Combinational usage of next generation sequencing and qPCR for the analysis of tumor samples

Robert P. Loewe

GeneWake GmbH, Floriansbogen 2-4, 82061 Neuried bei München, Germany

ABSTRACT

The combination of multiple techniques especially those adding complementary information have proven to be beneficial in terms of data consistency. The employment of quantitative PCR (qPCR) prior to next generation sequencing (NGS) methods such as RNA-Seq and mutational analysis presented here does not only enhance data in terms of CNV integration and sample choice, but also allows a faster and more efficient workflow. Correct analysis of libraries prior to sequencing has proven to be a vital step for specific assumption and to some extent for a more parallel testing. By illustrating the combination of qPCR and NGS in oncological examples, the potential of this approach is presented.

1. Introduction

Next generation sequencing (NGS) has been the driving force in molecular biology and biomedicine [1]. After completion of the human genome project, this technique arose and is now used to tackle the questions remaining after identification of the human genetic information. We have to answer the terms of regulation and transcriptional importance of specific genes. The method, where sequencing and gene expression come together is RNA-Seq [2]. This technique has as all methods benefits and shortcomings, but in combination with quantitative PCR (qPCR) and gene specific sequencing it is a powerful tool that can and will alter our procedure of analyzing patient samples. Possible pitfalls and solutions are provided to tackle problems not only for RNA-Seq but also for CNV (copy number variation) analysis, which mostly employ qPCR. Since oncology is a major field of interest and represents a good example as research area due to a multitude of starting material, it has been chosen to be the focus of this methodological overview.

The usage of RNA-Seq is quite promising in the field of oncology as not only most somatic mutations of interest are expressed also on the RNA level (e.g. KRAS Codon 12/13 and BRF Codon 600) [3,4] but also some expression levels are linked to therapeutic outcome such as RRM1 for gemcitabine [5] and ERCC1 for cisplatin [6]. Fusion genes like EML4-ALK have also proven to be valid targets for drug intervention [7]. In case of non-small cell lung cancer (NSCLC) the majority of used drugs are to some degree linked to the expression of genes or to genetic changes measurable on RNA level. The combination of both expression and genetic analysis holds quite some merit. In addition for quite a number of genes allelic discrepancies on gene expression level as well as to some extent loss of heterozygosity can be analyzed by RNA-Seq [8]. In other cancer types such as breast cancer the parallel measuring of EGFR, ERBB2, ERBB3 and ERBB4 has shown to be of clinical importance for pertuzumab [9]. Also the expression of CXCR4 [10] and TNFSF11 is reported to be linked to metastasis risk and for the later a therapeutic intervention with denosumab is clinically approved [11]. To perform a valid RNA-Seq experiment, certain quality criteria have to be met to insure reproducible and clinical relevant results. The following article aims to give examples how to employ quality steps into a RNA Seq workflow.

2. Description of methods

2.1. General overview

A number of steps included in NGS and qPCR can influence the experimental outcome and even more the general conclusion of analyzed data. When planning a set-up, the most straightforward method is often preferred, as this will give us an idea what to expect quite soon. Sometimes, however by adding control steps and giving us the possibility to go back to a certain time point in the experiment, we achieve additional value and in some cases a better reproducibility. In the field of NGS focusing on RNA-Seq and CNV mistakes can be costly. They can originate from handling of the starting material, quality control, library check, input amount control, running the sequencing itself down to data analy-
sis. This article concentrates on the issues at hand regarding source material, targeted RNA-Seq and qPCR integration.

2.2. Source materials and impact on subsequent analysis

Tumors and their cells as source material come in different variations, such as solid tumors excised, formalin fixed and paraffin embedded (FFPE), bladder cancer cells in urine and its sediment, circulating cells in body fluids and of course fresh tissue biopsies. It is vital to understand the nature and quality of the source, as this will influence subsequent analysis greatly. RNA as well as DNA is susceptible to degradation due to fixation and/or biological processes prior to taking the sample. For formalin fixation as an example, RNA is often shorter than 200 bp in length and only rarely exceeds 400 bp. Although DNA is more stable, it is also quite fragmented [12]. A reliable nucleic acid extraction method is therefore a paramount. A large number of commercial kits are available in a great range of usability for subsequent analysis. In our hands, Roche High Pure FFPE RNA Isolation Kit (version 2012) and Promega Relia (for DNA and RNA) as well as Maxwell 16 (altered Blood LEV DNA Kit with RNA lysis buffer) have yielded sufficient results (not shown).

RNA isolation from FFPE has to be reproducible in order to be integrated into the daily routine. RNA quality in terms of length highly depends on the respective histology block and can therefore not be the sole criteria for evaluation. RNA quantity as well as purity is linked to the kit chemistry. Therefore it is essential to know the requirements of the downstream application for forming a decisive end result.

To evaluate different kits in terms of outcome, it is important to use similar starting material. In first experiments to check kits for reproducibility, we use multiple blocks from the same tumor patient (e.g. granulosa cellular tumor) and analyze alternating sections. Theoretical example: Two isolation kits for RNA should be compared. From two blocks, four sections are made. The first and third from one block are used for kit A whereas the second and fourth are used for kit B. The same is done for the other block. So the samples chosen should be biological-wise similar for those from the same block. In addition, combining both block sections in one pool and splitting after each lysis would result in further almost identical samples providing the lysis is complete. It is not advisable to use sections from different tumors, different blocks or sections from one block but quite distant from each other as the biology of the sample can be locally quite different and comparisons are difficult. Prior to each experimental series, the nucleic acid isolation method should be re-evaluated on small sample numbers equivalent in type to study. In case of RNA, reverse transcription and target length are also of great importance. Shorter amplicons or regions of interest are usually more stable in varying samples. Sample types like urine and blood comprise high levels of inhibitors for PCR. Therefore, purity should be one of the great concerns. Additional washing steps even if not described in the commercial kit can sometimes improve PCR results. In regards to potential pitfalls, the subsequent analysis should be targeted (if possible and/or desired) and controllable to avoid piling of errors within the preparation. In addition, if qPCR is used to evaluate sample or kit quality the genes analyzed (initial analysis of reference genes is recommended) should vary in terms of estimated Cq. Meaning it is advisable to employ for example three reference genes with different Cq to check the samples for reproducibility (e.g. Cq 25, 28 and 32).

2.3. Targeted RNA-Seq – how to sequence difficult sample

One of the great advantages of NGS is throughput. This gives rise to a new dimension and in essence scale of experiments. However, the amount of data compiling in one run and the sum of unwanted information is increasing dramatically. To avoid this, multiple methods of target enrichment have been described to support the researcher. Multiple vendors offer solutions and some are quite tailor-made. As comparisons are frequently published, the decision is not an easy one. In terms of RNA, a possible alternative is only reverse-transcription of your genes of interest with gene specific primers. In some cases, entire gene families can be reverse transcribed with the same degenerated primer [13].

2.3.1. Quality control of the different steps of RNA-Seq

By incorporating qPCR in multiple steps it is possible to control the overall experiment quite stringently. Fig. 1 illustrates the steps from sample to final result of a 454 Sequencer (Roche Diagnostics) run.

2.3.2. Quality control of the second strand synthesis

After RNA isolation, samples are reverse transcribed with gene specific primers. It is advisable to include also primers for not high expressed reference genes to this mix in addition to the primers for the genes of interest to insure comparability of samples for data analysis. Reverse transcription (RT) can be done for example with mixes like the MMLV H-Point mutant by Promega (similar products are available from other vendors) or with Transcripter from Roche Applied Science. There are also kits on the market combining this step with the second strand generation such as the cDNA Synthesis System (Roche Applied Science). Otherwise, the second strand has to be generated separately (for example with the second strand kit from NEB). After RT and second strand synthesis, a control qPCR is advisable to insure that both reactions did perform

![Fig. 1. Flow diagram of the steps from nucleic acid isolation to the sequencing run itself.](image-url)
according to the specification. An exemplary control set-up is seen
in Table 1.

Ideally, the following Cq should be measured: Cq(3) \approx Cq(4) \approx Cq(5), if the second strand synthesis did work. In case of failure, the
following should be the case: Cq(4) > Cq(5). This means that the
second strand synthesis was not sufficient and this part should
be repeated. If all values are below detection limit or close to it, this
could be due to either low expression of the chosen reference gene
or failure of the RT. It is therefore vital to employ references that
have been tested prior on the sample type. If the second strand
synthesis was deemed successful after end repair, adaptors can
be linked. This can be done according to the manufacturing proto-
col. The subsequent sizing – meaning the exclusion of certain frag-
ment sizes and enrichment of other – is highly dependent on the
sequencing platform used and the targeted read length.

### 2.3.3. Quality control of the sizing of the library

The generated library has to be sized to avoid small-unwanted
fragments (adapter-multimers, primer-dimers and short double
strands generated by earlier reactions). The main method listed
by Roche 454 is sizing with a designated buffer and the employ-
ment of AmPure beads (Beckman Coulter) at different volume ra-
tios in comparison to the library. The beads show a lot-by-lot
variation and with the regular protocol an initial experiment
should show how the used lot compares to earlier ones. In addi-
tion, a sizing of different concentrations should be done to avoid
one-sided results. The easiest way to evaluate sizing and set
cut-offs for the sizing size is a sizing experiment with a specific
amplicon ladder (of any desired length and source; ideally ampli-
cons represent the range of size needed; all amplicons should be
different in terms of sequence to insure differential analysis). In
general, the sizing limits for 454 are 500–800 bp as wanted range.
FFPE and other material (such as urine derived RNA) however can
only rarely achieve such sizes, hence lower cut offs have to be cho-
sen such as 200 bp (meaning enrichment of everything above
200 bp and eliminating as much as possible of everything below)
keeping in mind that the adaptors add additional sequence to each
end. The fragment size of the library is also comparable on other
next generation sequencers such as the MiSeq and HiSeq (Illumina)
or the Ion Torrent Sequencer (Life Technologies).

The ratio of beads to library volume is essential to obtain a
proper binding capacity and the desired size cut off for short
fragments.

To insure the criteria for sizing are met, different ratios of beads
are compared with the mentioned amplicon ladder. This marker
ideally consists of at least two bands below (e.g. 100–150 bp) and
two bands above (e.g. 250–300 bp) the desired cut-off. It is
either possible to design primers for genomic areas that meet these
criteria or to recycle existing ones. As HRM (high resolution melt-
ing) for a different set-up might be of use, we designed primer
pairs to analyze exons of TP53, PIK3CA and others to scan for so-
matic mutations in tumor tissue. Therefore, we simply ran PCRs
for a certain number of exons that fit the size range and pooled
the products. After primer removal by clean-up with AmPure

---

**Table 1**

Control experiment for RT and second strand synthesis (in this case samples were reverse transcribed with the primer called HPRT1_reverse_primer_1).

<table>
<thead>
<tr>
<th>Step</th>
<th>Sample type</th>
<th>Reaction step</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample after second strand synthesis</td>
<td>Control for RT with a single primer HPRT1_forward_primer_1</td>
<td>25 cycles of a “PCR” (single primer) to amplify reverse transcribed cDNA</td>
</tr>
<tr>
<td>2</td>
<td>Sample after second strand synthesis</td>
<td>Control for second strand synthesis with a single primer HPRT1_reverse_primer_1</td>
<td>25 cycles of a “PCR” (single primer) to amplify the second strand of reverse transcribed cDNA</td>
</tr>
<tr>
<td>3</td>
<td>Sample after second strand synthesis</td>
<td>Control for RT and check of values after reaction with both primers HPRT1_forward_primer_1 HPRT1_reverse_primer_1</td>
<td>qPCR for e.g. 45 cycles with SYBR green master mix in a LightCycler 480</td>
</tr>
<tr>
<td>4</td>
<td>Sample after 1</td>
<td>For Comparison for second strand validation with both primers HPRT1_forward_primer_1 HPRT1_reverse_primer_1</td>
<td>qPCR for e.g. 45 cycles with SYBR green master mix in a LightCycler 480</td>
</tr>
<tr>
<td>5</td>
<td>Sample after 2</td>
<td>For Comparison for second strand validation with both primers HPRT1_forward_primer_1 HPRT1_reverse_primer_1</td>
<td>qPCR for e.g. 45 cycles with SYBR green master mix in a LightCycler 480</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** qPCR (SYBR green based) after sizing with different conditions. Output material was diluted $10^{-3}$, $10^{-4}$, $10^{-5}$ and $10^{-6}$. The test was performed on a LightCycler 480 (Roche Diagnostics).
Beads (ratio given by the manufacturer; Beckman Coulter) the result was a sufficient ladder. Then, multiple different bead ratios are compared to find the proper sizing mixture with the ladder as the sized material.

The volume of beads in conjunction with the volume of the library (in this case the ladder) decides the binding stringency. The sizing buffer used (e.g. from the rapid library preparation kit from Roche Diagnostics or the sizing buffer from New England Biolabs) allows a rebinding of partially eluted sample if the fragment size is high enough. This means by using the sizing buffer in the manner described by the manufacturer undesired fragment sizes are washed away.

The eluted marker is then measured with qPCR for the regions used to generate the ladder. The amplicon size for the qPCR should be between 80 and 120 bp and not the entire length of the sized ladder fragments itself. Ideally, amplicons representing bands below the cut-off should come at high Cqs (if at all) and PCR products above cut-off should have similar low Cq values. As the qPCR input is an amplicon, it is advisable to dilute the sample at least one to one thousand and higher. Different dilution steps (10-fold each time) add information about PCR inhibitions as too high concentrations lead to decreased PCR efficiency. In addition to this fact it is thereby possible to illustrate inhibition of PCR and insure a proper quantification. Fig. 2 is illustrating a sizing ratio comparison for a number of conditions in one amplicon.

This is exemplary and should be repeated with multiple other amplicons representing other ladder steps. The overall quantity illustrates the sizing efficiency for this particular amplicon. Clearly two groups of conditions are formed for this fragment which allow a choice and taken together with the results give the preferred option for sizing.

When a sufficient ratio is found, meaning undesired amplicon sizes are sufficiently reduced, this condition should be employed to yield the wanted outcome. If the Cqs for the unwanted amplicon sizes are higher than in one of the conditions and if the desired fragment sizes remain in their quantity, this ratio should be used for the sizing.

When the real library used for the RNA-Seq is sized according to the established protocol it is necessary to quantify it for calculating the input amount.

2.3.4. Quantification of the library with qPCR

The resulting sized library should again be quantified for input amount calculation of the emulsion PCR (emPCR) necessary for 454 sequencing. The manufacturers protocol states a titration to find the appropriate concentration for an ideal emPCR bead and library ratio. For some time now multiple authors such as Zheng et al. [14] described a titration free method to calculate input amount. QPCR is used to measure the library concentration. The primers are located on the adaptors and only quantify adaptor-linked product. Again it is advisable to run at least two different dilutions of the library for quantification. Fig. 3 depicts an example of such a qPCR run.

In theory, one concentration is enough to insure quantification, but given potential inhibition of the PCR and in some cases unknown usable dilution, it is advisable to test multiple dilution steps. In first experiments they should span from $10^{-2}$ to $10^{-9}$. 

![Amplification Curves](image_url)

**Fig. 3.** Sized library quantification with primers for the Y-Adaptor (SYBR green based qPCR) at two dilutions ($10^{-3}$ and $10^{-6}$). The test was performed on a LightCycler 480 (Roche Diagnostics). The green flat lines depict the non template control.

![Size distribution of a typical sequencing run](image_url)

**Fig. 4.** Size distribution of a typical sequencing run.
The library is then diluted (if necessary) to the appropriate concentration and used as input for the emPCR. From this step onward to the entire sequencing-run itself the manufacturers protocol was used. A typical sequencing output is illustrated in Fig. 4 showing size distribution of reads. Ideally the read lengths sequenced resemble the native size of the starting RNA of 200–400 bp (in case of the 454 Roche Technology this can be the case). Sequencers with regular maximum read lengths below the 200 bp mark should present reads in upper quarter of the maximum length. Subsequent data analysis can be done with multiple bioinformatics tools.

2.4. CNV and mutation analysis – a quick fix

In some case information like CNV brings extra merit to certain sample sets. NGS can under specific circumstances also evaluate copy numbers [15]. However, the experiment has to be perfectly controlled and amplification prior to sequencing introduces a substantial bias that can diminish this possibility. We therefore use qPCR for CNV prior to sequencing to calculate copy numbers and go with these samples for sequencing afterward. For mutational screening and CNV it is useful to sequence, if the number of exons is not unpractical high, amplicons generated with qPCR. A SYBR green or HRM dye based PCR is performed with designated primers. The Cq Values for the exons can be used to calculate CNV when compared to a single copy reference gene whereas the melting curve can give hints, which samples might carry a somatic mutation. Amplicons from qPCR can be used after clean up for subsequent analysis. If PCR products from one sample should be pooled prior to processing (e.g. amplicons of different exons from the same gene) pooling should ideally be done according to the end fluorescence level reached at the plateau phase. Subsequent pools can enter end repair and undergo the procedure similar to RNA-Seq as described before (equivalent to cDNA after second strand synthesis). As an example, initial experiment results prior to sequencing to illustrate HRM and CNV are given in Fig. 5 and 6 with one sample behaving quite differently each (not the same overall sample).

It is important to insure that similar plateaus are reached for the different samples in respect to the amplification curve (and ideally

Fig. 5. Amplification Curve of PIK3CA exon 9 (two wells /same sample show significant lower Cq). The test was performed on a LightCycler 480 (Roche Diagnostics). No distinct difference was to be found.

Fig. 6. Normalized and Temperature shifted difference plot of the same samples (two wells/same sample show significant different melting curve). The test was performed on a LightCycler 480 (Roche Diagnostics). The Yellow sample (two replicates) behaves differently compared to the rest of the patient contingent.
a similar Cq). This makes even minute changes (low percentages of somatic mutations) visible. Late Cqs close to the limit of detection often result in difference plots that are difficult to decipher. Therefore, the amount of input DNA should be sufficient. We usually use 10–50 ng per reaction depending on quality, integrity, total amount and presence of possible inhibitors.

The region in question (PIK3CA exon 9) was normalized and quantified by using POLQ as a reference. Measurement and interpretation is usually done according to [16] al. The described approach allows prior screening for potential changes and subsequent sequencing. In addition each alteration is analyzed with two methods.

2.5. Merging DNA and RNA analysis

In some cases a joined analysis of RNA and DNA at the same time provide additional insights, hence parallel analysis or merging of separate experiments for identical samples on the same sequencer run might be of interest. This is possible in multiple ways. One of the seemingly logical solutions would be total nucleic acid extraction in one elution and reverse transcription in the same volume where the genomic DNA remained. Ideally targeted RNA segments would be analyzed with genomic regions of interest. Proper unbiased RNA and DNA isolation at the same time is not feasible, as most kits seem to favor one analyte type (data not shown). In addition even in case of gene specific RT the genomic DNA would be not enriched for certain areas, but would be present as a whole giving rise to a large quantity of unwanted data during sequencing. Hence an analysis in parallel and later combining for the sequencing run itself, appears to be more feasible solution. By employment of barcodes or MID (multiplex identifier) it is possible to pool different samples. If DNA and RNA is analyzed within the same run using the identical MID for the same source for nucleic isolation is advisable.

3. Materials and methods information

RNA isolation was performed with the Maxwell Blood LEV Kit Promega (with adjusted protocol) [17] and the Roche High Pure FFPE RNA Isolation Kit (version 2012). In general, SYBR green and HRM kit were used from either Bioline (SensiFast SYBR green with no ROX and SensiFast HRM Master) or Roche Diagnostics (LightCycler SYBR green LC480 and LightCycler HRM Master). The Ampure Beads XP were provided from Beckman Coulter. The chemistry for the sequencing itself was completely used from Roche Diagnostics according to the sequencing manual. qPCR and HRM was performed on a LightCycler480 (Roche Diagnostics) and sequencing was done on a 454 GS FLX (Roche Diagnostics). Chemicals like Taq-polymerase and MMLV were provided from Promega.

4. Concluding remarks

Next generation sequencing has become an integral part of molecular biology and specifically oncology. With number of labs performing such experiments however standardisation and quality control have become even more important to insure inter-lab comparability. A technique like RNA-Seq seem to be quite a challenge, but by utilizing qPCR at varies steps the workflow can be significantly improved, either in terms of reproducibility or overall speed. In some cases the use of qPCR and NGS improved and enriched data is a possible outcome.

5. Additional remarks

The presented workflow is optimized for the 454 GS platform (Roche Diagnostic). It can be adjusted to the requirements for other NGS instruments. As Illumina provided sequencers (e.g. MiSeq and HiSeq) are generally shorter in terms of read length, it is advisable to lower the cut-off to 100 bp for the sizing to accommodate this fact. In addition, for RNA-Seq of FFPE no fragmentation should be needed to insure maximum output. The ion torrent platform (Life Technologies) has similar requirements as the Illumina platforms. All instruments use especially designed sequencing primers. This should be taken into account for control primer design as well as when the adapters are custom made.

Acknowledgments

Samples for RNA Seq were isolated with the HighPure FFPET RNA isolation kit (version 2012) in collaboration with Roche Diagnostics (for shown quality control illustration). Samples mentioned for CNV and HRM came from IBIDELL and CIMA for a EU FP7 project called CURELUNG.

References