Whole-transcriptome analysis of FFPE samples using Clariom D Assays

Key observations

- Streamlined workflow and extensive transcriptome coverage of Applied Biosystems™ Clariom™ D Assays make them an ideal choice for whole-transcriptome analysis of FFPE samples
- Applied Biosystems™ Transcriptome Analysis Console (TAC) 4.0 Software is an easy-to-use software that allows analysis of whole-transcriptome microarray data, including alternative splicing events
- Applied Biosystems™ TaqMan® Gene Expression Assays confirm differential gene expression as well as alternative splicing events from microarray expression studies in FFPE samples

Introduction

Formalin-fixed, paraffin-embedded (FFPE) tissues represent a large repository of clinically relevant samples that are routinely archived during disease studies. These tissue samples can be linked to pathological outcomes and are therefore useful for long-term follow-up analyses. FFPE samples are hence considered extremely important, especially in retrospective studies [1]. Preservation of tissues by formalin fixation followed by embedment in paraffin has long been the method of choice because this process largely preserves the morphological features of tissues. The process of formalin fixation, however, chemically modifies and partially degrades DNA and RNA, making downstream molecular analysis of these samples a challenge. Even DNA and RNA extracted from well-preserved FFPE samples are limited in quantity and quality, making whole-transcriptome analysis of FFPE samples difficult. Nevertheless, it is clear that when tissues are properly fixed, it is possible to get valuable molecular data from these samples even after several decades of preservation [1,2].

Whole-transcriptome analyses have begun to unravel the complexity of transcripts that are present in tissues and cells. One of the ways this complexity arises is through differential splicing, which can generate a large number of mRNA and protein isoforms from a single gene. It has been reported that more than 90% of all genes undergo some form of alternative splicing [3]. Alternative splicing events and resulting mRNA and protein isoforms have been found to be associated with conditions such as Alzheimer's disease, cystic fibrosis, many cancers, and heritable disorders [4,5]. Understanding alternative mRNA splicing outcomes provides an opportunity to identify these sequences as biomarkers for pathological states [5]. It is therefore important to include alternative splicing events as part of any whole-transcriptome analysis.

Long noncoding RNAs, or lncRNAs, are a class of large and diverse noncoding RNA molecules that are greater than 200 bp in length. lncRNAs have been found to be either repressed or induced in many disease conditions, implying their role in a wide range of biological processes. More recently, a large number of lncRNAs were also found to be associated with modulation of alternative splicing events [6]. The exact role of lncRNAs is not yet clear, but because they constitute a large and important percentage of the transcriptome, their analysis should be an integral part of whole-transcriptome analysis.

ThermoFisher SCIENTIFIC

In this study, we demonstrate a whole-transcriptome analysis of FFPE samples using the Applied Biosystems™ GeneChip™ WT Pico Kit, Clariom D Assays, and TAC 4.0 Software (Figure 1). The GeneChip WT Pico Kit enables whole-transcriptome analysis of FFPE samples on Clariom D Assays with as little as 500 pg of total RNA. Clariom D Assays use an updated design to provide unprecedented coverage of transcriptome-wide gene- and exon-level expression profiles. Sequences on Clariom D Assays are built using the latest transcriptome knowledge from multiple databases. These arrays cover more than 500,000 transcripts and include probes that detect alternative splicing events of coding and lncRNA. TAC 4.0 Software enables analysis of transcriptomic microarray data using a variety of statistical, visualization, and quality control tools. One of the unique features of TAC 4.0 Software is that it integrates differential splicing analysis into the interpretation of microarray data. TAC 4.0 Software is designed for end users and offers meaningful insights into transcriptomic data in a very short period of time.

A complete workflow for whole-transcriptome analysis of archived samples

We illustrate how a retrospective whole-transcriptome analysis could be performed by making use of archived tumor samples from different tissues. Three samples each from brain, kidney, lung, ovary, and stomach tumor biopsies were purchased from BioIVT. We extracted total RNA and DNA from these samples using the Invitrogen™ RecoverAll™ Multi-Sample RNA/DNA Isolation Kit. Extracted RNA was initially quantified using the Invitrogen™ Qubit™ 4 Fluorometer.

Since RNA is often degraded in FFPE samples, fluorometric or spectrophotometric techniques only provide partial insight into sample quality and suitability for whole-transcriptome analysis. We therefore developed a method based on RT-qPCR to more reliably quantitate amplifiable RNA extracted from FFPE samples. Briefly, total RNA extracted from each FFPE sample was reverse-transcribed using the Invitrogen™ Superscript™ VILO™ cDNA Synthesis Kit. 2 µL of cDNA was then used in a qPCR reaction with a TaqMan Gene Expression Assay targeting 18S rRNA. Simultaneously, a standard curve was generated with C_t values corresponding to 18S rRNA from 6 known concentrations of Universal Human Reference RNA (Agilent Technologies, Cat. No. 740000). The concentration of amplifiable RNA was determined by plotting the C_t values obtained from the FFPE samples on the standard curve for Universal Human Reference RNA.

Targets for hybridization to Clariom D Assays were prepared using the GeneChip WT Pico Kit according to the kit instructions. 50 ng of amplifiable RNA from 3 samples each from brain, kidney, lung, ovary, and stomach tumor biopsies was first converted to cDNA, followed by *in vitro* transcription to make cRNA. Biotinylated cRNA (15 samples total) and controls were hybridized to Clariom D Assays for 16–18 hr at 45°C. Standard posthybridization washing and staining were done on the Applied Biosystems™ GeneChip™ Fluidics Station 450, followed by scanning on the Applied Biosystems™ GeneChip™ Scanner 3000 7G. Data were analyzed using TAC 4.0 Software.

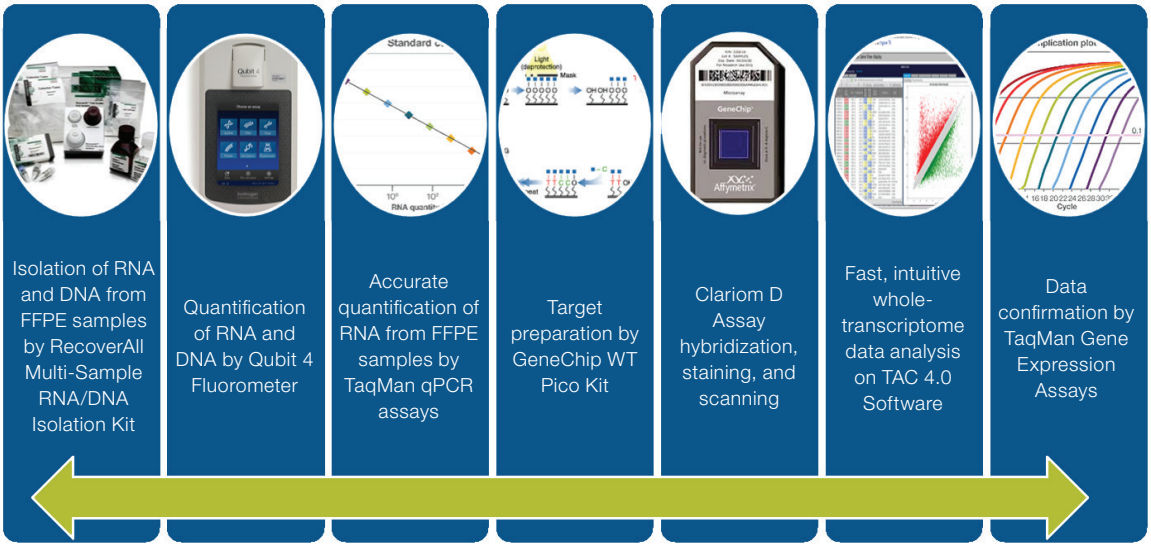

Figure 1. Complete workflow for whole-transcriptome analysis of FFPE samples. Thermo Fisher Scientific offers the necessary reagents, tools, and expertise for studying gene expression in FFPE samples.

Tissue- and tumor-specific analysis of gene expression

Our first test was to verify that transcriptomic results obtained with Clariom D Assays recapitulate known biological aspects of these samples. To analyze differences in gene expression between the different FFPE tissues, expression signals from the 15 Clariom D Assays were analyzed on TAC 4.0 Software. This flexible and powerful yet intuitive data analysis software provides insight into statistically significant expression data across multiple samples within a few minutes. TAC 4.0 Software was used to normalize signal intensities of probes and identify differentially expressed genes (DEGs) between all 15 samples [7]. All 3 replicates of each tissue type clustered together, indicating similar DEGs among samples belonging to the same tissue types (Figure 2). Because gene expression is known to be not only tissue specific but also tumor specific, we confirmed the tissue and tumor specificity of genes in this set of FFPE samples by comparing DEGs in our samples to previously published data. Consistent with data published in the Human Protein Atlas [8], we found *LIPF*, *GIF*, *GKN1*, and *PGA4* to have higher transcript levels in stomach tumor samples (Figure 3 and Table 1). Similarly, we confirmed that *SLC1A2* and *GFAP* were overrepresented in brain tumor samples, *GIPC2* and *SLC16A4* had higher transcript levels in kidney tumor samples, *IGHG4* and *SCGB3A2* had higher transcript levels in lung tumor samples, and *CAS4* and *MEIS1* transcripts were present at elevated levels in ovary tumor samples. In addition, genes associated with tissue-specific tumors showed tissue-specific overrepresentation in our sample set [8]. *SLC1A2* was overrepresented in brain tumor samples, *TMEM91* had higher transcript levels in kidney tumor samples, *PVRL1* had higher transcript levels in lung tumor samples, *VTCN1* and *THSD4* transcripts were present at elevated levels in ovary tumor samples, and *SLC12A2* and *SLC12A1* had higher transcript levels in stomach tumor samples. We also found that genes commonly associated with any tumor type such as *MLLT3*, *ESR1*, and *SFTB* had higher transcript levels in all 5 samples. Representative data are shown in Table 1.

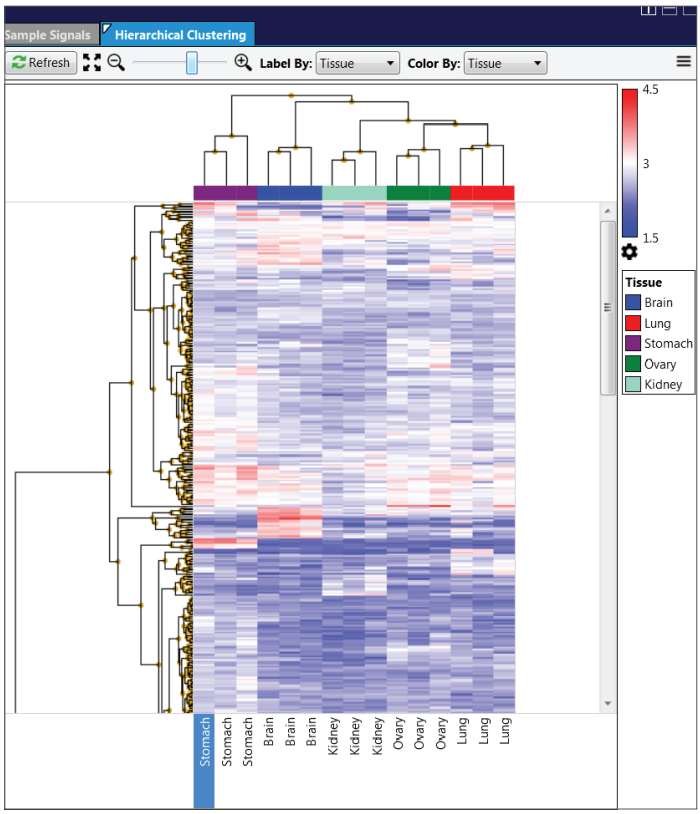

Figure 2. Unsupervised hierarchical cluster analysis based on Clariom D microarray expression data. The cluster map was generated on TAC 4.0 Software using transcriptomic data from FFPE samples belonging to brain, kidney, lung, ovary, and stomach tumor tissues. The heat map represents cluster analysis of 2,742 DEGs, filtered for those where confidences of differences are highly significant (P <0.001). DEGs from these samples are clustered by tissue type. Each row represents a single gene and each column represents a single sample. Colors represent $\log_2$ of relative expression levels (blue: low relative expression, red: high relative expression).

Since Clariom D Assays also include probes for annotated lncRNAs, we investigated expression profiles of lncRNAs in FFPE tissues. In accordance with previously published data, we found that LINC00461, which is associated with proliferation of glioma cells [9], was overexpressed in brain tumor samples (Table 1). In addition, we found that LINC00260 and LINC01381 were consistently expressed in all 5 tissues. The high concordance of data from Clariom D Assays to previously published data suggests that the Clariom D Assays can be reliably used to study gene expression profiles in FFPE samples.

A **B**

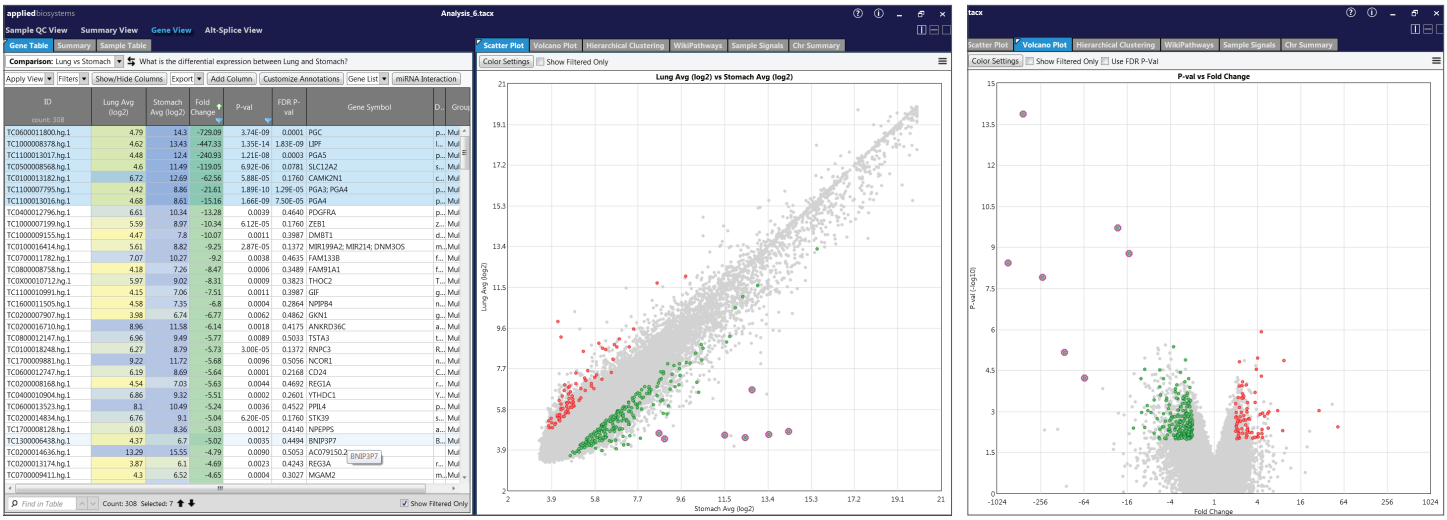

Figure 3. Visualization of tissue-specific gene expression. (A) Table of expression values and scatter plot showing relative expression of genes between FFPE stomach tumor and lung tumor samples. TAC 4.0 Software was used to examine pairwise relative expression of genes. The highlighted genes at the top of the table (blue rows) and in the plot (purple bubbles) were overexpressed in stomach tumor tissue relative to lung tumor tissue, consistent with previously published data. **(B)** The same data can also be viewed as a volcano plot, with the same genes highlighted by purple bubbles.

Table 1. Average expression level of representative genes from tumor samples measured by Clariom D Assays.*

Gene symbol	Brain average (log_2)	Kidney average (log_2)	Lung average (log_2)	Ovary average (log_2)	Stomach average (log_2)
Tissue-specific genes					
LIPF	4.22	4.44	4.62	4.14	13.43
GIF	4.52	4.54	4.15	4.5	7.06
GKN1	4.19	4.03	3.98	4.02	6.74
PGA4	4.55	4.54	4.68	4.46	8.61
SLC1A2	10.23	5.09	5.77	5.57	5.53
GFAP	6.86	3.72	3.7	4.06	4
GIPC2	4.21	9.84	4.57	5.15	11.78
SLC16A4	6.96	9.45	6.59	6.7	6.9
IGHG4	4.74	5.5	8.05	4.52	5.71
SCGB3A2	5.26	5.38	6.23	5.33	4.23
Tumor-specific genes					
PVRL1	4.58	4.65	7	4.5	4.97
VTCN1	3.96	4.25	5.15	8.62	4.62
THSD4	6.11	6.58	6.73	7.25	5.98
SLC1A2	10.23	5.09	5.77	5.57	5.53
SLC12A2	4.88	4.89	4.6	5.15	11.49
TMEM91	4.33	5.84	4.04	4.25	4.43
MLLT3	6.03	5.96	7.47	5.87	6.89
Long noncoding RNA					
LINC00260	4.45	4.3	4.34	5.4	5.02
LINC00461; MIR9-2	13.45	5.8	6.14	6.17	5.96
LINC01381	4.75	3.83	4.27	4.13	4

* Log_2 values are normalized signal intensities for each sample.

Verification of array data using TaqMan Gene Expression Assays

To confirm the differential expression results, we selected 18 genes that showed varying fold-change differences between brain tumor and kidney tumor samples [7]. Gene symbols corresponding to the 18 genes from microarray data were used to pick "best coverage" TaqMan Gene Expression Assays using the online search tool (**thermofisher.com/taqman**) [10]. Best coverage TaqMan Gene Expression Assays have been designed to provide gene-level expression profiles (that is, across all exons) and hence are a good choice to confirm gene-level fold changes from Clariom D Assays. TaqMan Gene Expression Assays have also been designed and mapped to the various transcript clusters (TCs) and probe selection regions (PSRs) from Clariom D Assays. TCs are a group of probes covering a region of the genome reflecting all the exonic transcription evidence known for the region and corresponding to a known or putative gene. PSRs are a group of one of more probes that are associated with a particular exon or exon junction. Another easy way to find TaqMan Assays corresponding to either TCs or PSRs is to use the TC or PSR IDs in the online search tool. Signal differences between brain tumor and kidney tumor samples from Clariom D Assays were compared to C_t differences from qPCR (Figure 4). We found a high concordance of data with respect to directionality of change. The magnitude of the expression changes also correlated well on the two platforms; the differences between Clariom D and TaqMan Assays can be attributed to the differences in the way signal calculations are averaged between the two platforms. Nevertheless, these data confirm that TaqMan Gene Expression Assays can be used to reliably confirm fold-change differences measured by Clariom D Assays.

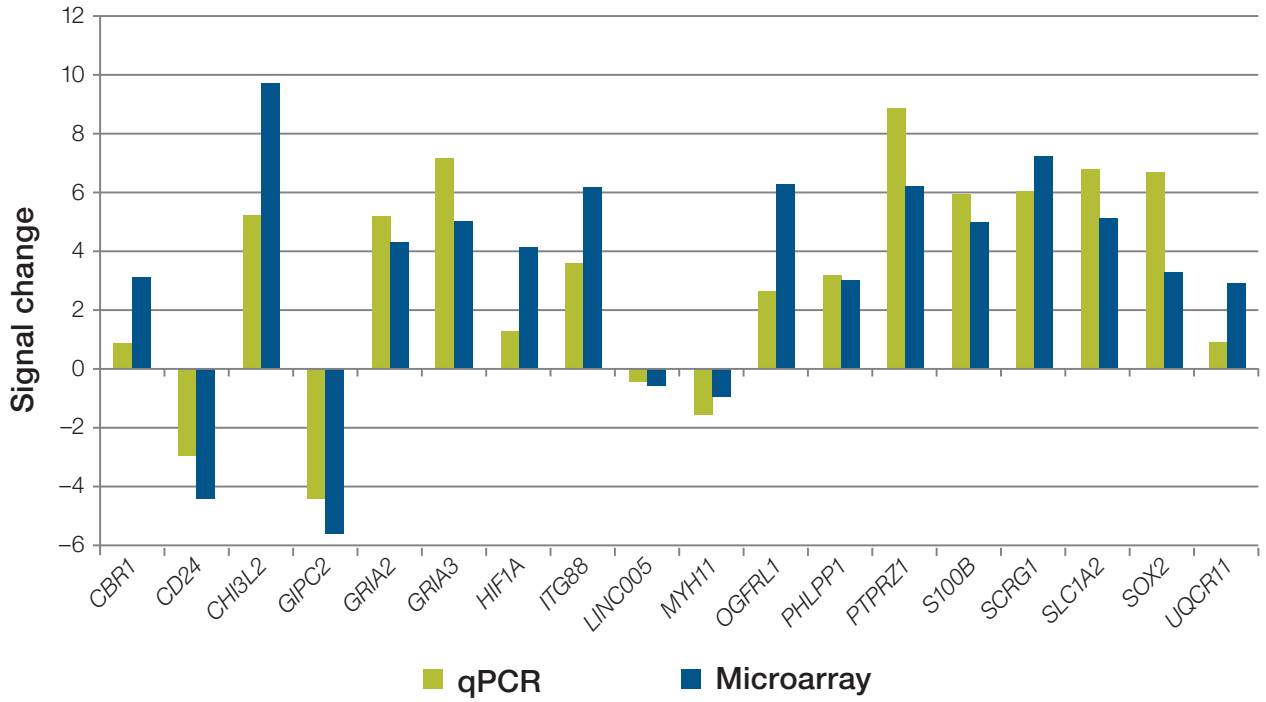

Figure 4. TaqMan Gene Expression Assays were used to confirm the differences in gene expression between brain tumor and kidney tumor samples from Clariom D Assays. The directionality of change between brain and kidney samples was the same on TaqMan and Clariom D Assays. The difference in the magnitude can be attributed to the differences in the way signal calculations are averaged between the two platforms.

Detection and confirmation of alternative splicing events

Finding differential splicing events is very easy using Clariom D Assays with TAC 4.0 Software. The software provides a visual representation of all isoforms associated with a gene between any two samples or conditions. To do this, TAC 4.0 Software uses a splicing index algorithm to measure how much exon-specific expression differs between two conditions after excluding gene-level influences. The algorithm first normalizes the exon and junction expression values by the level of gene expression and creates a ratio of normalized signal estimates from one condition relative to another.

To illustrate how this works, we chose one gene, *PHLPP1*, and analyzed the difference in isoforms present in brain tumor and kidney tumor samples. *PHLPP1* is a member of the serine/threonine phosphatase family and is known for its tumor suppressor role in many different cancers [8]. When we examined *PHLPP1* expression, we found that the sum of the expression of all transcripts was about 3-fold higher in kidney versus brain (Table 2). These results were confirmed using a TaqMan Assay that targets all transcript isoforms (best coverage choice). However, *PHLPP1* is represented by 18 different transcript isoforms on the Clariom D Assay. All 18 isoforms of this gene and their different abundances (splice index) between brain and kidney samples can be visualized using TAC 4.0 Software (Figure 5) [11]. We noted that one transcript, represented by PSR1800151620.hg.1, was significantly different between brain tumor and kidney tumor samples. To confirm this differential splicing event, we used TaqMan Assay Hs01597866_m1, corresponding to PSR1800151620.hg.1 and designed for the exon EX1800131017 within the gene *PHLPP1*. qPCR data showed a high correlation of gene expression levels between Clariom D and TaqMan Assays (Table 2). These results demonstrate that TAC 4.0 Software and Clariom D Assays provide an easy and fast solution for investigating alternative splicing events between samples. Furthermore, TaqMan Gene Expression Assays, which are designed and mapped to the probes on Clariom D Assays, can be used to confirm alternative splicing events between samples.

Table 2. Differential expression of *PHLPP1* at the gene and exon level (kidney tumor vs. brain tumor).

TC or PSR ID	Gene	Signal difference, microarray	C_t difference, qPCR
TC1800007523.hg.1	*PHLPP1* (gene level)	2.99	3.196
PSR1800151620.hg.1	*PHLPP1* (exon level)	0.63	0.492

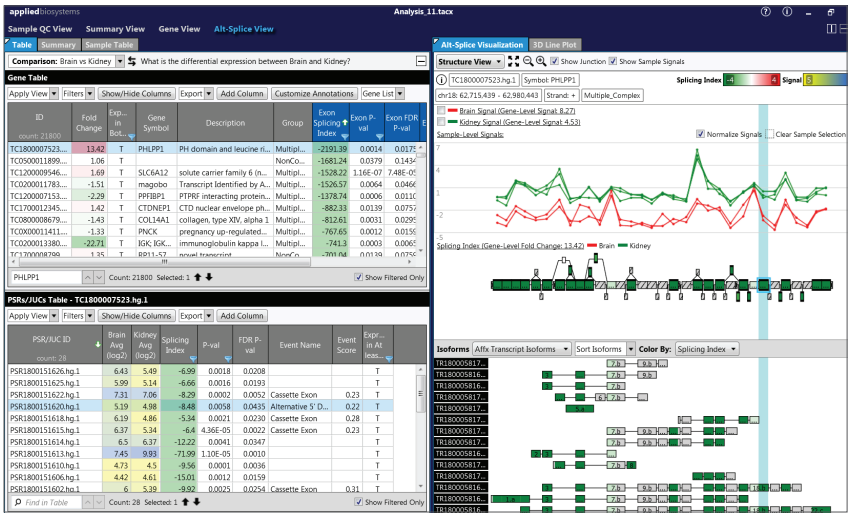

Figure 5. Alt-Splice View in TAC 4.0 Software provides a detailed record of splicing events for all genes between any two samples or conditions. Genes can be filtered by Exon Splicing Index values in the gene table. Selection of a gene by highlighting it in the Alt-Splice table results in display of all known isoforms associated with it in the PSRs/JUCs (junction probe sets) table on the bottom left and Alt-Splice graphs and 3D plots on the right. The PSRs/JUCs table is a list of all PSRs or JUCs associated with all known isoforms of the gene that can also be sorted by splice index. The line graph on the right shows the expression signal of each PSR or JUC in brain and kidney samples. The structure view displays gene structure. All PSRs and JUCs are represented in the structure view with boxes that have the same size. The gene map is colored by splice index, where shades of green and red are indicative of expression differences. Deeper shades are indicative of higher splice indices between the two samples. White, on the other hand, is indicative of low or no differences in expression of PSRs and junctions between the two samples. The bottom-right panel shows a graphical representation of all the isoforms of *PHLPP1* corresponding to the PSRs/JUCs table. These are also colored by splice index as explained above. Note that exons colored deep green or red show the highest expression variation between these brain tumor and kidney tumor samples. The blue bar on the right panel highlights the exon EX1800131017, represented by PSR1800151620.hg.1.

Conclusion

In this application note, we demonstrate successful whole-transcriptome analysis of FFPE samples using Clariom D Assays, the GeneChip WT Pico Kit, and TAC 4.0 Software. The streamlined workflow enables whole-transcriptome analysis, from sample to answer, in less than 4 days. The RecoverAll Multi-Sample RNA/DNA Isolation Kit allows extraction of both total RNA and total DNA from FFPE samples, making it an ideal choice for complete genomic studies. To quantitate the isolated RNA, we use a simple and reliable RT-qPCR method that measures the amount of amplifiable RNA in a sample, and is therefore ideal for FFPE samples. TAC 4.0 Software facilitated the analysis of DEGs across the 5 different tissues. Although degraded RNA samples can be difficult to analyze, Thermo Fisher Scientific has developed options for whole-transcriptome analyses of FFPE samples, including Clariom D microarrays and Ion AmpliSeq™ Transcriptome kits [12].

We confirmed that DEGs cluster by tissue type. Genes in this study also show tumor and tissue specificity with remarkable agreement to previously published data. TAC 4.0 Software also provides an elegant solution for studying splice variants associated with the genes and relative expression of each isoform in the different samples. TaqMan Gene Expression Assays mapped to Clariom D transcript clusters and PSRs can be readily used to confirm Clariom D Assay data. Together, Clariom D Assays, the GeneChip WT Pico Kit, and TAC 4.0 Software provide a complete solution for studying whole-transcriptome analysis in FFPE samples.

References

1. Lewis F, Maughan NJ, Smith V et al. (2001) Unlocking the archive–gene expression in paraffin-embedded tissue. *J Pathol* 195:66–71.
2. Gilbert MT, Haselkorn T, Bunce M et al. (2007) The isolation of nucleic acids from paraffin-embedded tissues—which methods are useful when? *PLoS One* 2:e537.
3. Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336
4. Stamm S, Ben-Ari S, Rafalska I et al. (2005) Function of alternative splicing. *Gene* 344:1–20.
5. Seok J, Xu W, Davis RW, et al. (2015) RASA: robust alternative splicing analysis for human transcriptome arrays. *Sci Rep* 5:11917.
6. Romero-Barrios N, Legascue MF, Benhamed M (2018) Splicing regulation by long noncoding RNAs. *Nucleic Acids Res* 46:2169–2184.
7. Thermo Fisher Scientific. Microarray normalization using Signal Space Transformation with probe Guanine Cytosine Count Correction. https://tools.thermofisher.com/content/sfs/brochures/sst_gccn_whitepaper.pdf
8. The Human Protein Atlas. https://www.proteinatlas.org/humanproteome/tissue+specific
9. Yang Y, Ren M, Song C et al. (2017) LINC00461, a long noncoding RNA, is important for the proliferation and migration of glioma cells. *Oncotarget* 8:84123–84139.
10. Thermo Fisher Scientific. Gene expression analysis using TaqMan Assays. https://www.thermofisher.com/us/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/taqman-gene-expression.html
11. Thermo Fisher Scientific. Alternative transcript analysis methods for exon arrays. https://tools.thermofisher.com/content/sfs/brochures/exon_alt_transcript_analysis_whitepaper.pdf
12. FitzGerald LM, Jung CH, Wong EM et al. (2018) Obtaining high-quality transcriptome data from formalin-fixed, paraffin-embedded diagnostic prostate tumor specimens. *Lab Invest* 98:537–550.

applied
biosystems

Ordering information

Product	Quantity	Cat. No.
Clariom D Assay, human	10 reactions	902922
GeneChip WT Pico Kit	12 reactions	902622
GeneChip Fluidics Station 450	1 each	000079
GeneChip Scanner 3000 7G	1 each	000210
TaqMan Gene Expression Assay	250 reactions	4331182

Invitrogen products	Quantity	Cat. No.
RecoverAll Multi-Sample RNA/DNA Isolation Workflow	120 preps	A26069
Qubit 4 Fluorometer	1 each	Q33226
SuperScript VILO cDNA Synthesis Kit	50 reactions	11754250

Find out more at **thermofisher.com/clariomassays**

ThermoFisher
SCIENTIFIC