantly (to 1/10th or less). The obvious downside is that all non-targeted sequences will be invisible (although a probe can be made for almost any target). Additionally, the assay can be performed in any lab, requiring no specialized equipment beyond a thermocycler and access to a sequencing instrument (commercially, or in most university core facilities). Critically for the purpose of FFPE evaluation, TempO-Seq does not rely on RNA extraction and reverse transcription, which makes it relatively insensitive to fragmentation, as the probes can successfully bind to RNA targets that are <100 nt in length.

In this study, we provide data demonstrating the quality and reproducibility (Fig 2 and Table 1) of the TempO-Seq assay of a variety of FFPE tissue samples across three different species (human, mouse, and rat). The gene expression data agrees well with existing data from the

literature (Tables 2 and 3), and is highly sensitive, producing excellent gene expression read-outs from tissue inputs as small as 2 mm^2 areas of a 5 µm section (Fig 5). In comparison, most other methods require sacrifice of entire sections (or multiple sections) to obtain sufficient extracted RNA for a single attempt at measurement. This level of sensitivity is critical when dealing with precious and irreplaceable archival samples.

The assay is also insensitive to the time samples are stored (Fig 6), although some caveats apply. Namely, while we demonstrate the ability of the assay to detect highly precise and repro-ducible gene expression information from decades-old samples, the data presented here does not address the question of how good the preservation of biological information is in FFPE over such time periods. Further studies will be needed to confirm the full validity of such data-sets, especially when compared to matched frozen tissues. One variable that can be eliminated from consideration is the time of fixation (Fig 7), which does not significantly affect the assay results.

Equivalency between recently fixed and fresh samples was demonstrated in Fig 8A, where the data showed an excellent correlation (R^2 = 0.97). This demonstrates that fixation does not produce large data distortions, or other immediate problems for further analysis. This is true not only within the TempO-Seq platform, but as Fig 8B demonstrates, between differential expression measured using RNA-Seq and the TempO-Seq platform, producing a "between platform" R^2 = 0.84. This is notable, particularly considering that typical between-platform correlations profile the same sample (e.g. aliquots of same extracted RNA); while in this case we compared the RNA-Seq differential expression of RNA extracted from unfixed cells to the whole transcriptome TempO-Seq FFPE assay of cell pellets after they were fixed, embedded in paraffin, and then sectioned before assay. This combined cross-platform and cross-methodol-ogy consistency demonstrates that results from this assay are likely to reflect the true expres-sion profile of any assayed sample.

While the cross-platform consistency is good, additional sources of variation besides the platform differences may contribute to the observed R^2 value of 0.84. Two are immediately identifiable: first, the RNA-Seq data was derived using purified RNA from unfixed cells col-lected two years prior to growing and fixing the cells for our FFPE experiments, which means these are not identical samples. Secondly, purified RNA from unfixed cells will be different from that from fixed cells, and different in fragmentation state quality from RNA present in an unpurified lysate of FFPE. Thus, the comparison includes both the variability of cross-platform differences as well as between-sample differences. A cross-platform comparison between FFPE TempO-Seq and RNA-Seq performed on RNA purified from FFPE would be more closely equivalent. However, our attempts to purify RNA from fixed cell pellets never reached the minimum quality required for good performance in RNA-Seq, leading us to conclude that use of high quality RNA isolated from unfixed cells for RNA-Seq provided a more relevant com-parison to TempO-Seq assay of FFPE lysates then would use of low quality RNA isolated from FFPE.

The observed sensitivity to assay small areas of FFPE (Fig 5) becomes especially valuable when coupled with data showing that TempO-Seq can be performed on H&E stained tissues (Figs 1A and 9), as long as the staining is performed using RNase-free reagents. In practice, this means individual tissue sections can be stained, and the staining used to determine pre-cisely delimited areas of tissue to be profiled (e.g. separating epithelial cells from background, or stromal tissue from glands, etc.). Examples of H&E stained tissue are shown in Fig 1A, where the heterogeneity of the FFPE is evident (upper right panel) as well as the consistency of histology within the small scraped area (lower left and right panels). It is notable that such small areas can be profiled quickly and easily by hand, whereas existing alternatives that require the extraction of RNA are expensive, laborious, and typically require much more tissue

(and where small amounts of tissue are assayed, such as with laser capture microdissection, required specialized expensive hardware and complex procedures). Thus if 1 mm^2 spatial resolution is sufficient, TempO-Seq provides a high sample throughput, highly repeatable, simple solution for profiling FFPE, even after archiving for a long period of time. This approach should enable investigators to obtain highly histology-specific gene expression data to delineate not only disease states but also the complex interactions between cell types and histologies within a tissue.

The between donor biological repeatability of the TempO-Seq assay revealed an important observation: the biological variability of normal colorectal tissue is quite low in comparison to variability between cancer samples, demonstrating that biological differences in disease phenotype between patients can be clearly seen above any assay variability. Normal samples clustered tightly and very differently from cancer, showing that the regularly seen "noise" in expression profiles due to biological differences between the normal tissue of donors is easily distinguishable from disease processes inherent to oncogenesis. Thus, using a larger cohort of patients, it should be possible to identify phenotypic subtypes or molecular signatures among colon cancer patients.

The combined sensitivity, robustness, and consistency of expression profiling permit the TempO-Seq FFPE assay to be used over a wide variety of applications which would not otherwise be possible. By enabling the assay of many samples and study designs which were previously very technically difficult (or sometimes outright impossible), we believe that whole transcriptome profiling using the TempO-Seq assay of FFPE samples will lead to significant advancements in many fields of biological science.

Acknowledgments

Author Contributions

Conceptualization: Christy L. Trejo, Miloš Babić, Elliot Imler, Joanne M. Yeakley, Bruce E. Seligmann.

Data curation: Christy L. Trejo, Miloš Babić, Elliot Imler, Peter J. Shepard.

Formal analysis: Christy L. Trejo, Miloš Babić, Elliot Imler, Peter J. Shepard, Joanne M. Yeakley, Bruce E. Seligmann.

Funding acquisition: Joanne M. Yeakley, Bruce E. Seligmann.

Investigation: Christy L. Trejo, Elliot Imler, Migdalia Gonzalez, Sergei I. Bibikov.

Methodology: Christy L. Trejo, Miloš Babić, Elliot Imler, Joanne M. Yeakley, Bruce E. Seligmann.

Project administration: Christy L. Trejo, Joanne M. Yeakley, Bruce E. Seligmann.

Resources: Harper C. VanSteenhouse.

Software: Peter J. Shepard.

Supervision: Peter J. Shepard, Joanne M. Yeakley, Bruce E. Seligmann.

Validation: Christy L. Trejo, Miloš Babić, Harper C. VanSteenhouse, Joanne M. Yeakley.

Visualization: Christy L. Trejo, Miloš Babić, Peter J. Shepard.

Writing – original draft: Christy L. Trejo, Miloš Babić, Bruce E. Seligmann.

Writing – review & editing: Christy L. Trejo, Miloš Babić, Elliot Imler, Peter J. Shepard, Harper C. VanSteenhouse, Joanne M. Yeakley, Bruce E. Seligmann.

References

1. Esteve-Codina A, Arpi O, Martinez-García M, Pineda E, Mallo M, Gut M, et al. A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. PLoS One. 2017 Jan 25; 12(1):e0170632. https://doi.org/10.1371/journal.pone.0170632 PMID: 28122052

2. Lüder Ripoli F, Mohr A, Conradine Hammer S, Willenbrock S, Hewicker-Trautwein M, Hennecke S, Murua Escobar H, Nolte I. A Comparison of Fresh Frozen vs. Formalin-Fixed, Paraffin-Embedded Specimens of Canine Mammary Tumors via Branched-DNA Assay. Int J Mol Sci. 2016 May 13; 17(5). pii: E724. https://doi.org/10.3390/ijms17050724 PMID: 27187374

3. Kashofer K, Viertler C, Pichler M, Zatloukal K. Quality Control of RNA Preservation and Extraction from Paraffin-Embedded Tissue: Implications for RT-PCR and Microarray Analysis. PLoS One 2013 Jul 31; 8(7):e70714. https://doi.org/10.1371/journal.pone.0070714 PMID: 23936242

4. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA Quality from FFPE Samples. PLoS One. 2007 Dec 5; 2(12):e1261. https://doi.org/10.1371/journal.pone.0001261 PMID: 18060057

5. Castiglione F, Degl'Innocenti RD, Taddei A, Garbini F, Buccoliero AM, Raspollini MR, et al. Real-time PCR analysis of RNA extracted from formalin-fixed and paraffin-embeded tissues: effects of the fixation on outcome reliability. Appl Immunohistochem Mol Morphol. 2007 Sep; 15(3):338–42. https://doi.org/10.1097/01.pai.0000213119.81343.7b PMID: 17721281

6. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, et al. 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. 2014 May 30; 9(5):e98187. https://doi.org/10.1371/journal.pone.0098187 PMID: 24878701

7. Yeakley JM, Shepard PJ, Goyena DE, VanSteenhouse HC, McComb JD, Seligmann BE. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. PLoS One 2017 May 25; 12(5):e0178302. https://doi.org/10.1371/journal.pone.0178302 PMID: 28542535

8. Grimm FA, Iwata Y, Sirenko O, Chappell GA, Wright FA, Reif DM, et al. Chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. Green Chem. 2016; 18:4407–4419. https://doi.org/10.1039/c6gc01147k PMID: 28035192

9. Rooney JP, Ryan N, Chorley BN, Hester SD, Kenyon EM, Schmid JE, et al. Genomic effects of androstenedione and sex-specific liver cancer susceptibility in mice. Toxicol. Sci. 2017; 160:15–29. https://doi.org/10.1093/toxsci/kfx153 PMID: 28973534

10. House JS, Grimm FA, Jima DD, Zhou YH, Rusyn I, Wright FA. A pipeline for high throughput concentration response modeling of gene expression for toxicogenomics. Front. Genet. 2017; 8:168. https://doi.org/10.3389/fgene.2017.00168 PMID: 29163636

11. Grimm FA, Blanchette A, House JS, Ferguson K, Hsieh NH, Dalaijamts C, et al. A human population-based organotypic in vitro model for cardiotoxicity screening. ALTEX. 2018 Jul 8; https://doi.org/10.14573/altex.1805301 PMID: 29999168

12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

13. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11 (10):R106. https://doi.org/10.1186/gb-2010-11-10-r106 PMID: 20979621

14. Consortium GTEx. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun; 45(6):580–5. https://doi.org/10.1038/ng.2653 PMID: 23715323

15. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based

proteomics. Mol Cell Proteomics. 2014 Feb; 13(2):397–406. https://doi.org/10.1074/mcp.M113.035600 PMID: 24309898

16. Itoh T, Rai T, Kuwahara M, Ko SB, Uchida S, Sasaki S, Ishibashi K. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. Biochem Biophys Res Commun. 2005 May 13; 330 (3):832–8.–insert at 16 https://doi.org/10.1016/j.bbrc.2005.03.046 PMID: 15809071

17. Chung JY, Cho H, Hewitt SM. The paraffin-embedded RNA metric (PERM) for RNA isolated from formalin-fixed, paraffin-embedded tissue. Biotechniques. 2016 May 1; 60(5):239–44. https://doi.org/10.2144/000114415 PMID: 27177816

18. Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, Jordan RC. Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. Mod Pathol. 2002 Sep; 15(9):979–87. https://doi.org/10.1097/01.MP.0000026054.62220.FC PMID: 12218216

19. Cronin M, Pho M, Dutta D, Stephans JC, Shak S, Kiefer MC, Esteban JM, Baker JB. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol. 2004 Jan; 164(1):35–42. https://doi.org/10.1016/S0002-9440(10)63093-3 PMID: 14695316

20. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol. 2014 Sep; 32 (9):903–14. https://doi.org/10.1038/nbt.2957 PMID: 25150838

21. Illumina White Paper: Sequencing, Pub. No. 470-2011-004, 12 April 2011

22. Wolff A, Bayerlová M, Gaedcke J, Kube D, Beißbarth T. A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. PLoS One. 2018 May 16; 13(5):e0197162. https://doi.org/10.1371/journal.pone.0197162 PMID: 29768462

23. Feldman AT, Wolfe D. Tissue processing and hematoxylin and eosin staining. Methods Mol Biol. 2014; 1180:31–43. https://doi.org/10.1007/978-1-4939-1050-2_3 PMID: 25015141