AMP Europe 2018
Poster P07

Robust measurement of signal transduction pathway activity in cancer using RNA sequencing on cells and FFPE tissue

S. van den Bosch, A. van Brussel, F. van Hemert, W. Verhaegh, R. Bakker, S. Kuijpers, B. de Regt, M. van Driel, H. Feitsma

Precision & Decentralized Diagnostics, Philips Research, Eindhoven, The Netherlands, contact: harma.feitsma@philips.com

Introduction

With the introduction of high throughput genomics technologies, more and more molecular information of cancers has become available, but the translation into pathophysiological characteristics to improve treatment selection is still a large challenge. We have developed a method to quantitatively and reliably measure the activity of signal transduction pathways in cancer cells, based on the interpretation of expression levels of transcription factor target genes via a Bayesian computational model. We have previously shown that this works well on expression levels measured with RT-qPCR or microarrays[1]. Using RNA sequencing (RNAseq) to reliably measure expression levels has so far proven more difficult, because of the large variation in protocols and quality of results, especially for formalin-fixed, paraffin-embedded (FFPE) tissue. Here we show how to generate RNAseq expression data with sufficient quality to call signaling pathway activity.

Conclusions

- Upon proper selection and validation of protocols, RNAseq can be used to robustly measure expression levels in fresh as well as FFPE material.
- Thus determined expression levels of transcription factor target genes can be used to properly predict signaling pathway activity.

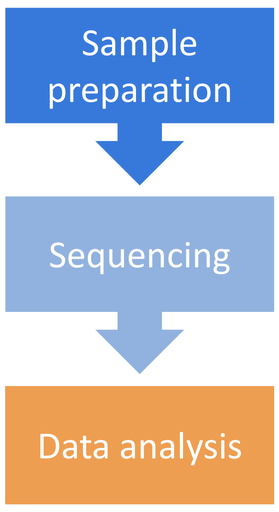
Discussion and next steps

These two well-defined cases of ER and Wnt, two well-studied pathways with proven ground truth samples, validate the approach of using RNAseq for signaling pathway activity prediction. Nevertheless, next steps are needed to standardize the approach:

- Quality control boundaries will be set to exclude poor quality samples
- ER and Wnt models will be validated in clinical samples of variable tissue
- Other signaling pathway models will be calibrated with RNAseq data

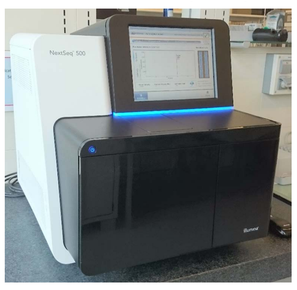
Results

RNA sequencing workflow



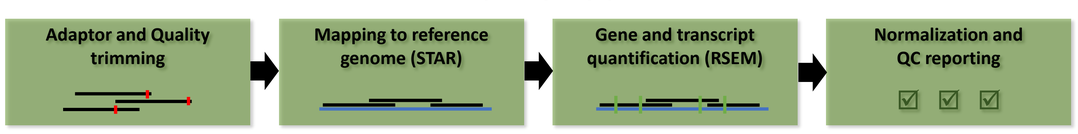
We have set up a complete RNA sequencing (RNAseq) workflow in house, including sample processing and data analysis. We saw a strong need to do this ourselves to be able to generate high quality genomics data, instead of depending on available datasets of unknown quality.

We have evaluated several RNAseq library preparation protocols for quality of expression quantification and have chosen two for further validation and testing: the TruSeq stranded mRNA and TruSeq RNA enrichment kits. The libraries are sequenced on an Illumina NextSeq 500 instrument.

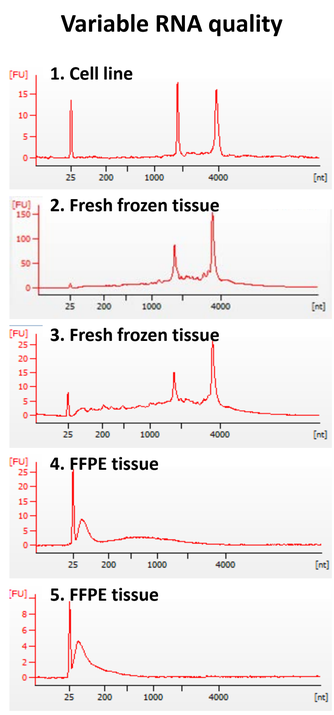


The RNAseq analysis pipeline uses STAR for mapping of reads and RSEM for quantification, exporting Transcripts Per Kilobase Million (TPM) values as quantitative measure of gene expression.

RNAseq analysis pipeline



Variable RNA quality



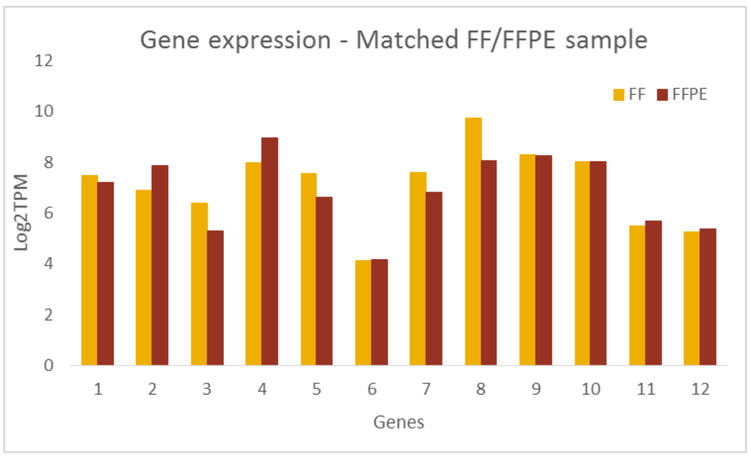
#	Material	RIN	DV200
1	Fresh	9.7	97%
2	Fresh frozen	7.0	95%
3	Fresh frozen	5.5	89%
4	FFPE	2.3	51%
5	FFPE	2.5	18%

RIN: RNA Integrity Number
DV_{200}: %fragments > 200 nucleotides

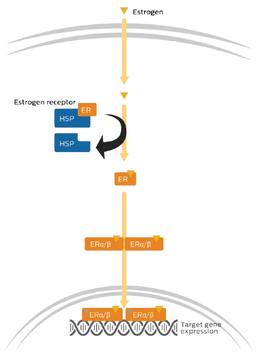
RNA sequencing of FFPE samples

Formalin-fixed, paraffin-embedded (FFPE) tissue is the standard for oncology samples in hospital workflows. Its use for genomics analysis is challenging due to degradation of DNA and RNA. However, to be able to match genomics to the pathology image, processing of FFPE material in a robust manner is key.

We selected and validated RNAseq workflows specifically for running FFPE samples. We have determined the amount of sequencing reads needed and optimized the normalization method to quantify the number of reads per gene. We observed good correlation in gene expression levels between FF and FFPE material of the same patient sample.



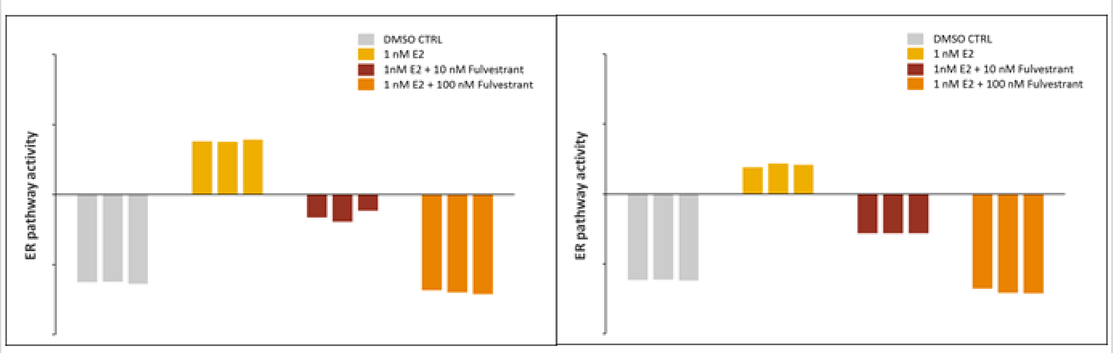
ER Pathway



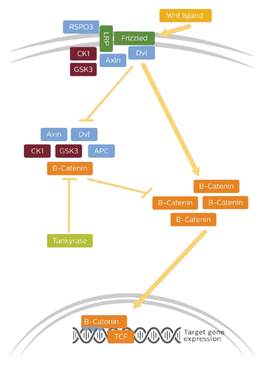
MCF7 breast cancer cell line

The OncoSignal model that predicts ER pathway activity was calibrated for RNAseq by sequencing ER inactivated and activated MCF7 cell line samples. TPM values of the target genes were used as input to train the model. Independent cell line samples were used to validate the model. Control samples (grey) are ER inactive, while estradiol (E2) treatment (yellow) activates the pathway. Fulvestrant treatment (dark and light orange) blocks ER activity in a dose-dependent manner.
The same samples were tested on the two validated library prep kits, and the ER model was calibrated for each kit separately. The kits gave similar, good quality results, and can therefore both be used. In future use, choice for either kit may depend on cost, effort, number of samples etc.

TruSeq® Stranded mRNA Library Prep | **TruSeq® RNA Enrichment Library Prep**



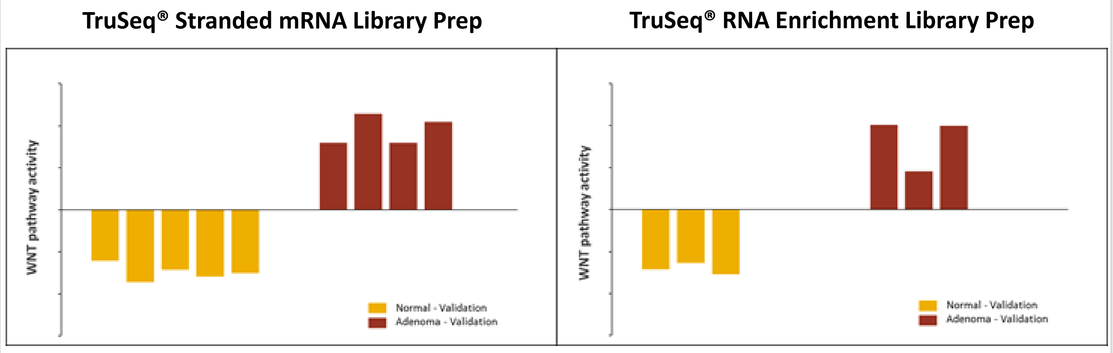
Wnt Pathway



Colon adenoma vs normal colon samples

With the optimized RNAseq workflow we sequenced normal colon and colon adenoma FFPE samples. Selected samples were use to calibrate a model for prediction of Wnt pathway activity. A second set of samples was used to validate the model. Wnt activity score is shown for normal colon samples (yellow), which are known to be Wnt inactive, and for colon adenoma (dark orange), which usually are Wnt active.
Again we tested this on the two different library prep kits, and again the kits gave similar, good quality results.

TruSeq® Stranded mRNA Library Prep | **TruSeq® RNA Enrichment Library Prep**



Materials and methods

RNA extraction from samples was done using standard methods. Library preparation for RNAseq was done following protocols from the supplier on 10-100 ng input RNA. Samples were sequenced on an Illumina NextSeq 500 instrument with 2x 75 bp paired-end reads, yielding 25-100M reads per sample, dependent on the library prep method and RNA quality. Reads were quality-filtered, mapped to hg38 with STAR and quantified with RSEM, resulting in gene expression values in TPM.

For creating OncoSignal models, known direct target genes of the pathway-associated transcription factors were identified and Bayesian computational models were made which infer pathway activity from mRNA expression levels of these target genes[1]. The models have been calibrated with samples with known pathway activity from a single model system and further validated with clinical samples.

References:
1 Selection of personalized patient therapy through the use of knowledge-based computational models that identify tumor-driving signal transduction pathways; W Verhaegh et al, Cancer Research 2014 74(11); 2936-45

PHILIPS