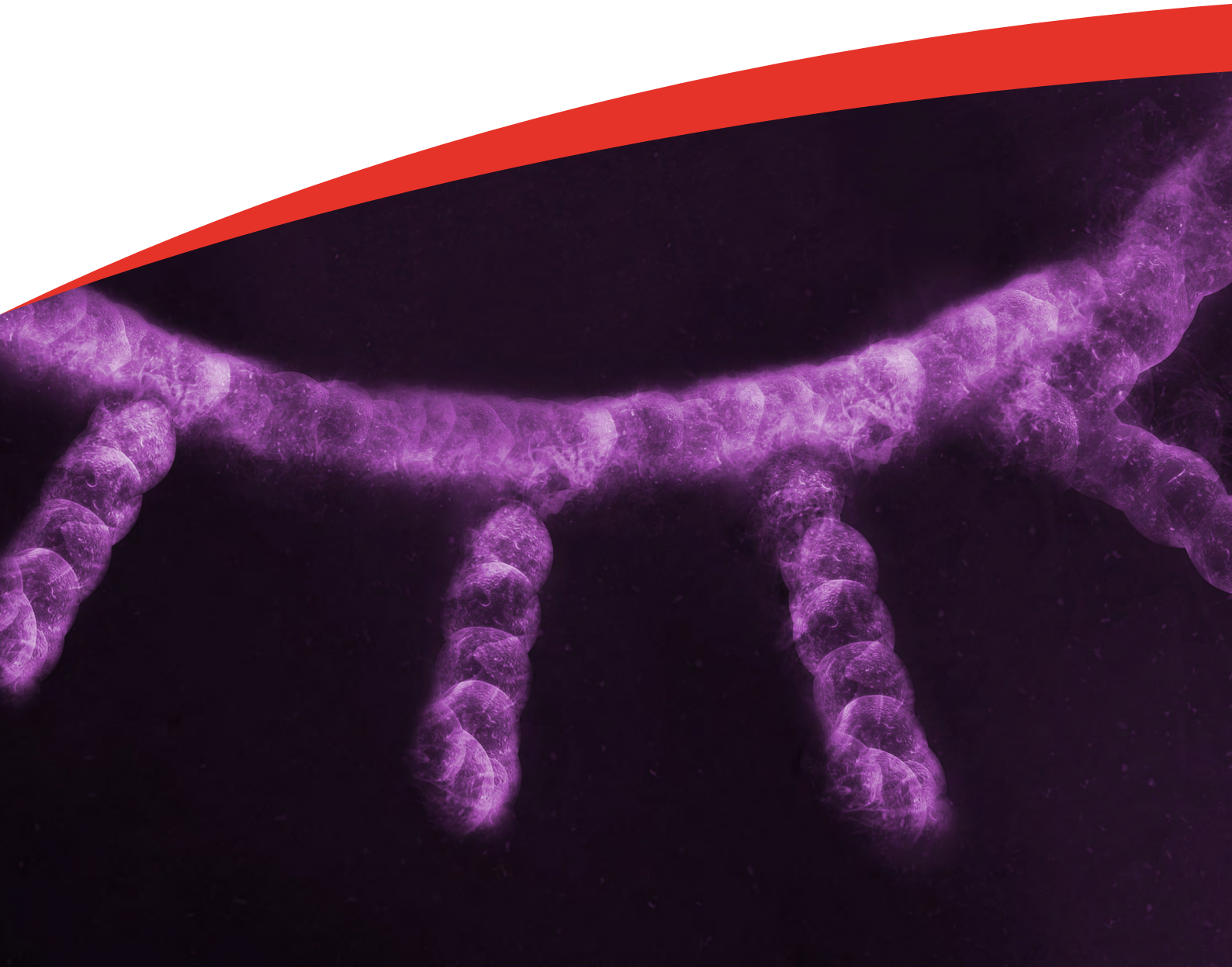
ALMAC

Partnering to Advance Human Health

White Paper

Suitability assessment and performance evaluation

Illumina's TruSeq® RNA Exome as a tool for Biomarker Discovery



Introduction:

Gene expression (GE) profiling from tissue samples using high content discovery platforms is a cornerstone of biomarker discovery and validation. Next Generation Sequencing (NGS) based RNA-Sequencing is a novel approach and has several advantages over established GE methodologies such as qPCR and gene expression arrays.

This is due to its high sensitivity and the ability to detect gene fusions and alternative transcripts in a single assay. TruSeq® RNA Exome utilises sequence-specific capture probes to known exons and does not rely on the presence of polyadenylated transcripts, therefore it is ideal for RNA-Seq with Formalin Fixed Paraffin Embedded (FFPE) or degraded samples and samples with limited starting material (input from 20 ng). By focusing the sequencing budget on only the coding regions of the genome (mRNA), this enables increased sample throughput and lower costs per sample whilst retaining the ability to discover novel features such as splicing and fusions. In this white paper, Almac Diagnostic Services describe the evaluation of the most appropriate nucleic acid extraction kit to be used with the TruSeq® RNA Exome, we demonstrate equivalent performance of TruSeq® RNA Exome compared to gene expression arrays, the establishment of robust QC metrics, defined process controls and a bespoke bioinformatics pipeline for data analysis.

Results:

RNA Extraction

In a preliminary screen, five extraction kits were compared using FFPE cell line samples which were processed according to routine clinical sampling. Three replicates per cell line (GM12878 and MCF-7 (manufactured into FFPE blocks at Almac Diagnostic Services) across five extraction kits were utilised for the analysis, each using 3 x 10 µm sections. One operator performed all extractions to minimise any operator bias. Quality control metrics analysed were yield (≥20 ng/µl), 260/280 ratio (~2.0), 260/230 ratio (~2.0) and the DV_{200} value (≥30%).

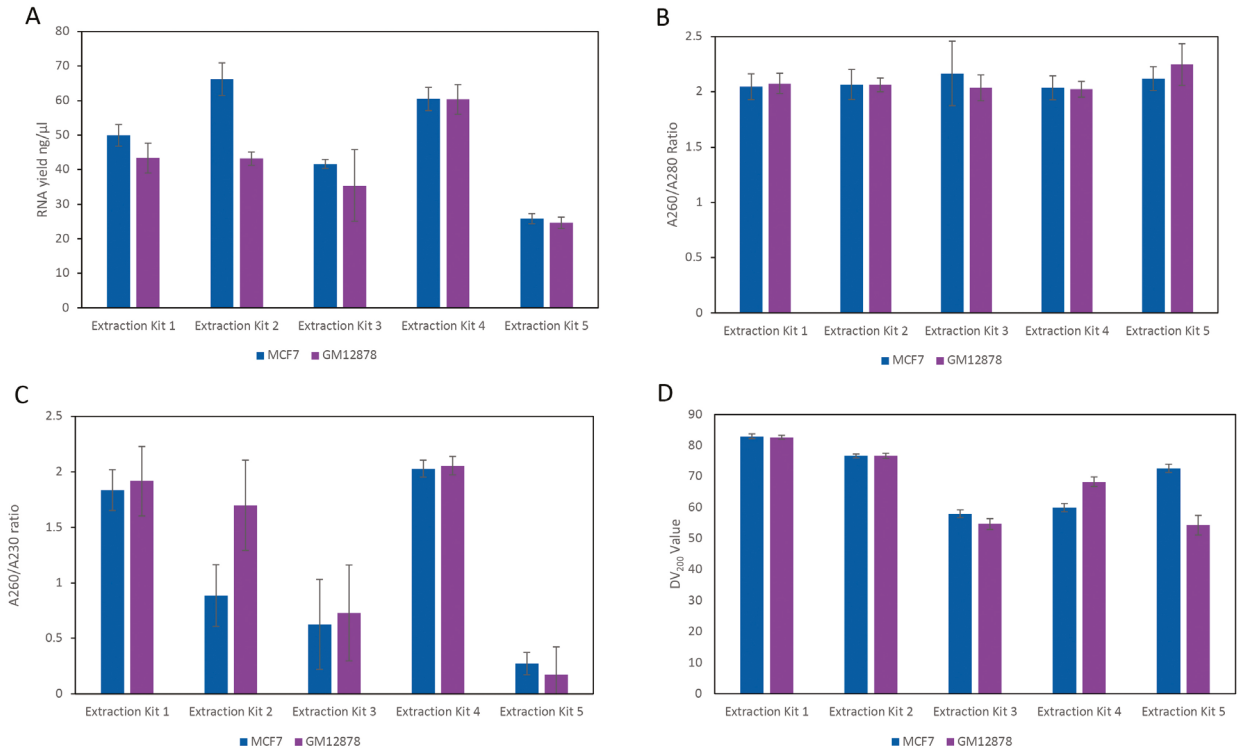


Figure 1.
Comparison of yield and quality from nucleic acid extraction kits utilising cell line FFPE samples. The RNA quantity (A) is expressed as ng/µl, 260/280 ratio (B), 260/230 ratio (C) and DV_{200} value (D).

The highest yields (ng/µl) were obtained with samples processed using kits 1, 2 and 4 (Figure 1A). Sample quality reflected by 260/280 and 260/230 ratios was closer to the optimal ratio of 2.0 in kits 1, 2 and 4 (Figure 1B and 1C). High levels of sample integrity, as assessed using DV_{200} metric were observed in all samples across all extraction kits, however kits 1 and 2 obtained a slightly higher DV_{200} value (not statistically significant - Figure 1D). Based on these results, three kits (kits 1, 2 and 4) were shortlisted for testing with clinical samples. Twenty macrodissected FFPE breast cancer surgical resections were extracted (2 x 10 µm sections each) with the 3 selected kits in duplicate. Sample FFPE blocks ranged in age from 3-20 years and had percentage viable tumour areas in the range of 15-100%.

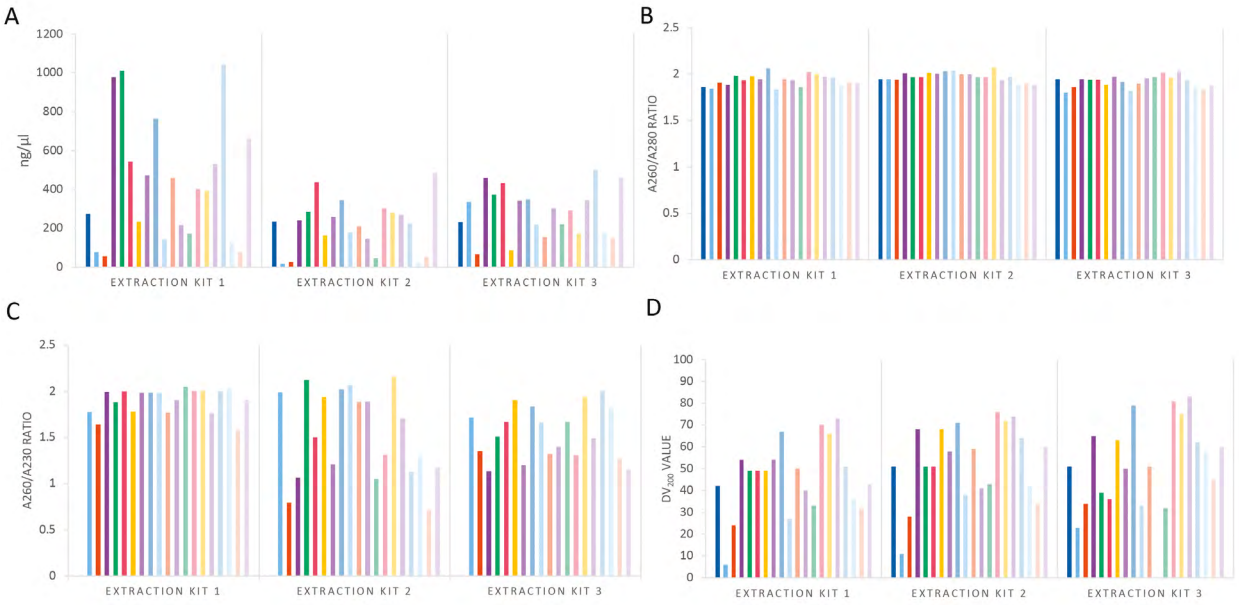


Figure 2.
Comparison of yield and quality from nucleic acid extraction kits utilising 20 FFPE breast cancer tumour samples. RNA quantity (A) expressed as ng/µl, 260/280 ratio (B), 260/230 ratio (C) and DV_{200} value (D). ng/µl, 260/280 ratio (B), 260/230 ratio (C) and DV_{200} value (D).

The largest concentrations (ng/µl) were observed in samples processed using kit 1. Furthermore, samples extracted using kit 1 obtained much higher 260/230 ratios, closer to the optimal 2.0, in comparison to the other two tested kits while all extraction kits showed a similar distribution in 260/280 ratios (Figure 2). No obvious difference in the distribution of DV_{200} scores was observed across all three kits (p= 0.257; ANOVA). Given its consistent yields and RNA quality, kit 1 (Qiagen RNeasy FFPE kit) is recommended to extract RNA from FFPE material for use with the TruSeq® RNA Exome chemistry (Figure 2).

Concordance of RNA Exome to microarray gene expression assays

Almac Diagnostic Services has discovered and validated a portfolio of predictive and prognostic gene expression assays based on distinct molecular subgroups. One key benefit of this approach is that it allows for migration of such gene expression assays across multiple platforms. These molecular assays were used to evaluate concordance of gene expression signatures between the microarray discovery platform and the TruSeq® RNA Exome platform. These include the DNA Damage Response Deficiency (DDRD) Assay (Mulligan et al. 2014), an Epithelial to Mesenchymal Transition (EMT) Assay (McCabe et al. 2017) and a prognostic signature in Prostate Cancer (Metastatic Assay) (Walker et al. 2017). A total of 39 FFPE clinical samples (17 breast cancer, 13 ovarian cancer and 9 prostate cancer) including small core needle biopsies were profiled using the Almac developed Xcel microarray and the TruSeq® RNA Exome kit. A high correlation was observed between both platforms for all three assays with Pearson correlation co-efficient of 0.93 (DDRD), 0.94 (EMT) and 0.97 (Prostate Metastatic Assay) (Figure 3).

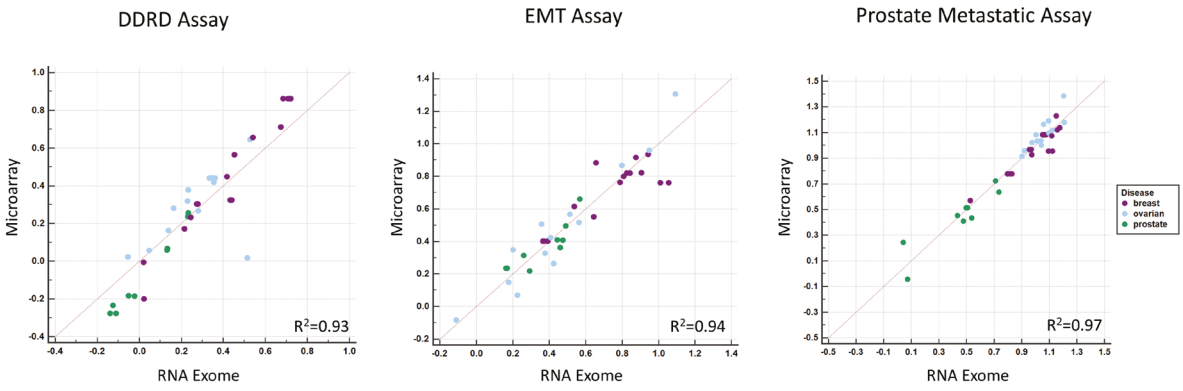


Figure 3.
Concordance of Almac gene expression assays (DDRD, EMT and Prostate Metastatic) between TruSeq® RNA Exome and microarray.

This data suggests that the TruSeq® RNA Exome kit is a suitable platform for migrating and delivering gene expression studies formerly run on a microarray platform.

Precision

Analytical precision of TruSeq® RNA Exome was assessed using the DDRD Assay. Two extractions of the MCF7 cell line control (representing DDRD low signature scores) and 3 homogeneous breast clinical RNA sample pools were used at an input of 100 ng. Clinical samples were pre-selected to include one DDRD low signature scoring sample, one DDRD high signature scoring sample and a sample which scored close to the medical decision point (central scoring). The experimental design enabled the assessment of repeatability (inter-assay and intra-assay) and included the determination of within-laboratory reproducibility through assessment of variability between operators (n = 2), reagent lots (n = 3) and equipment (n = 2 sequencers) across analytical runs (n = 10). Runs 1-3 were performed ensuring all controllable variables (reagent lot, equipment, and operator) remained constant to enable assessment of intra- and inter-assay repeatability. Runs 4-10 included a random distribution of reagent lots, operators and sequencers across runs to enable assessment of within-laboratory reproducibility (Figure 4).

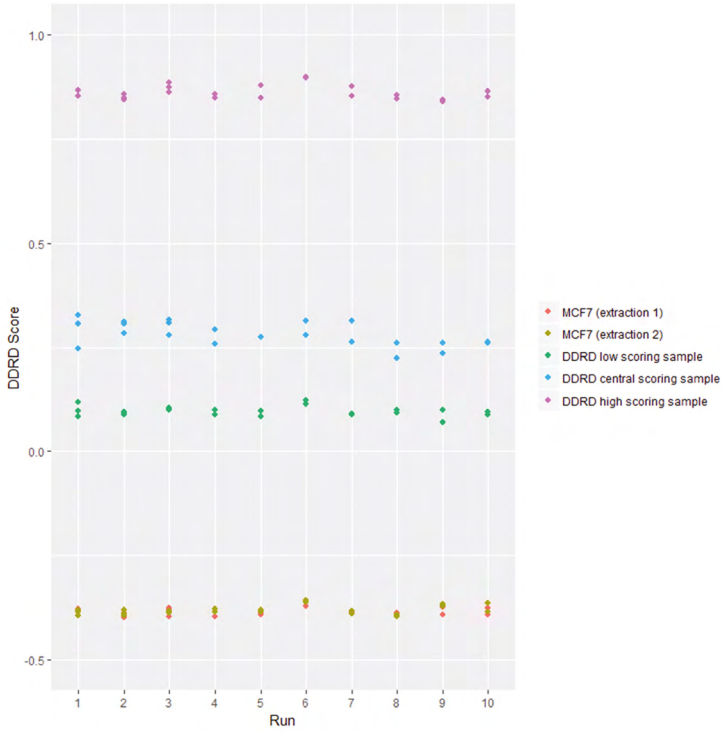


Figure 4.
Precision of TruSeq®RNA Exome as measured using the in-house developed DDRD signature. 10 sequencing runs were performed.

The precision study standard deviation (SD) estimates for the DDRD Assay were applied to a previously profiled breast cancer cohort to quantify the effects of imprecision (using inter-assay repeatability, intra-assay repeatability and within-laboratory reproducibility) on signature call. In each case the percentage of samples unaffected in terms of signature call was calculated (Table 1).

Table 1: Percent agreement in DDRD Assay call in clinical breast cancer cohort pre and post standard deviation adjustments

Sample	Precent agreement in signature call after adjustment for source of variablity		
	Inter-assay repeatability	Intra-assay repeatability	Within-laboratory reproducibility
MCF7 (extraction 1)	99.2%	98.4%	97.9%
MCF7 (extraction 2)	100.0%	98.9%	97.6%
DDRD low scoring sample	100.0%	97.8%	97.5%
DDRD central scoring sample	100.0%	94.8%	93.4%
DDRD high scoring sample	97.9%	98.1%	96.2%

The largest source of variation resulted in <7% of samples being misclassified. Hence, all sources of variation on this platform resulted in >93% of samples being correctly classified. This result suggests that the TruSeq® RNA Exome is a highly repeatable and reproducible platform for biomarker discovery and validation.

Process Control Monitoring

As part of the RNA Exome Discovery Tool, Almac Diagnostic Services has developed Statistical Process Control (SPC) charts and have established metrics for monitoring a number of key parameters including percentage total reads aligned, percentage read positions aligned to RNA exome targets, percentage ribosomal RNA and percentage read duplicates. Together with pre-defined acceptance limits, control samples are used to monitor sample processing within each batch. Control samples are monitored using a comprehensive set of Westgard rules, to ensure stability and detect variation therefore providing confidence in sample processing (Figure 5).

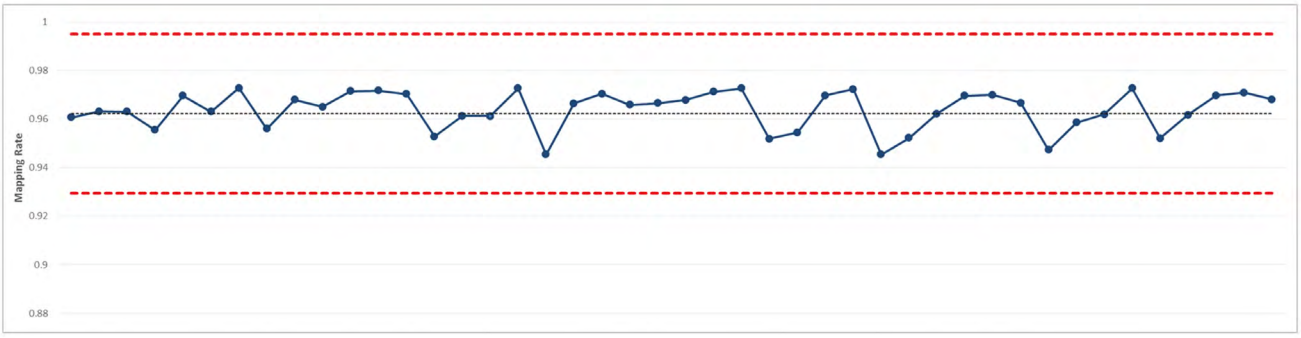


Figure 5.
An example Statistical Process Control (SPC) Chart for monitoring process quality over time. Standard deviations of Process Control metric (Mapping Rate) from FFPE samples recently analysed with TruSeq® RNA Exome.

Bioinformatics Pipeline

Our mature in-house pipeline uses widely accepted mainstream software tools (Figure 6), and is implemented on the Illumina Connected Analytics (ICA) cloud-based platform.

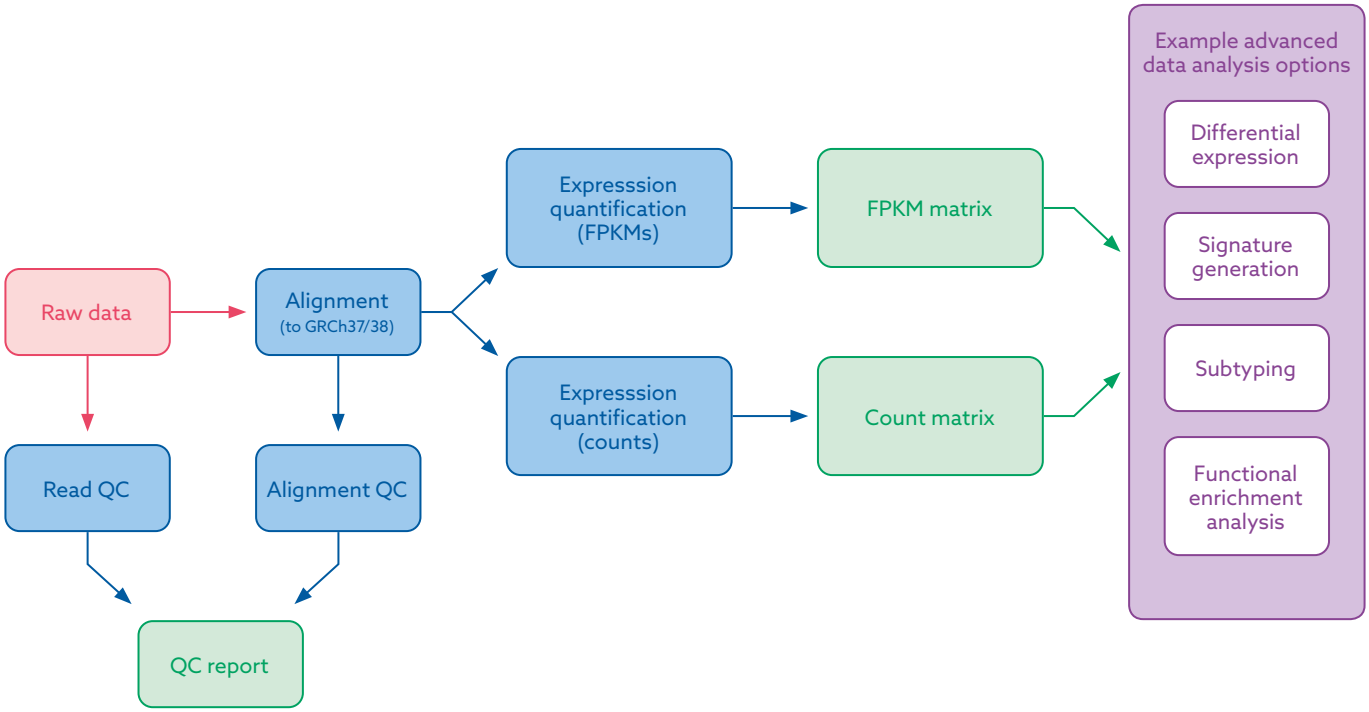


Figure 6.
RNA Exome expression profiling bioinformatics pipeline implemented on Illumina Connected Analytics (ICA) outlining key analysis steps and examples of advanced data analysis options.

Data analysis and reporting is performed using the latest DRAGEN suite of secondary analysis tools and/or Almac customised bioinformatics pipelines hosted on the Illumina Connected Analytics (ICA) platform. ICA is a cloud-based NGS solution which enables full integration with Illumina sequencing instruments, whilst ensuring scalability and security of project data. A significant benefit of utilising ICA within this offering is the flexibility of workflow optimisation and the ability to customise bespoke pipelines. Taking raw sequence data as input, the pipeline generates a set of quality control metrics to rigorously assess sequence and alignment quality that include mapping rate, ribosomal RNA contamination, on-target rate and duplication rate as well as distribution and principle component analysis (Table 2).

Table 2: QC parameters, and acceptance criteria

Parameter	Threshold
Mapping rate	>92.4%
Ribosomal RNA contamination	<20.2%
On target rate	>87.4%
Duplication rate	<43.1%

Expression estimates at both the count and fragments per kilobase million (FPKM) level are provided in convenient matrix format ideally suited for advanced downstream analyses such as molecular subtyping, differential expression or signature generation. Additional functionality such as fusion detection and isoform expression estimation is also available.

Summary

Here we describe the Qiagen RNeasy FFPE kit as the **optimal RNA extraction kit** for RNA-Seq using the TruSeq® RNA Exome platform. Our study demonstrates the **high reproducibility** of RNA Exome outputs, and **high concordance** in measurement of relative gene expression levels between the RNA Exome and microarrays in the Almac Laboratories. This was exemplified using a number of internally developed molecular signatures. We describe the establishment of a set of **rigorous quality control metrics, established process controls, metrics for surveillance and a robust, scalable bioinformatics pipeline**, which is fully customisable based on client requirements. Overall this workflow has been implemented by Almac, the TruSeq® RNA Exome is the ideal solution for **comprehensive** transcriptome profiling from degraded/or limited material including small biopsies at a **minimal sample input.** Almac are offering the TruSeq® RNA Exome panel as a solution for biomarker discovery with a protocol and QC pipeline optimised to **maximise on sample pass rates, sequencing data quality and biological relevant raw data provision.**

[1] https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf

almacgroup.com

GET IN TOUCH

UK
Almac Group
19 Seagoe Industrial Estate
Craigavon
BT63 5UA
United Kingdom

diagnostics@almacgroup.com
+44 28 3833 7575

US
Almac Group
4238 Technology Drive
Durham
NC 27704
United States of America

diagnostics@almacgroup.com
+1 (919) 294 0230