Methods 59 (2013) 89–100

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review Article

Application of next generation qPCR and sequencing platforms to mRNA biomarker analysis

Alison S. Devonshire [a], Rebecca Sanders [a], Timothy M. Wilkes [a], Martin S. Taylor [b], Carole A. Foy [a], Jim F. Huggett [a,*]

[a] Molecular and Cell Biology, LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY, UK
[b] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine at the University Of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

ARTICLE INFO

Article history:
Available online 24 July 2012
Communicated by Michael W. Pfaffl

Keywords:
Next generation sequencing
NGS
Digital PCR
mRNA biomarker
Nanofluidic PCR
RNA-Seq

ABSTRACT

Recent years have seen the emergence of new high-throughput PCR and sequencing platforms with the potential to bring analysis of transcriptional biomarkers to a broader range of clinical applications and to provide increasing depth to our understanding of the transcriptome. We present an overview of how to process clinical samples for RNA biomarker analysis in terms of RNA extraction and mRNA enrichment, and guidelines for sample analysis by RT–qPCR and digital PCR using nanofluidic real-time PCR platforms. The options for quantitative gene expression profiling and whole transcriptome sequencing by next generation sequencing are reviewed alongside the bioinformatic considerations for these approaches. Considering the diverse technologies now available for transcriptome analysis, methods for standardising measurements between platforms will be paramount if their diagnostic impact is to be maximised. Therefore, the use of RNA standards and other reference materials is also discussed.

1. Introduction

Changes in the expression of multiple genes are implicated in complex diseases such as breast cancer, type 2 diabetes mellitus and cardiovascular disease [1,2]. *In vitro* diagnostic multi-variate index assays (IVDMIAs) utilising gene expression measurements, such as OncotypeDx tests which predict cancer recurrence, have emerged in recent years [3,4]. The pipeline for RNA biomarker panel development involves screening the transcriptome for genes whose expression is associated positively or negatively with disease pathology. Multiple stages of potential marker refinement are required in order to define the best predictors of clinical outcomes coupled with expanded patient cohorts. For example, in the development of Oncotype Dx Colon Cancer Assay, 761 gene candidates were narrowed down to a panel of seven biomarkers and five reference genes, with over 3,000 patient samples screened [5,6].

The DNA microarray is a well-established technique, which has been used to screen for multiple potential gene expression biomarkers and drug targets, and microarray gene expression data continues to be a useful source for mining of potential biomarkers. However, DNA microarrays utilise probes containing known cDNA sequences and therefore do not enable the discovery of novel transcripts and sequence variants [7]. Additionally, limitations in microarray dynamic range make this platform less sensitive in the detection of transcripts of low abundance [8]. Recent technological innovations in the fields of DNA sequencing and PCR address these issues and provide an unprecedented level of information for the discovery and validation of novel RNA biomarkers [9,10].

Next generation sequencing (NGS – also referred to as second generation sequencing) platforms share the common technological feature of being capable of massively parallel sequencing on clonally amplified or single cDNA molecules. This design defines a major shift from "first generation" Sanger sequencing, which was based on the electrophoretic separation of chain-termination products, prepared in individual sequencing reactions. NGS technologies offer the possibility of hypothesis-neutral discovery of novel transcripts and isoforms in a fraction of the time required for genome-wide analysis performed by Sanger sequencing [11,12]. However, multiple template preparation stages, diverse sequencing chemistries and complex data processing of NGS experiments may impact on the verification of *bona fide* RNA biomarkers [13] (Section 4).

Reverse Transcription quantitative PCR (RT–qPCR) technology is central to biomarker validation where potential markers need to be measured with greater accuracy and precision in larger sample sets. A new generation of nanofluidic qPCR platforms has also emerged over recent years which can be used for the simultaneous screening of patient samples for the expression of 10s-100s of candidate biomarkers or enumeration of single copies of cDNA by

* Corresponding author. Fax: +44 020 8943 2767.
E-mail address: jim.huggett@lgcgroup.com (J.F. Huggett).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.07.021

digital PCR (dPCR) (Section 3). Considering the current use of qPCR for molecular microbiological testing in the clinical laboratory, such high-throughput RT–qPCR devices are also likely to be at the forefront of transcript-based diagnostics in the near-future.

The translation of gene expression biomarkers from validated panel to diagnostic test requires assurance of the accuracy and robustness of the developed multi-parametric assay through the use of QC materials and establishment of QA schemes. Potential means of standardisation of multi-parametric RNA biomarker measurements through the use of reference standards are also addressed (Section 5).

This article aims to summarise how this next generation of PCR and sequencing platforms can be applied to different stages of RNA biomarker analysis while highlighting key methodological differences between the varying approaches.

2. RNA as an analyte

2.1. RNA extraction

In order to investigate messenger RNA (mRNA) expression and biomarker profiles, mRNA first needs to be successfully extracted from source material. The variety of biological samples available for molecular analyses has given rise to a multitude of extraction methods, which may confer particular advantage in terms of yield and integrity when utilised for specific sample types. Current approaches include acid phenol/chloroform, silica-column and bead-based extraction methods. Generally, total RNA will be prepared from sample extractions, the majority of which will comprise ribosomal RNA populations [14].

Formalin-fixed paraffin-embedded (FFPE) tissues provide a useful historical source of disease specimens for screening of potential biomarkers [15], however FFPE sections are challenging samples for RNA extraction, due to RNA degradation, cross-linking of RNA to proteins and modification of bases [16]. FFPE RNA extraction methods require deparaffinisation and extended lysis treatment at elevated temperatures; developments in automation of these steps offer potential for high-throughput screening of FFPE samples [17]. FFPE material is amenable to mRNA analysis using established methods like RT–qPCR and microarrays [18–20] and methods for 3'-end digital gene expression profiling by NGS have been developed [21]. Assuming the problems associated with RNA quality do not cause too great a challenge, FFPE samples will provide a valuable source of material for identifying mRNA biomarkers using next generation PCR and sequencing platforms.

While working with FFPE material offers a number of unique challenges, sample sourcing must also be considered when designing gene expression studies to investigate potential biomarkers from 'fresh' material. Some clinical sources such as tissue biopsies are difficult and intrusive to obtain, or may be particularly difficult from which to extract nucleic acid material (e.g. bone) [22–24]. This may lead to great variability in extraction efficiency (yield and quality), particularly when tissue-specific extraction methods are not employed. Consequently, less invasive yet easily handled sources of biological samples, such as blood, urine and buccal swabs, are a popular focus for the development of diagnostic tools.

It is similarly important to take into account tissue variability when planning to obtain samples. Gene expression profiles differ not only between different tissue types, but also between different cell types within the same sample. Furthermore, gene expression can be cyclical and may be influenced by many different genetic and environmental factors; including stress, satiety, nutrition, diurnal fluctuation, exercise, cellular proliferation, disease state and by mitogenic stimuli (e.g. growth factors) [25–31]. When conducting specific gene expression studies it is therefore important to ensure like-for-like samples are used in comparative studies and where possible, only the specific cell-type of interest is collected (for example, separating cell populations by centrifugation, using primary cell culture or laser microdissection) [14,32–35].

It is well known that RNA is more labile than DNA, and as such, precautions must be made in order to achieve the most reliable results. It is recommended that samples be collected in a buffer/preserving reagent suitable for safeguarding RNA against degradation or, alternatively, samples may be snap frozen using liquid nitrogen. The selection of an appropriate reagent may be heavily influenced by the intended down-stream applications (see Section 2.2). Moreover, specific RNA-handling procedures should be applied to reduce the risk of RNase activity. During collection of multiple samples, appropriate fixatives should be employed. Depending on storage buffers/fixatives, samples are usually stored at −20 °C or below until required, then thawed and maintained on ice during the extraction process. Where appropriate, purified RNA should be diluted in a solution designed to maintain RNA integrity, which is free of RNases.

2.2. Inhibition

Extracted RNA samples may be compromised due to the co-extraction of sample components (such as DNA, proteins, bile salts or haeme) or carry-over of chemicals used in sample stabilisation (such as EDTA or heparin) or extraction process (such as chloroform or ethanol) [14,36–39]. Every effort should be made to eliminate these constituents from the final RNA sample. DNA contamination may be further reduced by the application of DNase enzymes. However, no enzymatic reaction can be assumed 100% efficient and as such the presence of Genomic DNA (gDNA) should be monitored and accounted for, otherwise measurement bias may be introduced. Furthermore, if contaminating elements including PCR inhibitors are at reasonably low quantities in the extracted RNA, sample dilution may minimise or effectively eliminate their effect on target measurement.

For clinical application of mRNA biomarker-based diagnostics, thorough characterisation of assay performance should be performed [40] and standards for calibration and QC developed (Section 5). In this context, it is important to remember the influence of matrix effects when choosing an appropriate reference material. Where external standards may be applied for quantification purposes, these must be appropriate to the chosen target and analysed in background material that sufficiently mimics the sample matrix. Ideally, selected external standards should possess similar responses to matrix effects as experienced by the target, and must be spiked into target samples to ensure equal matrices [41].

Matrix-associated inhibition of qPCR may be detected by several different means. The simplest way is to measure samples in serial dilution and monitor linearity of amplification. Reversible inhibition, which may usually be observed at higher concentrations, will materialise as an increase in quantification cycle (C_q) and a decrease in correlation coefficient (R^2) when C_q is plotted against $\log_{10}$ RNA quantity. The SPUD assay has been developed to more accurately determine the extent of qPCR inhibition by measuring an external spike-in from potato (*Solanum tuberosum*) in control (water) vs. target cDNA samples. Analysis of C_q and assay efficiency between control and target samples for the SPUD assay indicates the extent of matrix inhibition [42].

Inhibition of the RT reaction is typically less readily quantified in the course of an RT–qPCR experiment, a factor that is of concern particularly when performing two-step RT–qPCR, where the RT reaction usually contains a higher concentration of both RNA and co-purified inhibitors. Defining the matrix impact on the RT step should be paramount as this reaction is a key component of both RT–qPCR and the majority of current RNA-seq methodologies.

The application of RNA standards (Section 5) and digital PCR (Section 3.3) provide means of monitoring RT reaction efficiency and its impact on downstream processes. In addition to RT, RNA-seq library preparation involves multiple enzymatic stages including fragment end repair and adapter ligation with potential for inhibition by sample components or upstream reaction components. The use of spike-in control DNAs to measure the efficiency of individual steps provides the user with QC data for library preparation stages [43].

3. Next generation RT–qPCR

New innovations in miniaturisation of the PCR have taken real-time qPCR platforms to a new level in terms of sample throughput without the automated infrastructure required for 384 or 1536 well plate formats. Scaling of the PCR reaction from the microlitre down to the nanolitre range dramatically reduces the required volumes of reagents and samples, reducing reagent costs and conserving clinical samples [9]. The increased number of reactions possible in a single run increases the statistical power of an experiment through the use of replicate reactions for qPCR-based quantification or dPCR-based molecular counting [44]. High-throughput RT–qPCR may be particularly beneficial for rapid screening of multiple biomarkers, whereas RT–dPCR may provide unrivalled sensitivity for trace detection. The following two sections provide an overview of the latest generation of nanofluidic PCR technologies and describe their application to high throughput RT–qPCR and RT–dPCR.

3.1. High throughput RT–qPCR

A desirable goal for screening of potential biomarkers and the development of clinical diagnostic tools is to maximise on sample throughput in order to generate multiple, yet accurate, clinical results efficiently and rapidly. Industry has responded to this by the production of several high throughput qPCR chips, where multiple samples and assays can be processed simultaneously with minimal complexity. Examples of market providers include the Biomark Dynamic Array (Fluidigm), OpenArray (Life Technologies) and the SmartChip (Wafergen).

3.1.1. BioMark

BioMark 48.48 and 96.96 dynamic arrays (Fluidigm) are microfluidic chips consisting of a network of capillary channels, Nanoflex™ valves and reaction chambers (Fig. 1A) [9]. Samples and assays are distributed within the Integrated Fluidic Circuit (IFC) of the chip by the IFC Controller such that each of the samples loaded into the 48 or 96 sample inlets is assayed for each of the

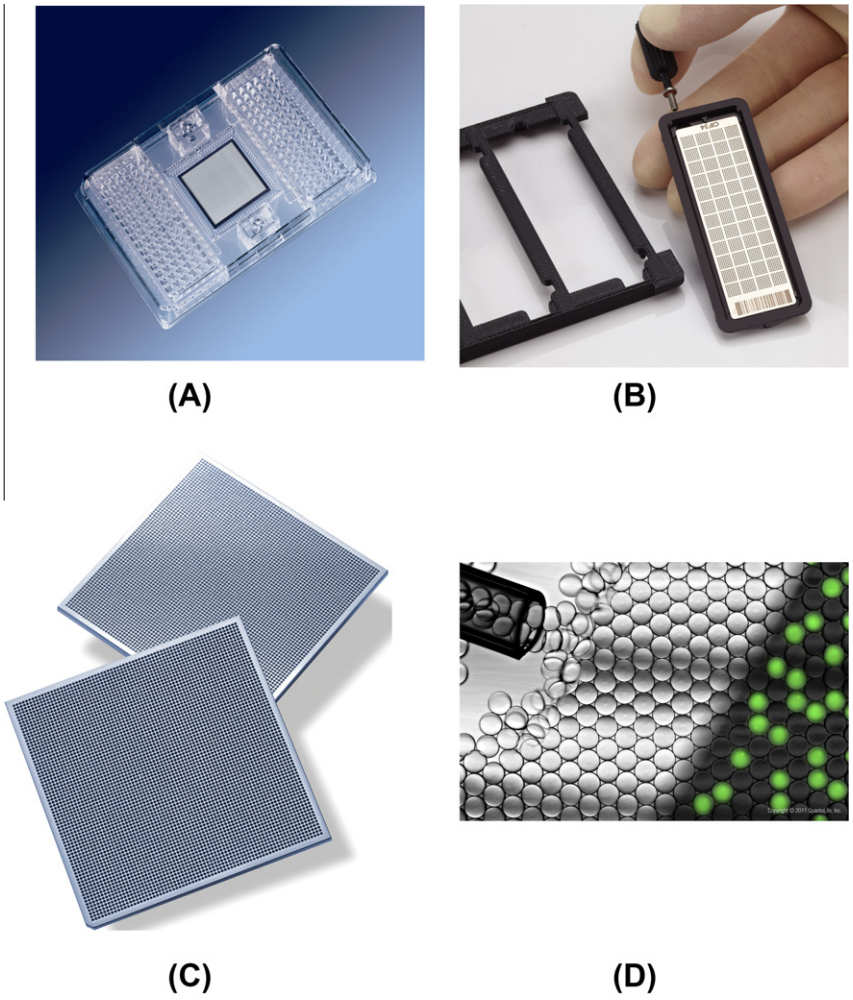

Fig. 1. Microfluidic PCR platforms. (A) Biomark 96.96 dynamic array with 96 assay inlets and 96 sample inlets with central integrated fluidic circuit (image courtesy of Fluidigm). (B) OpenArray stainless steel plate containing 48 sub-arrays of 8 × 8 reaction through-holes (image courtesy of Life Technologies). (C) Two SmartChip arrays (image courtesy of Wafergen). (D) QX100™ Droplet Digital PCR™ system: A microscopic image of a sample emulsified into tens of 1,000s of nanolitre volume droplets. Only droplets that contain a copy of the target molecule will fluoresce after PCR (image courtesy of Quantalife).

Table 1
Performance characteristics of selected high throughput nanofluidic RT–qPCR platforms currently on the market.

System	Reaction volume (nL)	Total no. of reactions	Maximum no. of gene targets per sample	Maximum no. of samples	Melt curve capacity
Biomark Dynamic Array (48.48 or 96.96)	10 or 6.75	2,304 or 9,216	48 or 96	48 or 96	Yes
OpenArray	33	3,072	224 (12 samples)	48 (64 assays)	Yes
SmartChip	100	5,184	1728	384	Yes

assays loaded into the 48 or 96 separate assay inlets, resulting in a total of 2,304 or 9,214 individual reactions per array for 48.48 or 96.96 arrays respectively (Table 1). The BioMark array qPCR chamber volume of 10 nL (48.48 arrays) or 6.75 nL (96.96 arrays) is the lowest of the nanofluidic platforms discussed here [45]. Both hydrolysis probe- and intercalating dye-based qPCR assays may be performed on the BioMark nanofluidic qPCR system and unlike other platforms the dynamic array assay sets are loaded by the user, enabling versatility.

3.1.2. OpenArray

The OpenArray comprises a microscope slide-sized stainless steel chip consisting of 48 sub-arrays each containing 64 (8×8) through-holes (reaction chambers), with a total of 3,072 qPCR chambers per array (Table 1, Fig. 1B) [46]. For qPCR applications, assays are pre-spotted on the interior surface during the manufacturing process. The associated PCR cycler can process three chips simultaneously. Both custom arrays and off-the-shelf target panels to a range of physiological pathways, including ADME/Tox, Cancer, Cardiovascular Disease and Inflammation, are available for screening samples [47].

3.1.3. SmartChip

SmartChip system (Wafergen) is the latest nanoscale qPCR platform to emerge onto the market (Fig. 1C) [48]. Assays are either pre-dispensed or customer-dispensed in PCR nanowell plates, like OpenArray or BioMark arrays respectively, on PCR arrays and the number of different gene targets may be user-defined. This allows flexibility in terms of profiling 100–1000s of genes for the purpose of biomarker discovery, or a small number of genes for screening large sample numbers. Off-the-shelf panels are also available for this platform, such as the Human Oncology Panel, which contains gene-specific assays for over 1200 targets in 16 cancer-related functional groupings [49].

3.2. dPCR

dPCR is a development on standard PCR that utilises single molecule amplification for absolute quantification [50]. In this method limiting dilutions of samples are used to facilitate sample distribution across multiple reaction chambers at single copy densities. The presence of target is indicated by amplification and positive reaction chambers are counted to obtain the number of target copies, thus converting the analogue signal associated with qPCR into a digital one [51]. Amplification of single target molecules reduces the signal-to-noise ratio offering increased assay sensitivity. This is particularly useful for detection of minority targets, for example early disease markers/splice variants, and trace RNA detection (low copy number) such as single cell samples [52]. dPCR measurement of DNA is often described as an absolute quantification technique and there is increasing evidence that this approach may indeed offer one of the most accurate estimations of DNA copy number [36,51]. This approach may be applied to RNA quantification by combining a reverse transcription step with dPCR and may provide a means of certifying RNA reference materials for application in biomarker diagnostics.

The advent of next generation PCR instruments has changed dPCR from a technically challenging and laborious technique to a very powerful method with considerable potential. There are several platforms currently available for dPCR analysis, based on microfluidic chip- or emulsion-based formats.

3.2.1. Microfluidic chip-based dPCR

Microfluidic PCR platforms which can be used for dPCR analysis include the Biomark (Fluidigm) and OpenArray (Life Technologies). For the BioMark, two dPCR chip formats are available (Table 2), working on similar principles to the related Dynamic arrays (Section 3.1.1.). For the OpenArray, the number of sub-arrays used to analyse a sample can be altered according to the level of sensitivity or precision required (Life Technologies, personal communication). Such dPCR platforms are characterised by single molecule amplification in pre-fabricated microfluidic reaction chambers and amplification can be monitored in real time. Results are generated by calculating the number of positive reaction chambers (Fig. 2A) and equating it to the number of target copies present.

3.2.2. Emulsion dPCR

Emulsion PCR employs individual vesicles, formed by water-in-oil emulsions, as reaction vessels for dPCR. The QX100™ Droplet Digital PCR™ system (Bio-Rad), formerly the QuantaLife Droplet Digital PCR™ platform, couples the emulsion vesicles to a capillary-based analysis system [53]. Samples are emulsified in an 8-channel disposable droplet generator cartridge then transferred to a standard 96-well plate for thermal cycling. Following PCR amplification, end-point detection of PCR products is achieved by streaming vesicles singularly past a fluorescence detector at a rate of 1,000 droplets per second, which determines the presence or absence of amplified product (Fig. 1D) [54]. The recently launched RainDrop™ dPCR platform (RainDance) uses a flow cytometry approach in order to count positive PCR reactions and higher order assay multiplexing is possible with this approach based on

Table 2
Performance characteristics of selected dPCR platforms currently on the market.

System	Reaction volume	Total no. of reactions	No. of reactions per sample	Maximum no. of samples	Melt curve capacity
Biomark Digital Array (12.765 or 48.770)	6 or 0.85 nL	9,180 or 36,960	765 or 770	12 or 48	Can use intercalating dye but no melt analysis
OpenArray	33 nL	3,072	3,072 (1 sample)	48 (64 reactions)	Yes
QX100	~1 nL	160,000	20,000	8	End-point analysis
RainDrop	~5 pL	80 million	1–10 million	8	End-point analysis

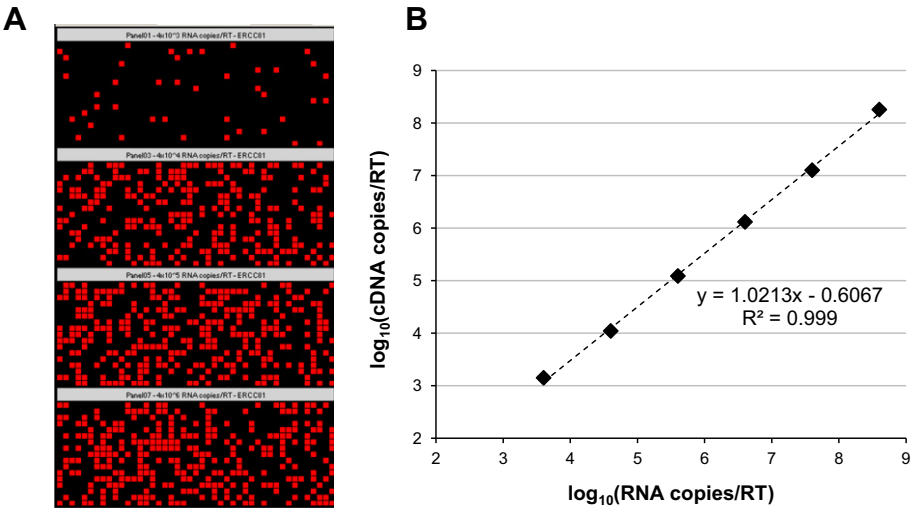

Fig. 2. dPCR quantification of RT efficiency. RT reactions were performed with 400 ng total Universal Human Reference RNA (Stratagene) spiked with different copies of an ERCC RNA standard in a volume of 40 μL as described in [56] and cDNA copies measured by dPCR (BioMark 12.765 dPCR arrays). (A) Heatmap of dPCR panels. Red and black squares represent positive and negative PCR reactions respectively. (B) Comparison of RNA input with cDNA copies produced per RT reaction. Mean cDNA copies per RT reaction ($n = 1$) were calculated from duplicate dPCR panels. Slope, intercept and R^2 values plotted for linear regression analysis of RNA vs. cDNA copies.

different concentrations of reporter dye [55]. Summary details of these technologies are also provided in Table 2.

3.3. Preparation and manipulation of RNA extracts for analysis

For most next generation PCR platforms, samples may be prepared using similar processes as those used for routine RT–qPCR, however as with standard approaches, the RT step is critical for both RT–qPCR and RT–dPCR. Target pre-amplification may be necessary for accurate measurement of low concentration targets by nanofluidic RT–qPCR due to the lower reaction volumes compared to standard microtitre plates (Table 1). These aspects are discussed below.

3.3.1. Reverse transcription

The RT component of gene expression studies is notoriously variable and is dependent on many factors, including RNA quality, choice of reverse transcriptase, priming strategy and inhibition [14,27,33,40]. Analysis by one-step RT–qPCR (or RT–dPCR) affords a greater scrutiny of the RT step and associated efficiency, automatically considering RT–qPCR sensitivity and linearity, important factors that are often neglected when performing two-step RT–qPCR, where calibration curves and replicates are often performed with cDNA instead of RNA.

Fig. 2 shows an example of how dPCR can be applied for absolute quantification of cDNA copies in order to monitor RT performance, in combination with an RNA standard developed through the External RNA Controls Consortium (ERCC) (for further information, see Section 5). A range of copy numbers of an ERCC RNA standard was spiked into human total RNA as described [56] and cDNA copies measured by dPCR (following dilution as required). These results demonstrate good linearity of RNA to cDNA conversion across the range of different transcript abundance levels ($R^2 > 0.99$) and suggest that less than 50% of RNA transcripts are converted into cDNA (based slope of 1.02 and negative intercept of −0.6) (Fig. 2B), in line with other reports of RT efficiency [57,58].

3.3.2. Pre-amplification

Pre-amplification of specific targets for applications using nanofluidic RT–qPCR is normally performed using a PCR-based approach (also known as specific target amplification (STA)), whereby cDNA undergoes 14–18 cycles of PCR-amplification with a low concentration mix of up to 100 different primer pairs [59,60]. Investigation of relative gene expression levels in pre-amplified samples using the Biomark dynamic array indicates that results are concordant with those obtained from established qPCR platforms using non-pre-amplified samples. Furthermore, these studies showed that substantial bias was not introduced by performing this pre-amplification step [45,61,62]. For BioMark 48.48 arrays (reaction volume 10 nL), it is possible to analyse samples containing ⩾250 ng/μL total RNA or to study high abundance transcripts (present at 10^4 copies or higher per sample inlet or 250 copies or more per reaction chamber) without pre-amplification [62]. As the reaction volumes of the OpenArray and Wafergen platforms are higher than those of the BioMark arrays (Table 1), quantification of lower concentration samples or targets may be possible without pre-amplification [63].

4. Next generation sequencing

Next generation RNA sequencing approaches fall into two main groups: full length RNA sequencing (RNA-Seq) and RNA-tag digital gene expression profiling (DGE). 3′-tag DGE is the next generation form of serial analysis of gene expression (SAGE) and measures the expression of transcripts based on sequencing of 3′ tags [64]. Whilst DGE enables high-throughput mRNA quantification, full-length RNA-Seq is capable of identification and quantification of promoter usage, alternative splicing and chimeric transcripts due to genomic rearrangements and non-polyadenylated transcripts [65–67].

Technologies used for RNA-Seq experiments include 454 (Roche), Solexa (Illumina) and SOLiD (Life Technologies). These approaches all depend on unique sequencing chemistries, which in turn necessitate different methods for library preparation and analysis of sequencing data. These differences also present significant challenges when attempting to compare platform performance and data quality, and to develop appropriate standardisation approaches. The following sections provide an overview of the three main stages in an RNA-Seq experiment: library preparation (Section 4.1), sequencing and imaging (Section 4.2), and data analysis (Section 4.4). Key methodological approaches are also

discussed which may impact on the quality and coverage of transcriptomic data.

4.1. Library preparation

4.1.1. mRNA enrichment

Most NGS library preparation methods require that RNA is enriched for the population of interest or include a polyA-enrichment step in the protocol (Table 3). While polyA-enrichment provides good coverage of mRNA transcripts, co-enrichment of non-polyadenylated RNAs facilitates the parallel study of non-coding transcripts and offers a more complete view of the transcriptome [68,69]. Ribodepletion methods such as RiboMinus (Invitrogen) or Ribo-Zero (Epicentre) remove 5S, 5.8S, 18S and 28S human rRNAs using 5′-biotin labelled oligonucleotide probes combined with their removal using streptavidin-coated magnetic beads. Alternatively rRNA and tRNA can be removed from cDNA synthesis using a Duplex-Specific Nuclease (e.g. Evrogen). The stringency of selection for RNAs of interest is an important step for maximising the utility of the reads produced, with protocols for SOLiD sequencing recommending two rounds of oligodT-based purification or a combination of both polyA-enrichment and ribodepletion methods [70].

Alternatively, 5′ exonuclease digestion of total RNA selectively removes ribosomal RNA, whilst mRNA bearing a 5′ methylguanosine cap is protected (e.g. Terminator™ 5′-Phosphate-Dependent Exonuclease, Epicentre). This offers the benefits of co-purification of both eukaryotic and prokaryotic mRNAs and no losses due to column- or magnetic bead-binding steps required for polyA-enrichment and ribodepletion methods. Cap analysis of gene expression (CAGE) approaches also target the 5′ end of transcripts by chemical ligation with biotin or using a cap-binding protein with the aim of mapping transcriptional start sites (TSS) and regulatory networks within the genome [71–73].

Further enrichment of the RNA sample in order to study allelic expression, splicing events or gene fusions in specific subsets of genes can be performed by hybridisation-based array (e.g. Nimblegen Sequence Capture Array) or in-solution (e.g. Nimblegen SeqCap EZ, Agilent Sureselect) capture methods [66]. Targeted amplicon sequencing of specific exons may be performed by PCR-based amplification of the regions of interest; examples of which include the emulsion PCR-based ThunderStorm system (Raindance Technologies) and nanofluidic Access Arrays (Fluidigm). Targeted sequencing of specific transcripts is amenable to focussed experiments using sequencing platforms which produce smaller numbers of reads such as Roche/454 (Section 4.2.1) or benchtop sequencers such as the Illumina MiSeq (Section 4.2.2) and Ion Torrent (Section 4.3.2). Nimblegen Sequence Capture Arrays targeted to a set of 50 transcribed loci were recently combined with 454 sequencing for in depth analysis of alternative exon usage and splicing patterns [74].

4.1.2. Fragmentation

For RNA-Seq, fragmentation may be performed with RNA or post reverse transcription using double-stranded cDNA (ds cDNA) using chemical (e.g. hydrolysis with zinc chloride), enzymatic (e.g. RNase or DNase) or physical (e.g. heat, sonication or nebulisation) methods [11]. RNA fragmentation methods such as those for 454 and Illumina RNA-Seq library preparation, employ heat treatment of the sample, which offers the advantage of denaturing RNA secondary structure prior to RT [75]. Enzymatic methods of RNA or ds cDNA cleavage are also employed for the SOLiD RNA-Seq protocol and in 3′ DGE methods for generation of 3′ tags (Table 3). For RNA-Seq analysis of gene fusion events, restriction enzyme digestion of cDNA also provides a means of identification of spurious gene fusion reads by the presence of enzyme recognition site within the sequence [76].

4.1.3. Reverse transcription

Oligo(dT)-priming of RT provides another level of polyA-RNA selection, however this priming method has been reported to lead to under-representation of 5′ end of transcripts [75], although methods which select for full length cDNA can ameliorate this issue [77]. For whole transcriptome RNA-Seq, the majority of the manufacturers' protocols employ random-primed RT (Table 3). DGE profiling methods normally employ a 3′ oligo(dT)-primed RT step following binding of mRNAs to magnetic capture beads (Table 3).

4.1.4. Strand-specificity

A number of methods have been published which enable knowledge of the strandedness of RNA molecules to be maintained (summarised and compared by Levin et al. [83]). Information on whether the sequenced molecule originates from the sense or antisense strand allows the identification of regulatory non-coding RNAs and overlapping transcripts originating from opposite strands [83]. SOLiD RNA-Seq protocols utilise the ligation of adapters prior to RT (Table 3) and whilst official strand-specific protocols for Illumina and 454 are not available, other commercially available methods for these platforms such as ScriptSeq (Epicentre) utilise tagging of 5′ and 3′ ends with specific sequences. Alternative open methods, which adapt a dUTP-based second strand-marking method [84], have been developed in order to reduce the cost of library preparation [85].

Two alternative library preparation methods can be used in order to validate whether differentially expressed transcript

Table 3
Summary of library preparation methods for RNA-Seq for major platforms.

Platform	Amount	Enrichment	Fragmentation substrate(Method)	Strand-specificity	Adapter ligation	RT enzyme (priming)	Ref.
SOLiD RNA-Seq	Poly(A) RNA: 100–500 ng rRNA-depleted total RNA: 200–500 ng (5–25 ng of poly(A) RNA low input)	Flexible Recommend 2×Oligo (dT) enrichment	RNA (RNase III or chemical)	Addition of adapters in directional manner	Pre-RT	ArrayScript engineered MMLV (random)	[78]
SOLiD SAGE	1–10 μg total RNA	Dynabeads® Oligo(dT)	ds cDNA (*Nla* III restriction digestion)	Yes	Post-RT	SuperScript III (oligodT)	[79]
Roche 454	200 ng enriched RNA	Flexible	RNA ($ZnCl_2$ sol^n, 70 °C, 30 s)	No	Post-RT	AMV (random)	[80]
Illumina TruSeq	0.1–4 μg total RNA	Oligo(dT)	RNA (chemical, 94 °C, 5 min)	No	Post-RT	SuperScript II (random)	[43]
Illumina 3′ DGE	1–2 μg	Oligo(dT)	cDNA (NlaIII/ DpnII)	Yes	Post-RT	SuperScript II (oligodT)	[81,82]

sequences are artefacts due to library preparation methods. For example, a strand-specific RNA fragmentation protocol was used to validate novel transcriptionally active regions (nTARs) expressed in mouse intestine identified using a polyA-enriched cDNA fragmentation method [77].

4.1.5. Library quantification

Although NGS platform manufacturers advocate spectrophotometric methods, including the pico-green assay [86], for library quantification, there have been reports that this technique is a source of major inconsistency in the template preparation process [87]. Spectrophotometric methods are unable to distinguish between any adapter or genomic DNA contamination and the intended cDNA target. Quantification using capillary electrophoresis methods (for example, the Agilent Bioanalyzer) have been described in library preparation studies by the Sanger sequencing group [87] and it is worth noting that analysis of fragment size distribution by platforms such as the Bioanalyzer also functions as a quality control step for library preparation prior to sequencing.

Alternative strategies for library quantification by qPCR have been developed employing PCR primers binding sequencing adapters [87–89]. qPCR-based approaches may be particularly more accurate for quantification of low copy numbers in 1–100 pM range [87]. Standards for the latter approach have included a previously sequenced library or the generation of plasmid standards. However, questions have been raised that the amplification of the designated plasmid standard by qPCR may not reflect that of molecules of differing length and GC content in the fragment library [89]. Since dPCR (Section 3.2) does not require a calibrant material for absolute quantification, approaches such as the Slingshot kit (Fluidigm) may enhance the accuracy of library quantification [90].

4.2. Platforms

The most commonly used NGS platforms for RNA-Seq vary in the method chemistries used for massive parallel sequencing. The sequencing chemistries of four platforms are summarised below alongside advantages and disadvantages of the platforms for RNA-Seq; readers are referred to other reviews for further information on NGS technologies [12]. Specifications of the main NGS platforms are given in Table 4 for the purpose of comparison; this is a rapidly evolving field with advances in output in terms of number of reads and read length constantly being reported. Short read platforms (e.g. Illumina, SOLiD) typically produce a greater number of reads and hence enable higher coverage of the transcriptome, which is ideal for discovery of potential biomarkers [91,92]. Approaches offering longer read length combined with higher accuracy provide a means of validation of biomarkers and confirmation of splice variant structure and sequence [76].

4.2.1. Roche 454

The Roche 454 NGS systems work on the principle of pyrosequencing, where the release of pyrophosphate upon nucleotide incorporation results in luminescent signal output [97]. RNA-Seq libraries consisting of fragmented target cDNA are amplified en masse on the surfaces of hundreds of thousands of droplet encapsulated agarose beads using emulsion PCR (Section 3.2). These are then applied to the surface of the 454 picotiter plate (PTP) which consists of single wells in the tips of fused fibre optic strands that can each hold a single agarose bead. Imaging of the PTP following cyclical addition of each of the four base nucleotides serves to measure light emission as a consequence of nucleotide incorporation.

The relatively long read length of the 454 system (Table 4) compared to the Illumina and SOLiD systems offers advantages for confirmation of alternative splicing events in RNA-Seq experiments [76]. Accurate quantification of homopolymeric sequences may be problematic for 454 sequencing as the linearity of response can exceed the level of detector sensitivity, a recognised issue with pyrosequencing, leading to insertion/deletion (indel) errors [98].

4.2.2. Illumina

Unlike the 454 and SOLiD technologies which employ emPCR, the Illumina NGS platforms achieve target amplification in the flow cell by "bridge" amplification which relies on captured DNA strands "arching" over and hybridising to adjacent oligonucleotide anchors. Multiple amplification rounds convert single-molecule DNA template to clonally amplified arching clusters, with each cluster containing in the region of 1000 clonally amplified molecules. Illumina sequencing works on the principle of reversible termination with each sequencing cycle involving the addition of DNA polymerase and a mixture of four differently coloured reversible dye terminators followed by imaging of the flow cell. The terminators are then unblocked and the reporter dyes cleaved and washed away. Following sequencing from a single end of the template, paired-end sequencing can be achieved by sequencing from an alternate primer on the reverse stand of the template molecule.

Illumina NGS technology offers the advantage for RNA-Seq and DGE of the highest data output of the major platforms (Table 4). This enables the interrogation of low-abundance transcripts [64]. For RNA-Seq, paired-end read data derived from sequencing from both ends of the sequencing template enables information regarding splice junctions and fusion transcripts to be obtained [99]. Base-call accuracy decreases with increasing read length on Illumina NGS platforms due to "dephasing noise" due to under- or over-incorporation of nucleotides or failed terminator removal with successive cycles leading to the generation of a heterogeneous target-strand population within the cluster. This heterogeneity decreases signal purity and reduces precision in base calling, particularly towards the 3′ ends of a read [100].

The recent addition of the MiSeq benchtop sequencer to the Illumina portfolio enables smaller-scale discovery for individual

Table 4
Specifications of NGS platforms commonly used for RNA-Seq.

Platform	Clonal amplification method	NGS chemistry	No. of reads/run	Read length (bases)	Run time (days)	Reference
Roche/454 GS FLX Titanium XL+	emPCR	Pyrosequencing	1 million	700 (mode) 1000 (max)	23 h	[93]
Illumina Genome Analyzer IIx	Solid-phase bridge amplification	Sequencing by reversible termination	320 million single-end/ 640 million paired	50[a] 150 (max)	7 days (SE), 14 (PE) (151 cycles)	[94]
Illumina HiSeq 2000			3 billion single-end/ 6 billion paired	50[a] 100 (max)	8.5 days (2×100 bp)	[95]
Life Technologies SOLiD 4 (5500)	emPCR	Cleavable probe sequencing by ligation	100 million	50×35 (PE) (average) 75×35 (PE) (max)	3.5 days	[96]

[a] 85% of base calls from 50 base paired-end (PE) read pass quality score Q30 (chance of wrong base call: 1 error in 1,000).

research laboratories, generating up to 6–7 Gb data output and 30 million paired end reads per run (compared to 95 Gb and 600 Gb for Genome AnalyzerIIx and HiSeq respectively). Library preparation for this platform can be condensed to only 90 min using Nextera sample preparations kits (Illumina) which combine fragmentation, adapter ligation and barcoding of samples [101], an approach which may be suitable for targeted sequencing of transcripts.

4.2.3. SOLiD

The SOLiD (Sequencing by Oligo Ligation and Detection) platform employs a sequencing process catalysed by DNA ligase. Similar to the 454 technology, DNA fragments are amplified by emPCR while bound to the beads, after which the beads are covalently bound to the surface of a specially treated slide which is then placed into the fluidics cell of the instrument. Sequencing is initiated by the annealing of a universal sequencing primer to the adapters of the fragment library followed by addition of semi-degenerate fluorescently-tagged 8mer-oligonucleotides, which are ligated to the universal primer by DNA ligase when complementary to the sequence of interest. Following imaging of the slide, a subsequent cleavage step removes the sixth through to the eighth base, plus the fluorescent tag of the ligated 8-mer and a further nine ligation rounds performed [12]. The sequencing strand is then denatured and washed away and a second round of sequencing is performed using a universal primer of $(n-1)$ bp in length. A further three rounds are performed so that each base of the interrogated fragment is sequenced twice.

SOLiD technology has a high accuracy rate for raw reads (>99.9%) due to the double interrogation of each base, and that procedure requires a lower volume of oversampling in order to reach a threshold value of confidence for base calling. This offers advantages for SNP discrimination by RNA-Seq [65].

4.3. Developments in sequencing technologies

A third generation of sequencing platforms which offer further advantages for transcriptional profiling in terms of reductions in sequencing time and cost are emerging [102]. Examples of single molecule sequencing and non-optical sequencing technologies are described below.

4.3.1. Single Molecule Sequencing

In comparison to the 454, Illumina and SOLiD platforms, which sequence clonally amplified templates, single molecule sequencing technologies such as the PacBio RS (Pacific Biosciences) and Heliscope (Helicos Biosciences) sequencers are capable of single-molecule sequencing. The PacBio RS system uses "single molecule real-time detection" (SMRT) which detects the fluorescence of a labelled nucleotide as it is incorporated into the growing DNA strand. Fluorescence from a single DNA polymerase molecule is measured per perforation in a metal sheet containing 75,000 such perforations. As the fluorescent label is initially attached to the dNTP phosphate group, it is cleaved during nucleotide incorporation. Therefore there is no need to reverse terminators and, as each nucleotide is separately labelled, no need to cyclically alternate the availability of nucleotides [103]. The PacBio RS is designed to produce average read lengths greater than 1,000 base pairs [104] providing a useful complement to existing technologies for establishing long-range contiguity.

The Helicos system performs "true single molecule sequencing" (tSMS) of DNA or RNA molecules captured on its flow-cell surface. In principal the Helicos approach is similar to that of Illumina, where reversible cy-5 labelled terminators for each of the four nucleotides are cyclically presented for incorporation into the extending DNA strand, a complete cycle of four nucleotides is termed a "quad". Typically 20 to 30 quads are performed resulting in read-lengths of 25–55 bases (average 35). The key differences with Illumina being that there is no target amplification and single molecule fluorescence is detected [105]. The single molecule approach eliminates the dephasing problem inherent in the Illumina platform but the small signal from single molecule fluorescence leads to an increased per-nucleotide error rate and a much elevated frequency of missing nucleotide calls that manifest as single nucleotide deletions in the resulting sequence.

For RNA-Seq using the Helicos system, single-stranded cDNA or RNA template are hybridised to poly (dT) oligonucleotides, which have been immobilised onto a flow-cell surface at a high density [106,107]. Alternatively, low quantity samples can be hybridised directly for 3′ digital gene expression profiling [108]. Single molecule cDNA or RNA sequencing avoids the need for adapter ligation, clonal amplification and therefore these approaches are free from biases associated with these steps [109]. The Helicos Direct RNA Sequencing (DRS) technology does not require prior conversion of RNA to cDNA, since transcripts hybridise directly to poly(dT) on the sequencing flow cell [107,110]. Single molecule sequencing is therefore particularly suitable for low-quantity RNA samples which would otherwise require pre-amplification, such as those from clinical biopsies [108,110]. However, the efficiency of hybridisation of the template to the flow cells (estimated to be 10–20%) and efficiency of the SMS chemistry in producing useable reads (15–25%) are also limitations for this platform, especially for low quantity samples [108].

4.3.2. Ion Torrent

The Ion Torrent (Life Technologies) platforms constitute a shift in technology from optical-based sequencing systems which measure fluorescence or luminescence output, to monitoring release of hydrogen ions during DNA synthesis in a semiconductor-sensing device [111]. The Personal Genome Machine (PGM)™ sequencer can generate up to 8 million reads and is suitable for analysing small transcriptomes or expression of targeted transcripts. RNA-Seq libraries for this platform are prepared with as little as 200 ng total RNA using the Ion Total RNA-Seq Kit (Ambion) and sequencing runs are performed in less than one hour with read lengths anticipated to exceed 400 bp by 2012 [112,113].

The recently launched Ion Proton platform brings personalised genomic analysis within the realm of clinical medicine with the promise of $1000 whole genome sequencing [114]. This platform also offers possibilities for individualised transcriptome analysis [115].

4.4. RNA-Seq data analysis and bioinformatics

Analysis of RNA-Seq experiments poses challenges in both qualitative and quantitative data interpretation [75]. In contrast to conventional Sanger sequencing, current NGS platforms typically generate shorter reads with higher per-nucleotide miss-call rates and on some platforms markedly more insertions or deletions. They also produce many more sequence reads (Table 4). Benchtop sequencers such as the Illumina MiSeq (Section 4.2.2) or Ion Torrent PGM (Section 4.3.2) offer more streamlined workflows for RNA-Seq and generate more manageable quantities of data, which may more suitable to the clinical laboratory end-user for future diagnostic applications [116].

A key objective in most NGS analysis is to assign equivalence to sequence reads, i.e. identification of specific transcripts, RNA processing events, DNA binding events or polymorphisms. This either entails an 'all versus all' comparison of sequence reads, or aligning reads to a common comparator such as the reference human genome or a database of splice junctions. The need to accommodate the high rate of sequence errors in NGS data and also genuine

sequence differences such as polymorphisms substantially add to the complexity and thus computational time of the task of sequence alignment.

The majority of RNA-Seq analysis currently depends on a strategy of reference genome alignment. This provides a common framework for comparisons between studies and allows for simple comparison to reference annotation but it is not without its limitations. The human genome is highly repetitive and some regions are not unique, so it is impossible to unambiguously assign a sequence read to a locus, or worse for quantitative analysis, some regions are systematically depleted but not devoid of aligning reads. Methods have been developed that attempt to correct for this in transcriptome [117] and RNA tag sequencing [118]. Pre-computed genome uniqueness measures are available from the UCSC genome browser (http://genome.ucsc.edu) as the mappability track and are useful in identifying such problems. It is also known that polymorphisms and mutations cluster in the genome [119], a cluster of differences from the reference genome is likely to be both of considerable biological interest and systematically missing or under-represented from alignments of short read sequence data. In the absence of a reference genome, or in order to overcome some of the limitations in reference alignment, there are several tools available for *de novo* RNA-Seq assembly [120].

Qualitative conclusions which can be drawn from an RNA-Seq experiment include identification of previously unrecognised transcripts and exons, alternative splicing patterns, discovery of SNPs and the identification of allele-specific expression [121,122]. However, transcriptome sequence is a complex dataset where multiple splice isoforms may be represented, transcription units often overlap and antisense transcription is abundant [123]. Deconvolving these mutually confounding signals is a major and ongoing challenge, but there have already been a number of encouraging successes, particularly in the discrimination of distinct transcript isoforms [124–126]. There are also advantages from using less complex NGS datasets such as RNA tag sequencing methods like SAGE [64] and CAGE [73]. These can provide a quantitative measure of expression along with strand specificity and also encode information about transcript processing such as poly-adenylation and 5′-capping which can be of great importance in discriminating overlapping transcription units [127]. In principal, splice junction traversing reads from transcriptome sequencing could also be used in a similar way.

Quantitative interpretation of RNA-Seq experiments enables digital measurement of transcript abundance and inference of differential expression of biomarkers between experimental groups or clinical conditions. Mapped reads are quantified based on number of reads mapping to a transcribed region. Numbers of reads are dependent on both transcript abundance and length, necessitating normalisation of read counts to transcript length. The most commonly used metrics are Reads Per Kilobase per Million of mapped reads (RPKM) for single-end read data and Fragments Per Kilobase of exon per Million fragments mapped (FKPM) for paired-end read data [7,128]. Due to their length, short transcripts have fewer mapped reads than longer transcripts which can be a confounding factor for statistical analysis of differential expression [129].

Finally, differential gene expression analysis of normalised reads from RNA-Seq experiments requires the application of statistical models such as the Poisson or empirically calibrated distributions to estimate sampling probabilities contributing to read frequencies [8]. Similar to the development of methods for assessing microarray data normalisation and false discovery rates a decade ago, these issues are being addressed with new software packages for RNA-Seq data, however standardised procedures for RNA-Seq have not yet been established (reviewed in [130]).

5. Standardisation of RNA biomarker measurements

Increased translation of RNA biomarkers from research into clinical practice requires discovery of new biomarkers, validation of potential candidates and development of standardised methods for the clinical laboratory [56,131,132].

RNA-Seq offers the possibilities of discovery of novel transcriptionally active regions and splice variants with roles in disease pathology [68,76,133]. However despite good concordance between RNA-Seq results and those obtained from microarray and RT–qPCR platforms [7,8,134], recent reports suggest that technical variability and sampling issues can potentially cloud conclusions drawn regarding differentially regulated transcripts [135]. Efforts to determine the performance of different NGS methodologies are underway, with Phase III of the Microarray Quality Control (MAQC) consortium, the Sequencing Experiment Quality Control consortium (SEQC), coordinating evaluation of the major platforms by producing benchmark data sets using reference samples, and by comparing different bioinformatics pipelines. The use of reference samples were fundamental to microarray performance characterisation in the initial phases of MAQC [136], and one of these reference samples (Universal Human Reference RNA, Agilent) is recommended as a control for Illumina RNA-Seq [43].

Synthetic spike-in materials are recognised to be useful in testing of new RNA-Seq methods [85,110], however application of the same set of universal reference standards are required in order for cross-platform, inter-laboratory or inter-sample comparisons to be meaningful. ERCC RNA standards have been developed to meet this measurement need of the bio-analytical community through the initiative of the External RNA Controls Consortium, an *ad hoc* group of 70 members from private, public and academic organisations led by the National Institute of Standards (NIST) [137]. Panels of ERCC RNA standards have already been employed to evaluate the accuracy, limits of detection and biases due to GC content or transcript length of RNA-Seq using the Illumina Genome Analyzer II [138]. Commercially available pools of ERCC RNA standards (Ambion) are also recommended controls for SOLiD library preparation [78]. RNA standards are a useful source for calibration by relating read counts to RNA copy numbers and can provide information on the efficiency of processing steps during library preparation and template amplification [138].

In addition to universal reference materials for ensuring that future diagnostic assays meet defined general performance criteria for accuracy, precision and robustness, reference standards for the specific RNA biomarkers being assayed will also be required for routine QC of patient results [131]. In the area of molecular oncology, the CDC Genetic Testing Reference Materials Coordination Program (GeT-RM) provides information on sources of cell lines and DNA which are available for testing of genetic alterations observed in leukaemias, lymphomas and solid tumours, as well as listing sources of certified and standardized reference materials for genetic diagnosis [139]. Hopefully the emergence of more clinically validated RNA biomarker profiles will see the growth of disease-specific reference materials for RNA quantification.

6. Conclusion

The last decade has seen the emergence of high throughput PCR and sequencing platforms which will hopefully result in an expansion of RNA biomarker-based diagnostic tools. With the advances in sequencing technologies anticipated to bring about the realisation of the $1000 genome in 2012 [140], affordable individualised transcriptomic measurements are also on the horizon. However the majority of current applications of next generation instruments are focussed on achieving a better understanding of complex

diseases such as colorectal cancer (readers are referred to article by Murphy and Bustin in this issue of Methods) prior to the translation of biomarker panels into clinical practice; for example through single cell qPCR analysis of cell sub-types present in normal and diseased colon [141]. A combination of both genomic and transcriptomic biomarkers may be most useful for personalised cancer medicine, as although whole genome sequencing may reveal a range of genetic alterations present in diseased tissue, the causative mutations or abnormalities may be discerned by analysis of gene expression. Pilot studies are already evaluating this combined approach for appropriate selection of patients into clinical trials [142].

Whilst the diagnostic or prognostic validity of panels of transcriptional biomarkers requires testing through large scale prospective randomised trials before their inclusion alongside or as replacement for routine clinical tests [143], a pre-requisite of such trials is the technical validity of the obtained results. Benchmarking of the performance of next generation platforms and associated methodologies through the application of universal controls and reference materials will help to form the foundation for the translation of potential RNA biomarkers into authorised clinical assays for personalised medicine.

Acknowledgements

We would like to thank Dr. Ramnath Elaswarapu for his detailed assessment of the manuscript. This manuscript was funded by the UK National Measurement System.

References

[1] J.S. Ross, C. Hatzis, W.F. Symmans, L. Pusztai, G.N. Hortobagyi, Oncologist 13 (2008) 477–493.
[2] C. Herder, M. Karakas, W. Koenig, Clin. Pharmacol. Ther. 90 (2011) 52–66.
[3] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, N. Wolmark, N. Engl, J. Med. 351 (2004) 2817–2826.
[4] M. Cronin, C. Sangli, M.L. Liu, M. Pho, D. Dutta, A. Nguyen, J. Jeong, J. Wu, K.C. Langone, D. Watson, Clin. Chem. 53 (2007) 1084–1091.
[5] K.M. Clark-Langone, C. Sangli, J. Krishnakumar, D. Watson, BMC Cancer 10 (2010) 691.
[6] <www.oncotypedx.com>, accessed 08 November 2011.
[7] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Nat. Methods 5 (2008) 621–628.
[8] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, Genome Res. 18 (2008) 1509–1517.
[9] J. Liu, C. Hansen, S.R. Quake, Anal. Chem. 75 (2003) 4718–4723.
[10] D.A. Wheeler, M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.J. Chen, V. Makhijani, G.T. Roth, X. Gomes, K. Tartaro, F. Niazi, C.L. Turcotte, G.P. Irzyk, J.R. Lupski, C. Chinault, X.Z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D.M. Muzny, M. Margulies, G.M. Weinstock, R.A. Gibbs, J.M. Rothberg, Nature 452 (2008) 872–876.
[11] Z. Wang, M. Gerstein, M. Snyder, Nat. Rev. Genet. 10 (2009) 57–63.
[12] M.L. Metzker, Nat. Rev. Genet. 11 (2010) 31–46.
[13] N.C. Roy, E. Altermann, Z.A. Park, W.C. McNabb, Brief Funct. Genomics 10 (2011) 135–150.
[14] S.A. Bustin, J. Mol. Endocrinol. 29 (2002) 23–39.
[15] F. Lewis, N.J. Maughan, V. Smith, K. Hillan, P. Quirke, J. Pathol. 195 (2001) 66–71.
[16] N. Masuda, T. Ohnishi, S. Kawamoto, M. Monden, K. Okubo, Nucleic Acids Res. 27 (1999) 4436–4443.
[17] K. Bohmann, G. Hennig, U. Rogel, C. Poremba, B.M. Mueller, P. Fritz, S. Stoerkel, K.L. Schaefer, Clin. Chem. 55 (2009) 1719–1727.
[18] M. Cronin, M. Pho, D. Dutta, J.C. Stephans, S. Shak, M.C. Kiefer, J.M. Esteban, J.B. Baker, Am. J. Pathol. 164 (2004) 35–42.
[19] B. Sadikovic, P. Thorner, S. Chilton-Macneill, J.W. Martin, N.K. Cervigne, J. Squire, M. Zielenska, BMC Cancer 10 (2010) 202.
[20] S. Duenwald, M. Zhou, Y. Wang, S. Lejnine, A. Kulkarni, J. Graves, R. Smith, J. Castle, G. Tokiwa, B. Fine, H. Dai, T. Fare, M. Marton, J. Transl. Med. 7 (2009) 65.
[21] A.H. Beck, Z. Weng, D.M. Witten, S. Zhu, J.W. Foley, P. Lacroute, C.L. Smith, R. Tibshirani, M. van de Rijn, A. Sidow, R.B. West, PLoS ONE 5 (2010) e8768.
[22] J. Davoren, D. Vanek, R. Konjhodzic, J. Crews, E. Huffine, T.J. Parsons, Croat. Med. J. 48 (2007) 478–485.
[23] O.M. Loreille, T.M. Diegoli, J.A. Irwin, M.D. Coble, T.J. Parsons, Forensic Sci. Int. Genet. 1 (2007) 191–195.
[24] M. Salamon, N. Tuross, B. Arensburg, S. Weiner, Proc. Natl. Acad. Sci. USA 102 (2005) 13783–13788.
[25] C.A. Perez-Novo, C. Claeys, F. Speleman, P. Van Cauwenberge, C. Bachert, J. Vandesompele, Biotechniques 39 (2005). 52, 54, 56.
[26] R.V. Gutala, P.H. Reddy, J. Neurosci. Methods 132 (2004) 101–107.
[27] S.A. Bustin, T. Nolan, J. Biomol. Tech. 15 (2004) 155–166.
[28] E.M. Glare, M. Divjak, M.J. Bailey, E.H. Walters, Thorax 57 (2002) 765–770.
[29] E.M. Tunbridge, S.L. Eastwood, P.J. Harrison, Biol. Psychiatry 69 (2011) 173–179.
[30] B.C. Fox, A.S. Devonshire, M.E. Schutte, C.A. Foy, J. Minguez, S. Przyborski, D. Maltman, M. Bokhari, D. Marshall, Toxicol. In Vitro 24 (2010) 1962–1970.
[31] J. Kulka, A.M. Tokes, P. Kaposi-Novak, N. Udvarhelyi, A. Keller, Z. Schaff, Pathol. Oncol. Res. 12 (2006) 197–204.
[32] S.A. Bustin, J. Mol. Endocrinol. 25 (2000) 169–193.
[33] M.A. Valasek, J.J. Repa, Adv. Physiol. Educ. 29 (2005) 151–159.
[34] J. Chelly, J.P. Concordet, J.C. Kaplan, A. Kahn, Proc. Natl. Acad. Sci. USA 86 (1989) 2617–2621.
[35] Y. Sugiyama, B. Farrow, C. Murillo, J. Li, H. Watanabe, K. Sugiyama, B.M. Evers, Gastroenterology 128 (2005) 480–486.
[36] S. Bhat, N. Curach, T. Mostyn, G.S. Bains, K.R. Griffiths, K.R. Emslie, Anal. Chem. 82 (2010) 7185–7192.
[37] I. Blotta, F. Prestinaci, S. Mirante, A. Cantafora, Ann. Ist. Super. Sanita. 41 (2005) 119–123.
[38] M.J. Cavaluzzi, P.N. Borer, Nucleic Acids Res. 32 (2004) e13.
[39] K.A. Haque, R.M. Pfeiffer, M.B. Beerman, J.P. Struewing, S.J. Chanock, A.W. Bergen, BMC Biotechnol. 3 (2003) 20.
[40] S.A. Bustin, J.F. Beaulieu, J. Huggett, R. Jaggi, F.S. Kibenge, P.A. Olsvik, L.C. Penning, S. Toegel, BMC Mol. Biol. 11 (2010) 74.
[41] J.F. Huggett, T. Novak, J.A. Garson, C. Green, S.D. Morris-Jones, R.F. Miller, A. Zumla, BMC Res. Notes 1 (2008) 70.
[42] T. Nolan, R.E. Hands, W. Ogunkolade, S.A. Bustin, Anal. Biochem. 351 (2006) 308–310.
[43] Illumina TruSeq™ RNA Sample Preparation Guide Catalog # RS-930-20 01 Part # 15008136 Rev. A (November 2010).
[44] S. Weaver, S. Dube, A. Mir, J. Qin, G. Sun, R. Ramakrishnan, R.C. Jones, K.J. Livak, Methods 50 (2010) 271–276.
[45] S.L. Spurgeon, R.C. Jones, R. Ramakrishnan, PLoS ONE 3 (2008) e1662.
[46] T. Morrison, J. Hurley, J. Garcia, K. Yoder, A. Katz, D. Roberts, J. Cho, T. Kanigan, S.E. Ilyin, D. Horowitz, J.M. Dixon, C.J. Brenan, Nucleic Acids Res. 34 (2006) e123.
[47] <www.appliedbiosystems.com> accessed 24 October 2011.
[48] A. Keller, P. Leidinger, A. Bauer, A. Elsharawy, J. Haas, C. Backes, A. Wendschlag, N. Giese, C. Tjaden, K. Ott, J. Werner, T. Hackert, K. Ruprecht, H. Huwer, J. Huebers, G. Jacobs, P. Rosenstiel, H. Dommisch, A. Schaefer, J. Muller-Quernheim, B. Wullich, B. Keck, N. Graf, J. Reichrath, B. Vogel, A. Nebel, S.U. Jager, P. Staehler, I. Amarantos, V. Boisguerin, C. Staehler, M. Beier, M. Scheffler, M.W. Buchler, J. Wischhusen, S.F. Haeusler, J. Dietl, S. Hofmann, H.P. Lenhof, S. Schreiber, H.A. Katus, W. Rottbauer, B. Meder, J.D. Hoheisel, A. Franke, E. Meese, Nat. Methods 8 (2011) 841–843.
[49] <www.wafergen.com>, accessed 19 October 2011.
[50] B. Vogelstein, K.W. Kinzler, Proc. Natl. Acad. Sci. USA 96 (1999) 9236–9241.
[51] R. Sanders, J.F. Huggett, C.A. Bushell, S. Cowen, D.J. Scott, C.A. Foy, Anal. Chem. 83 (2011) 6474–6484.
[52] L. Warren, D. Bryder, I.L. Weissman, S.R. Quake, Proc. Natl. Acad. Sci. USA 103 (2006) 17807–17812.
[53] N.R. Beer, B.J. Hindson, E.K. Wheeler, S.B. Hall, K.A. Rose, I.M. Kennedy, B.W. Colston, Anal. Chem. 79 (2007) 8471–8475.
[54] <www.quantalife.com/product/ddpcr>, accessed 29 November 2011.
[55] Q. Zhong, S. Bhattacharya, S. Kotsopoulos, J. Olson, V. Taly, A.D. Griffiths, D.R. Link, J.W. Larson, Lab Chip 11 (2011) 2167–2174.
[56] A.S. Devonshire, R. Elaswarapu, C.A. Foy, BMC Genomics 11 (2010) 662.
[57] A. Stahlberg, M. Kubista, M. Pfaffl, Clin. Chem. 50 (2004) 1678–1680.
[58] J.P. Levesque-Sergerie, M. Duquette, C. Thibault, L. Delbecchi, N. Bissonnette, BMC Mol. Biol. 8 (2007) 93.
[59] BioMark™ Advanced Development Protocol Number 5: Single-Cell Gene Expression Protocol for the BioMark 48.48 Dynamic Array–Real-Time PCR Part Number 68000107.
[60] BioMark™ Advanced Development Protocol Number 8: Gene Expression Analysis Using Assays Designed with Probes from the Universal Probe Library (Roche Applied Sciences) Part Number 68000116.
[61] J.N. Orina, A.M. Calcagno, C.P. Wu, S. Varma, J. Shih, M. Lin, G. Eichler, J.N. Weinstein, Y. Pommier, S.V. Ambudkar, M.M. Gottesman, J.P. Gillet, Mol. Cancer Ther. 8 (2009) 2057–2066.
[62] A.S. Devonshire, R. Elaswarapu, C.A. Foy, BMC Genomics 12 (2011) 118.
[63] J.M. Dixon, M. Lubomirski, D. Amaratunga, T.B. Morrison, C.J. Brenan, S.E. Ilyin, Biotechniques 46 (2009) ii–viii.
[64] Y.W. Asmann, E.W. Klee, E.A. Thompson, E.A. Perez, S. Middha, A.L. Oberg, T.M. Therneau, D.I. Smith, G.A. Poland, E.D. Wieben, J.P. Kocher, BMC Genomics 10 (2009) 531.
[65] N. Cloonan, A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, D.F. Taylor, A.L. Steptoe, S. Wani, G. Bethel, A.J. Robertson, A.C. Perkins, S.J. Bruce, C.C. Lee, S.S. Ranade, H.E. Peckham, J.M. Manning, K.J. McKernan, S.M. Grimmond, Nat. Methods 5 (2008) 613–619.
[66] J.Z. Levin, M.F. Berger, X. Adiconis, P. Rogov, A. Melnikov, T. Fennell, C. Nusbaum, L.A. Garraway, A. Gnirke, Genome Biol. 10 (2009) R115.

[67] M. Sultan, M.H. Schulz, H. Richard, A. Magen, A. Klingenhoff, M. Scherf, M. Seifert, T. Borodina, A. Soldatov, D. Parkhomchuk, D. Schmidt, S. O'Keeffe, S. Haas, M. Vingron, H. Lehrach, M.L. Yaspo, Science 321 (2008) 956–960.
[68] V. Costa, C. Angelini, L. D'Apice, M. Mutarelli, A. Casamassimi, L. Sommese, M.A. Gallo, M. Aprile, R. Esposito, L. Leone, A. Donizetti, S. Crispi, M. Rienzo, B. Sarubbi, R. Calabro, M. Picardi, P. Salvatore, T. Infante, P. De Berardinis, C. Napoli, A. Ciccodicola, PLoS ONE 6 (2011) e18493.
[69] T. Raz, P. Kapranov, D. Lipson, S. Letovsky, P.M. Milos, J.F. Thompson, PLoS ONE 6 (2011) e19287.
[70] R.D. Thiagarajan, N. Cloonan, B.B. Gardiner, T.R. Mercer, G. Kolle, E. Nourbakhsh, S. Wani, D. Tang, K. Krishnan, K.M. Georgas, B.A. Rumballe, H.S. Chiu, J.A. Steen, J.S. Mattick, M.H. Little, S.M. Grimmond, BMC Genomics 12 (2011) 441.
[71] E. Valen, G. Pascarella, A. Chalk, N. Maeda, M. Kojima, C. Kawazu, M. Murata, H. Nishiyori, D. Lazarevic, D. Motti, T.T. Marstrand, M.H. Tang, X. Zhao, A. Krogh, O. Winther, T. Arakawa, J. Kawai, C. Wells, C. Daub, M. Harbers, Y. Hayashizaki, S. Gustincich, A. Sandelin, P. Carninci, Genome Res. 19 (2009) 255–265.
[72] M. Harbers, P. Carninci, Nat. Methods 2 (2005) 495–502.
[73] M. Kanamori-Katayama, M. Itoh, H. Kawaji, T. Lassmann, S. Katayama, M. Kojima, N. Bertin, A. Kaiho, N. Ninomiya, C.O. Daub, P. Carninci, A.R. Forrest, Y. Hayashizaki, Genome Res. 21 (2011) 1150–1159.
[74] T.R. Mercer, D.J. Gerhardt, M.E. Dinger, J. Crawford, C. Trapnell, J.A. Jeddeloh, J.S. Mattick, J.L. Rinn, Nat. Biotechnol. 30 (2012) 99–104.
[75] V. Costa, C. Angelini, I. De Feis, A. Ciccodicola, J. Biomed. Biotechnol. 2010 (2010) 853916.
[76] D.M. Carraro, E.N. Ferreira, G. de Campos Molina, R.D. Puga, E.F. Abrantes, A.P. Trape, B.L. Eckhardt, D.N. Nunes, M.M. Brentani, W. Arap, R. Pasqualini, H. Brentani, E. Dias-Neto, R.R. Brentani, PLoS ONE 6 (2011) e21022.
[77] U.C. Klostermeier, M. Barann, M. Wittig, R. Hasler, A. Franke, O. Gavrilova, B. Kreck, C. Sina, M.B. Schilhabel, S. Schreiber, P. Rosenstiel, BMC Genomics 12 (2011) 305.
[78] Applied Biosystems SOLiD™ Total RNA-Seq Kit Protocol Publication Part Number 4452437 Rev. B (July 2011).
[79] Applied Biosystems SOLiD™ SAGE™ Kit with Barcoding Adaptor Module Guide Part no. 4456596 Rev. date 13 April 2011.
[80] Roche cDNA Rapid Library Preparation Method Manual (GS Junior Titanium Series) May 2010 (Rev. June 2010).
[81] Illumina Preparing Samples for Digital Gene Expression-Tag Profiling with NlaIII Part # 1004240 Rev. A (March 2008).
[82] Illumina Preparing Samples for Digital Gene Expression-Tag Profiling with DpnII Part # 1004241 Rev. A March 2008.
[83] J.Z. Levin, M. Yassour, X. Adiconis, C. Nusbaum, D.A. Thompson, N. Friedman, A. Gnirke, A. Regev, Nat. Methods 7 (2010) 709–715.
[84] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, A. Soldatov, Nucleic Acids Res. 37 (2009) e123.
[85] L. Wang, Y. Si, L.K. Dedow, Y. Shao, P. Liu, T.P. Brutnell, PLoS ONE 6 (2011) e26426.
[86] S.J. Ahn, J. Costa, J.R. Emanuel, Nucleic Acids Res. 24 (1996) 2623–2625.
[87] M.A. Quail, I. Kozarewa, F. Smith, A. Scally, P.J. Stephens, R. Durbin, H. Swerdlow, D.J. Turner, Nat. Methods 5 (2008) 1005–1010.
[88] M. Meyer, A.W. Briggs, T. Maricic, B. Hober, B. Hoffner, J. Krause, A. Weihmann, S. Paabo, M. Hofreiter, Nucleic Acids Res. 36 (2008) e5.
[89] B. Buehler, H.H. Hogrefe, G. Scott, H. Ravi, C. Pabon-Pena, S. O'Brien, R. Formosa, S. Happe, Methods 50 (2010) S15–18.
[90] D. Pushkarev, N.F. Neff, S.R. Quake, Nat. Biotechnol. 27 (2009) 847–850.
[91] Q. Huang, B. Lin, H. Liu, X. Ma, F. Mo, W. Yu, L. Li, H. Li, T. Tian, D. Wu, F. Shen, J. Xing, Z.N. Chen, PLoS ONE 6 (2011) e26168.
[92] N. Palanisamy, B. Ateeq, S. Kalyana-Sundaram, D. Pflueger, K. Ramnarayanan, S. Shankar, B. Han, Q. Cao, X. Cao, K. Suleman, C. Kumar-Sinha, S.M. Dhanasekaran, Y.B. Chen, R. Esgueva, S. Banerjee, C.J. LaFargue, J. Siddiqui, F. Demichelis, P. Moeller, T.A. Bismar, R. Kuefer, D.R. Fullen, T.M. Johnson, J.K. Greenson, T.J. Giordano, P. Tan, S.A. Tomlins, S. Varambally, M.A. Rubin, C.A. Maher, A.M. Chinnaiyan, Nat. Med. 16 (2010) 793–798.
[93] <www.my454.com/products/gs-flx-system/index.asp> accessed 21 November 2011.
[94] <www.illumina.com/support/sequencing/sequencing_instruments/genome_analyzer_iix/perf_specs.ilmn>, accessed 21 November 2011.
[95] <www.illumina.com/systems/hiseq_2000/performance_specifications.ilmn> accessed 21 November 2011.
[96] Applied BioSystems Specification Sheet 5500 Series Genetic Analysis Systems CO18235 0511 (2011).
[97] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Nature 437 (2005) 376–380.
[98] S. Balzer, K. Malde, I. Jonassen, Bioinformatics 27 (2011) i304–309.
[99] K.C. Ha, E. Lalonde, L. Li, L. Cavallone, R. Natrajan, M.B. Lambros, C. Mitsopoulos, J. Hakas, I. Kozarewa, K. Fenwick, C.J. Lord, A. Ashworth, A. Vincent-Salomon, M. Basik, J.S. Reis-Filho, J. Majewski, W.D. Foulkes, BMC Med. Genomics 4 (2011) 75.
[100] J.C. Dohm, C. Lottaz, T. Borodina, H. Himmelbauer, Nucleic Acids Res. 36 (2008) e105.
[101] Nextera DNA Sample Preparation Kits Illumina Publication No. 770–2011-021 (07 October 2011).
[102] E.E. Schadt, S. Turner, A. Kasarskis, Hum. Mol. Genet. 19 (2010) R227–240.
[103] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Science 323 (2009) 133–138.
[104] <www.pacificbiosciences.com> accessed 21 November 2011.
[105] <www.helicosbio.com/Technology/TrueSingleMoleculeSequencing/tSMStradePerformance/tabid/151/Default.aspx> accessed 21 November 2011.
[106] D. Lipson, T. Raz, A. Kieu, D.R. Jones, E. Giladi, E. Thayer, J.F. Thompson, S. Letovsky, P. Milos, M. Causey, Nat. Biotechnol. 27 (2009) 652–658.
[107] F. Ozsolak, A.R. Platt, D.R. Jones, J.G. Reifenberger, L.E. Sass, P. McInerney, J.F. Thompson, J. Bowers, M. Jarosz, P.M. Milos, Nature 461 (2009) 814–818.
[108] F. Ozsolak, D.T. Ting, B.S. Wittner, B.W. Brannigan, S. Paul, N. Bardeesy, S. Ramaswamy, P.M. Milos, D.A. Haber, Nat. Methods 7 (2010) 619–621.
[109] L.T. Sam, D. Lipson, T. Raz, X. Cao, J. Thompson, P.M. Milos, D. Robinson, A.M. Chinnaiyan, C. Kumar-Sinha, C.A. Maher, PLoS ONE 6 (2011) e17305.
[110] F. Ozsolak, P.M. Milos, Nat. Rev. Genet. 12 (2011) 87–98.
[111] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo, Nature 475 (2011) 348–352.
[112] IonTorrent Technical Note No.: CO23050 (2011).
[113] <www.iontorrent.com/technology-how-does-it-perform/> accessed 21 November 2011.
[114] <www.technologyreview.com/biomedicine/39458/>, accessed 31 May 2012.
[115] The Ion Proton™ System Specification Sheet Publication No. CO25043 0412.
[116] U. Vogel, R. Szczepanowski, H. Claus, S. Junemann, K. Prior, D. Harmsen, J. Clin. Microbiol. 50 (2012) 1889–1894.
[117] S. Lee, C.H. Seo, B. Lim, J.O. Yang, J. Oh, M. Kim, B. Lee, C. Kang, Nucleic Acids Res. 39 (2011) e9.
[118] D. Tian, Q. Wang, P. Zhang, H. Araki, S. Yang, M. Kreitman, T. Nagylaki, R. Hudson, J. Bergelson, J.Q. Chen, Nature 455 (2008) 105–108.
[119] S. Kumar, M.L. Blaxter, BMC Genomics 11 (2010) 571.
[120] D.A. Skelly, M. Johansson, J. Madeoy, J. Wakefield, J.M. Akey, Genome Res. 21 (2011) 1728–1737.
[121] J.F. Degner, J.C. Marioni, A.A. Pai, J.K. Pickrell, E. Nkadori, Y. Gilad, J.K. Pritchard, Bioinformatics 25 (2009) 3207–3212.
[122] M.B. Clark, P.P. Amaral, F.J. Schlesinger, M.E. Dinger, R.J. Taft, J.L. Rinn, C.P. Ponting, P.F. Stadler, K.V. Morris, A. Morillon, J.S. Rozowsky, M.B. Gerstein, C. Wahlestedt, Y. Hayashizaki, P. Carninci, T.R. Gingeras, J.S. Mattick, PLoS Biol. 9 (2011) e1000625. discussion e1001102.
[123] J. Feng, W. Li, T. Jiang, J. Comput. Biol. 18 (2011) 305–321.
[124] L. Wang, X. Wang, Y. Liang, X. Zhang, Biochem. Biophys. Res. Commun. 409 (2011) 299–303.
[125] G.R. Grant, M.H. Farkas, A.D. Pizarro, N.F. Lahens, J. Schug, B.P. Brunk, C.J. Stoeckert, J.B. Hogenesch, E.A. Pierce, Bioinformatics 27 (2011) 2518–2528.
[126] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V.B. Bajic, S.E. Brenner, S. Batalov, A.R. Forrest, M. Zavolan, M.J. Davis, L.G. Wilming, V. Aidinis, J.E. Allen, A. Ambesi-Impiombato, R. Apweiler, R.N. Aturaliya, T.L. Bailey, M. Bansal, L. Baxter, K.W. Beisel, T. Bersano, H. Bono, A.M. Chalk, K.P. Chiu, V. Choudhary, A. Christoffels, D.R. Clutterbuck, M.L. Crowe, E. Dalla, B.P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C.F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T.R. Gingeras, T. Gojobori, R.E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T.K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S.P. Krishnan, A. Kruger, S.K. Kummerfeld, I.V. Kurochkin, L.F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, F. Nori, O. Ohara, Y. Okazaki, V. Orlando, K.C. Pang, W.J. Pavan, G. Pavesi, G. Pesole, N. Petrovsky, S. Piazza, J. Reed, J.F. Reid, B.Z. Ring, M. Ringwald, B. Rost, Y. Ruan, S.L. Salzberg, A. Sandelin, C. Schneider, C. Schonbach, K. Sekiguchi, C.A. Semple, S. Seno, L. Sessa, Y. Sheng, Y. Shibata, H. Shimada, K. Shimada, D. Silva, B. Sinclair, S. Sperling, E. Stupka, K. Sugiura, R. Sultana, Y. Takenaka, K. Taki, K. Tammoja, S.L. Tan, S. Tang, M.S. Taylor, J. Tegner, S.A. Teichmann, H.R. Ueda, E. van Nimwegen, R. Verardo, C.L. Wei, K. Yagi, H. Yamanishi, E. Zabarovsky, S. Zhu, A. Zimmer, W. Hide, C. Bult, S.M. Grimmond, R.D. Teasdale, E.T. Liu, V. Brusic, J. Quackenbush, C. Wahlestedt, J.S. Mattick, D.A. Hume, C. Kai, D. Sasaki, Y. Tomaru, S. Fukuda, M. Kanamori-Katayama, M. Suzuki, J. Aoki, T. Arakawa, J. Iida, K. Imamura, M. Itoh, T. Kato, H. Kawaji, N. Kawagashira, T. Kawashima, M. Kojima, S. Kondo,

H. Konno, K. Nakano, N. Ninomiya, T. Nishio, M. Okada, C. Plessy, K. Shibata, T. Shiraki, S. Suzuki, M. Tagami, K. Waki, A. Watahiki, Y. Okamura-Oho, H. Suzuki, J. Kawai, Y. Hayashizaki, Science 309 (2005) 1559–1563.
[127] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Nat. Biotechnol. 28 (2010) 511–515.
[128] A. Oshlack, M.J. Wakefield, Biol. Direct 4 (2009) 14.
[129] A. Oshlack, M.D. Robinson, M.D. Young, Genome Biol. 11 (2010) 220.
[130] M.J. Holden, R.M. Madej, P. Minor, L.V. Kalman, Expert Rev. Mol. Diagn. 11 (2011) 741–755.
[131] M. Cronin, K. Ghosh, F. Sistare, J. Quackenbush, V. Vilker, C. O'Connell, Clin. Chem. 50 (2004) 1464–1471.
[132] N.A. Twine, K. Janitz, M.R. Wilkins, M. Janitz, PLoS ONE 6 (2011) e16266.
[133] A. Agarwal, D. Koppstein, J. Rozowsky, A. Sboner, L. Habegger, L.W. Hillier, R. Sasidharan, V. Reinke, R.H. Waterston, M. Gerstein, BMC Genomics 11 (2010) 383.
[134] L.M. McIntyre, K.K. Lopiano, A.M. Morse, V. Amin, A.L. Oberg, L.J. Young, S.V. Nuzhdin, BMC Genomics 12 (2011) 293.
[135] L. Shi, L.H. Reid, W.D. Jones, R. Shippy, J.A. Warrington, S.C. Baker, P.J. Collins, F. de Longueville, E.S. Kawasaki, K.Y. Lee, Y. Luo, Y.A. Sun, J.C. Willey, R.A. Setterquist, G.M. Fischer, W. Tong, Y.P. Dragan, D.J. Dix, F.W. Frueh, F.M. Goodsaid, D. Herman, R.V. Jensen, C.D. Johnson, E.K. Lobenhofer, R.K. Puri, U. Schrf, J. Thierry-Mieg, C. Wang, M. Wilson, P.K. Wolber, L. Zhang, S. Amur, W. Bao, C.C. Barbacioru, A.B. Lucas, V. Bertholet, C. Boysen, B. Bromley, D. Brown, A. Brunner, R. Canales, X.M. Cao, T.A. Cebula, J.J. Chen, J. Cheng, T.M. Chu, E. Chudin, J. Corson, J.C. Corton, L.J. Croner, C. Davies, T.S. Davison, G. Delenstarr, X. Deng, D. Dorris, A.C. Eklund, X.H. Fan, H. Fang, S. Fulmer-Smentek, J.C. Fuscoe, K. Gallagher, W. Ge, L. Guo, X. Guo, J. Hager, P.K. Haje, J. Han, T. Han, H.C. Harbottle, S.C. Harris, E. Hatchwell, C.A. Hauser, S. Hester, H. Hong, P. Hurban, S.A. Jackson, H. Ji, C.R. Knight, W.P. Kuo, J.E. LeClerc, S. Levy, Q.Z. Li, C. Liu, Y. Liu, M.J. Lombardi, Y. Ma, S.R. Magnuson, B. Maqsodi, T. McDaniel, N. Mei, O. Myklebost, B. Ning, N. Novoradovskaya, M.S. Orr, T.W. Osborn, A. Papallo, T.A. Patterson, R.G. Perkins, E.H. Peters, R. Peterson, K.L. Philips, P.S. Pine, L. Pusztai, F. Qian, H. Ren, M. Rosen, B.A. Rosenzweig, R.R. Samaha, M. Schena, G.P. Schroth, S. Shchegrova, D.D. Smith, F. Staedtler, Z. Su, H. Sun, Z. Szallasi, Z. Tezak, D. Thierry-Mieg, K.L. Thompson, I. Tikhonova, Y. Turpaz, B. Vallanat, C. Van, S.J. Walker, S.J. Wang, Y. Wang, R. Wolfinger, A. Wong, J. Wu, C. Xiao, Q. Xie, J. Xu, W. Yang, S. Zhong, Y. Zong, W. Slikker Jr., Nat. Biotechnol. 24 (2006) 1151–1161.
[136] S.C. Baker, S.R. Bauer, R.P. Beyer, J.D. Brenton, B. Bromley, J. Burrill, H. Causton, M.P. Conley, R. Elespuru, M. Fero, C. Foy, J. Fuscoe, X. Gao, D.L. Gerhold, P. Gilles, F. Goodsaid, X. Guo, J. Hackett, R.D. Hockett, P. Ikonomi, R.A. Irizarry, E.S. Kawasaki, T. Kaysser-Kranich, K. Kerr, G. Kiser, W.H. Koch, K.Y. Lee, C. Liu, Z.L. Liu, A. Lucas, C.F. Manohar, G. Miyada, Z. Modrusan, H. Parkes, R.K. Puri, L. Reid, T.B. Ryder, M. Salit, R.R. Samaha, U. Scherf, T.J. Sendera, R.A. Setterquist, L. Shi, R. Shippy, J.V. Soriano, E.A. Wagar, J.A. Warrington, M. Williams, F. Wilmer, M. Wilson, P.K. Wolber, X. Wu, R. Zadro, Nat. Methods 2 (2005) 731–734.
[137] L. Jiang, F. Schlesinger, C.A. Davis, Y. Zhang, R. Li, M. Salit, T.R. Gingeras, B. Oliver, Genome Res. 21 (2011) 1543–1551.
[138] <http://wwwn.cdc.gov/dls/genetics/rmmaterials/default.aspx>, accessed 22 December 2011.
[139] G. Sinha, Nat. Biotechnol. 29 (2011) 960.
[140] P. Dalerba, T. Kalisky, D. Sahoo, P.S. Rajendran, M.E. Rothenberg, A.A. Leyrat, S. Sim, J. Okamoto, D.M. Johnston, D. Qian, M. Zabala, J. Bueno, N.F. Neff, J. Wang, A.A. Shelton, B. Visser, S. Hisamori, Y. Shimono, M. van de Wetering, H. Clevers, M.F. Clarke, S.R. Quake, Nat. Biotechnol. 29 (2011) 1120–1127.
[141] S. Roychowdhury, M.K. Iyer, D.R. Robinson, R.J. Lonigro, Y.M. Wu, X. Cao, S. Kalyana-Sundaram, L. Sam, O.A. Balbin, M.J. Quist, T. Barrette, J. Everett, J. Siddiqui, L.P. Kunju, N. Navone, J.C. Araujo, P. Troncoso, C.J. Logothetis, J.W. Innis, D.C. Smith, C.D. Lao, S.Y. Kim, J.S. Roberts, S.B. Gruber, K.J. Pienta, M. Talpaz, A.M. Chinnaiyan, Sci. Transl. Med. 3 (2011) 111–121.
[142] C. Wright, H. Burton, A. Hall, S. Moorthie, A. Pokorska-Bocci, G. Sagoo, S. Simon Sanderson, R. Skinner, Next steps in the sequence. The implications of whole genome sequencing for health in the UK (PHG Foundation report) (2011).