EXPERIMENTAL CELL RESEARCH 316 (2010) 1339–1343

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/yexcr

Review

Recent advances in DNA sequencing methods – general principles of sample preparation

Sten Linnarsson

Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 1, SE-171 77 Stockholm, Sweden

ARTICLE INFORMATION

Article Chronology:
Received 12 February 2010
Accepted 28 February 2010
Available online 6 March 2010

Keywords:
DNA sequencing
Sample preparation
RNA-Seq

ABSTRACT

DNA sequencing has revolutionized biomedicine, and progress in the field has been unrelenting since it was invented over 30 years ago. The complete DNA sequence of the human genome was obtained as the culmination of a decade of work by a large number of scientists. Less than ten years later, so-called 'next-generation' instruments now make it possible for a single lab to produce the same amount of data in a week. But while the instruments are increasingly automated, upstream sample processing remains a challenge. Here I review the current state of the art in preparing genomic and RNA samples for high throughput sequencing.

© 2010 Elsevier Inc. All rights reserved.

Contents

Introduction . . . 1339
Requirements for a sequencing-ready library . . . 1340
General sample preparation procedures . . . 1340
Fragmentation . . . 1340
Adapter ligation . . . 1341
Size selection . . . 1341
Amplification . . . 1341
The end-game: quality control and quantification . . . 1341
Preparing RNA for sequencing. . . . 1342
Discussion . . . 1342
Appendix A. Supplementary data . . . 1343
References . . . 1343

Introduction

Ever since DNA sequencing was invented in the mid-seventies by Fredrick Sanger (the chain-terminating method [1]) and by Allan Maxam and Walter Gilbert (the chemical method [2]), sequencing has been a solid foundation for research in all branches of biology and medicine. In recent years, so-called 'next-generation' sequencing instruments have been developed, which are in general based

E-mail address: sten.linnarsson@ki.se.

0014-4827/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yexcr.2010.02.036

on a clonal amplification method to directly generate clusters of DNA templates at high densities on a glass slide. Templates are then sequenced by a stepwise incorporation of nucleotides (Illumina Genome Analyzer, 454 Genome Sequencer) or short oligonucleotides (Applied Biosystems SOLiD). Through impressive incremental improvements the throghput of these instruments has increased to currently about 200 gigabases per week per instrument (Illumina HiSeq 2000), while costs have dropped to where a human genome can be sequenced at 30-fold coverage for less than $10,000. These developments have recently been the subject of several excellent reviews [3–5].

However, while the actual sequencing has been streamlined and automated, the upstream sample preparation procedure remains an area of active and fertile research. For every application of DNA sequencing, such as metagenomics, transcriptome sequencing (RNA-Seq), genome sequencing or resequencing, chromatin immunoprecipitation sequencing (ChIP-Seq) and so on, there is the need for a specific protocol to convert the source material into a standard DNA library suitable for loading onto the sequencing instrument.

Here I will review recent developments in sample preparation with a view to providing some useful advice to the researcher who wishes to convert any kind of raw nucleic acid into a sequencing-ready library. The focus will be on sample preparation for the commercial platforms already on the market, in particular for the 454 Genome Sequencer FLX (henceforth termed the 'FLX'), the Illumina Genome Analyzer IIx ('GA', used loosely to include the recent HiSeq 2000) and the Applied Biosystems SOLiD 3 Plus ('SOLiD'), although many general principles will apply also to emerging and future systems. Enrichment methods, used to selectively enrich desired target regions in a genome, have been reviewed elsewhere [3,6] and will not be covered here. Finally, I will briefly review the way these general principles apply to one specific objective, sequencing RNA. In the accompanying **Supplementary Methods**, a detailed best-practices protocol is provided, as currently implemented in our lab.

Requirements for a sequencing-ready library

For each of the current platforms, the input is a double-stranded DNA library consisting of short fragments flanked by adapters of known (and platform-specific) sequence. Given the great differences between the GA, FLX and SOLiD platforms in both how they generate clonally amplified templates on the flowcell surface, and how these clones are interrogated to reveal their sequence, the input sample requirements are remarkably similar (Table 1). The principal difference is that FLX and SOLiD clonally amplify the templates by emulsion PCR [7], whereas the GA uses surface PCR, also known as bridge amplification [8]. The SOLiD system uses an emulsion with smaller droplets compared with FLX, and works best with shorter fragments. In contrast, the GA accepts a greater range of fragment lengths, but the yield drops as fragments get longer because longer fragments result in larger surface clusters that must be spaced less densely.

Table 1

	GA	FLX	SOLiD
Concentration (nM)	2-10	1	1
Volume (μL)	2	1.2	5-10
Fragment length (bp)	100-600	200-600	150-200

It's instructive to consider how much sequence information is contained in a typical sequencing library. As an example, 2 μL of 2 nM double-stranded DNA with an insert length of 300 bp contains 2.4 billion molecules weighing less than one nanogram. The 720 gigabases of sequence contained in these molecules are equivalent to about one hundred human diploid genomes. This is more than enough complexity for current instruments, but not by a great margin; the most recent GA instruments can generate 200 gigabases per run.

Thus we can see that the minimum unamplified starting material must be about one nanogram. For genome sequencing, at least a hundred diploid cells would be required if there were no losses in sample preparation. However, if parts of the genome are targeted using enrichment methods, then proportionately larger amounts of starting material will be required. Similarly, since mRNA comprises only about 300,000 molecules per mammalian cell, each about 2 kb long, more than a thousand cells would be required even if mRNA could be converted to a sequence-ready library without losses.

At present, therefore, all sample preparation protocols for next-generation sequencing require either a large amount of starting material (which blurs distinctions between individual cells), or some form of amplification (which introduces inevitable bias), or both. Recent developments in sample preparation have sought to reduce the impact of both of these requirements.

General sample preparation procedures

In the general case, common to almost all sample preparation protocols for next-generation sequencing, the starting material is double-stranded DNA in the form of e.g. isolated genomic DNA, reverse-transcribed cDNA, or immunoprecipitated chromatin. To convert this into a sequenceable library, the source DNA must be fragmented, polished, size selected, adapter ligated, purified and quantified.

In most cases there will be no need for specialized equipment. However, great care should be taken to avoid cross-contamination, as this will be very difficult to detect. In addition, whenever working with low concentration of nucleic acid, it is necessary to use low-adsorbing plasticware. We and others [9] have found that standard polypropylene microfuge tubes can adsorb as much as 100 ng DNA, which can lead to complete loss of a sample. We recommend using polyallomer tubes (Beckman Coulter; 'non-stick' tubes from other manufacturers may be equivalent) and to add some detergent in every reaction step (e.g. 0.02% Tween-20) to reduce adsorption to tube walls.

Fragmentation

Fragmentation is typically achieved by mechanical force. The official protocols provided by the manufacturers call for fragmentation by nebulization using a disposable nabulizer driven by pressurized air (GA, FLX), or by the Covaris AFA ultrasound device (SOLiD). Other shearing instruments include the Hydroshear (GeneMachines) and the BioRuptor (Diagenode).

Enzymatic digestion is an effective alternative. We have successfully used highly diluted DNaseI in the presence of Mn^{2+}, which helps induce double-strand breaks [10]. Recently, two commercial enzymatic fragmentation kits were introduced with performance claimed to be similar to the physical methods. Fragmentase (New England Biolabs) is based on a *V. vulnificus* nuclease that generates random nicks, blended with a modified T7 endonuclease that recognizes the nicks and cleaves the opposite strand. In our hands, it performs similar to the DNaseI/Mn^{2+} combination (Supplementary Methods). In contrast, Nextera (Epicentre), is based on random transposon insertion. It has the distinct advantage of introducing adapter sequences simultaneously with fragmentation, so that the library is obtained directly in finished form. However, insertion bias is a concern that remains to be addressed.

In any case, while the choice of fragmentation method has not been considered a major source of bias, it can significantly affect the recovery of desired fragments and hence the amount of starting material required. For example, the Covaris instrument generally produces narrower fragment distributions than nebulization, leading to more than four-fold better recovery [11]. We have found DNaseI and Fragmentase to be intermediate: the fragment length distribution is not as narrow as that obtained by the Covaris, but in contrast to nebulization very little material is lost in the process.

Finally, the physical methods are necessarily more costly upfront, may be prone to cross-contamination and may often be difficult to scale to large numbers of samples. The simplicity of enzymatic digestion may be attractive to the small lab, or when preparing large numbers of samples in parallel, but batch-to-batch reproducibility is a concern.

Adapter ligation

After the fragmented library is repaired (that is, protruding 3' and 5' ends are removed or filled in), adapters must be ligated. All current systems require each fragment to have distinct upstream and downstream adapters, which we may call A and B. The SOLiD and FLX systems ligate double-stranded A and B adapters and then selectively remove A-A and B-B fragments, which leads to a 50% loss of material. The GA uses a single Y-shaped adapter which ensures that every fragment gets A-B or B-A adapter sequences.

In all cases, there is a risk that two fragments are accidentally joined together, forming a chimera which generates misleading paired-end reads. To avoid this, a double gel selection can be used [10]. If, say, 200 bp fragments are selected before ligation of 100 bp adapters, then 300 bp fragments can be selected after ligation. Any chimeras must be about 500 bp and can be easily avoided. Note that even A-tailed fragments can ligate to form chimeras at surprisingly high frequencies [11].

Size selection

Size selection is generally performed by agarose gel electrophoresis. However, the standard protocols for gel extraction include a heating step that may denature some AT-rich sequences and render them unligatable. To avoid the resulting GC-bias, it is advisable to avoid heating. For single selections, we routinely use 2% SizeSelect E-gels (Invitrogen; the Lonza FlashGel system is similar but is currently only available with 1.2% agarose), where the band of interest can be simply pipetted out of the gel. In addition to being quicker to use than standard agarose gels, these systems avoid ethidium bromide and UV light, and are thus not mutageneic either to the sample or to the researcher. In contrast, even a 30 second UV exposure of a standard ethidium bromide gel reduces cloning efficiency by 80% (Invitrogen SYBR Safe product documentation). For double selection, a regular agarose gel is preferrable, as a greater amount of DNA can be recovered and the size can be chosen more precisely. The excised agarose gel plug is dissolved by vigorous agitation at room temperature, not by heating.

Amplification

The single most problematic step of most sample preparation procedures is the amplification, which results in loss of specific regions of the template DNA as other regions are more efficiently amplified [10–12]. The standard FLX sample preparation protocol is PCR-free, but requires 2-5 μg starting material (i.e. millions of cells). Both the GA and SOLiD use PCR, but mainly as a means to enrich their libraries for properly adaptered fragments. If an accurate quantification method is used (see below), the amplification step can be avoided, at least when starting with several hundred nanograms of DNA, or micrograms of RNA [13]. For most applications, the result will be greatly improved coverage and quantitative accuracy.

However, as noted above, amplification is absolutely necessary whenever the starting material is less than a hundred (DNA) or a thousand (RNA) mammalian cells, because current sequencers require several billion input molecules. Although initial efforts at single-cell genome [14,15] and transcriptome [16] sequencing are promising, much work remains to develop scalable and quantitatively accurate methods that minimize amplification bias.

The end-game: quality control and quantification

An often overlooked but crucial sample preparation step is the final quality control and quantification. Given the high cost of sequencing, it is prudent to subclone and Sanger sequence a small number of clones to verify the integrity of the library. As a side effect, this permits a more accurate estimate of the average fragment length, which is necessary to convert mass concentration measurements into molar concentrations.

All the sequencing instruments discussed here are highly sensitive to the molar concentration of the sequencing library. Too low, and much of the capacity of the instruments is wasted, leading to higher cost per base sequenced. Too high, and many if not most detection sites will overlap or will be non-clonal. The optimum concentration may only be reached with very accurate quantification. Since yield drops rapidly when the optimum is exceeded, it is prudent to err on the low side.

DNA concentration has traditionally been estimated by measuring the absorbance at 260 nm. However, this is not enough to get an accurate estimate, since contamination by proteins, phenol and other organic compounds will contribute an unknown fraction to the absorbance. The practice of using an OD_{260}/OD_{280} ratio of

2.00 as an indication of sample purity is nearly useless, as even a 25% protein contamination only reduces this ratio to 1.97. Furthermore, leftover primers, free nucleotides (e.g. from upstream enzymatic reactions) and improperly adaptered fragments are all indistinguishable from the desired, productive fragments.

Instead, an intercalating fluorescent dye such as SYBR or PicoGreen can be used to measure double-stranded DNA specifically. Inexpensive and convenient benchtop fluorometers are available together with streamlined kits and protocols (e.g. Invitrogen's Qubit). This method is very robust to contaminating proteins, RNA, salts and organic compunds, and is recommended as a general-purpose initial quantification of sequencing libraries [11].

Fluorimetry gives the mass concentration (ng/μL), which can be converted to molar concentration (nM) if the average fragment length is known. Fragment lengths can be estimated from a diagnostic agarose gel, using capillary electrophoresis (e.g. Agilent BioAnalyzer) or from a set of cloned and Sanger-sequenced fragments. We have found that Qubit quantification together with Sanger-sequencing gives an adequate estimate of sample molar concentration for most routine purposes, at least for libraries prepared using identical procedures.

However, in order to achieve maximum throughput it is necessary to estimate specifically the fraction of fragments that are amplifiable in the emulsion or bridge PCR. Only this fraction will generate visible clusters, and thus only this fraction contributes to template density on the sequencing instrument. A simple quantitative PCR reaction can be used for this purpose, with primers identical to those used by the sequencing instrument [17]. If a sample previously determined to be optimal is used as reference, the highest yields can be consistently achieved [11]. An even more accurate quantification can be obtained by digital single-molecule PCR [18], but this requires a specialized instrument (e.g. the Fluidigm BioMark).

Preparing RNA for sequencing

We will now consider how the general principles above have been applied to the specific case of sequencing RNA. Ultimately, the goal of RNA sequencing is to determine the complete RNA content of a sample with full-length coverage and quantitative accuracy. Since transcription occurs in cells, the single cell is the relevant unit of analysis, although sometimes it is even desirable to isolate subcellular compartments such as synapses. A single mammalian cell may contain about 10 pg of RNA, or about 10 million molecules, of which 97% represent ribosomal RNA. Most of the time, we are interested in the non-ribosomal fraction, which includes mRNA, micro-RNA, tRNA and other non-coding RNAs. Sometimes it may be desirable to suppress other highly expressed transcripts, e.g. α- and β-globin in blood.

A straightforward approach is to isolate the RNA fraction of interest (or selectively remove the offending fractions), convert the remainder into double-stranded cDNA, and then use the general procedure above to generate a sequencing library. This was the approach taken in some of the initial RNA-Seq experiments [19,20], and is still the commercially supported protocol for the GA. However, strand information is not preserved, and thus one cannot accurately quantify the approximately 3000 genes that overlap on opposite strands in the human genome [21].

Alternatively, adapter sequences can be introduced directionally, using the template-switching activity of reverse transcriptase [22], by directional linker ligation [23,24], or by selectively labelling one cDNA strand for destruction or removal [25]. The latter approach is particularly straightforward: uracil is incoporated during second-strand synthesis and the second strand is then selectively destroyed by uracil-N-glycosylase just prior to amplification. Only minimal modifications to the standard protocols are required and the procedure adds only a 15 minute incubation step. Nevertheless, the procedure still suffers from biases introduced by amplification.

A recently described protocol, FRT-Seq [13], uses directional single-strand linker ligation directly on RNA to generate a library comprising DNA adaptors flanking RNA inserts. The library is added directly to a GA flowcell, reverse transcribed on the surface, and then amplified by bridge PCR. As a result, there is no need for prior amplification of the sample, and the resulting sequences maintain strand information. By this simple protocol, the authors demonstrate impressive accuracy and reproducibility, and FRT-Seq may well be nearly optimal for routine RNA sequencing. It's major drawback is the need for several micrograms of input RNA, and the need to use a modified instrument workflow not supported by the manufacturer.

An amplification-free direct RNA sequencing protocol [26] has been described for the Helicos HeliScope single-molecule sequencing platform. In the most recent version of the approach, selective hexamers (devoid of hexamers complementary to rRNA) were used to prime first-strand cDNA from 250 pg total RNA [27], corresponding to just 25 mammalian cells.

A number of protocols have been developed for tag-based gene expression by sequencing, similar to serial analysis of gene expression (SAGE [28]), and kits are available from the instrument manufacturers. However, published protocols would not seem to offer any advantages over full-transcript RNA-Seq.

Finally, it should be noted that all these RNA-Seq protocols attempt to measure RNA concentrations (copy numbers). In contrast, it is often useful to measure instead the *rate* of transcription, as this will give a more direct measure of the transcriptional activity of a locus. This can be achieved by the GRO-Seq method [29], which uses a nuclear run-on assay with affinity-labeled nucleotides to isolate RNA only from transcriptionally active genes.

Discussion

We have described the general principles and best practices in preparing nucleic acid samples for high-throughput sequencing. The field has matured to the point where sample preparation from micrograms of genomic DNA is a well-understood process, and a standard procedure can be described that avoids the major causes of bias. **Supplementary Methods** provides our protocol, with annotations and a detailed description of the procedure and the materials needed. Libraries prepared in this way will generally be less expensive and of superior quality to those prepared using kits and protocols supplied by the manufacturers.

In contrast, sample preparation from RNA remains an area of active investigation. While RNA-Seq has already surpassed microarrays as the method of choice for expression analysis, no published protocol yet combines the desired characteristics of (i) uniform transcript coverage, (ii) single-cell sensitivity, (iii) lack

of amplification bias and (iv) a high level of multiplexing. Currently, the choice is between highly accurate, reproducible and unamplified methods that require micrograms of RNA (e.g. FRT-Seq [13]), on the one hand; and less accurate, highly amplified methods with single-cell sensitivity [16], on the other. In the future, we expect the gap between these extremes to close, as our understanding of the manipulation of small amounts of RNA is improved. To this end, microfluidic sample preparation techniques [30] and single-molecule sequencing [27,31] show great promise for the future.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.yexcr.2010.02.036.

REFERENCES

[1] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463–5467.
[2] A.M. Maxam, W. Gilbert, A new method for sequencing DNA, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 560–564.
[3] M.L. Metzker, Sequencing technologies - the next generation, Nat. Rev. Genet. 11 (2010) 31–46.
[4] C.W. Fuller, L.R. Middendorf, S.A. Benner, G.M. Church, T. Harris, X. Huang, S.B. Jovanovich, J.R. Nelson, J.A. Schloss, D.C. Schwartz, D.V. Vezenov, The challenges of sequencing by synthesis, Nat. Biotechnol. 27 (2009) 1013–1023.
[5] J.M. Rothberg, J.H. Leamon, The development and impact of 454 sequencing, Nat. Biotechnol. 26 (2008) 1117–1124.
[6] L. Mamanova, A.J. Coffey, C.E. Scott, I. Kozarewa, E.H. Turner, A. Kumar, E. Howard, J. Shendure, D.J. Turner, Target-enrichment strategies for next-generation sequencing, Nat. Methods 7 (2010) 111–118.
[7] F.J. Ghadessy, J.L. Ong, P. Holliger, Directed evolution of polymerase function by compartmentalized self-replication, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 4552–4557.
[8] D.H. Bing, C. Boles, M. Rehman, M. Audeh, B. Belmarsh, B. Kelley, C.P. Adams, Bridge amplification: a solid phase PCR system for the amplification and detection of allelic differences in single copy genes, Genetic Identity Conference Proceedings, Seventh International Symposium on Human Identification, , 1996.
[9] C. Gaillard, F. Strauss, Avoiding adsorption of DNA to polypropylene tubes and denaturation of short DNA fragments, Elsevier Trends Journals Technical Tips Online, 1998, pp. 1–2.
[10] A. Pihlak, G. Bauren, E. Hersoug, P. Lonnerberg, A. Metsis, S. Linnarsson, Rapid genome sequencing with short universal tiling probes, Nat. Biotechnol. 26 (2008) 676–684.
[11] M.A. Quail, I. Kozarewa, F. Smith, A. Scally, P.J. Stephens, R. Durbin, H. Swerdlow, D.J. Turner, A large genome center's improvements to the Illumina sequencing system, Nat. Methods 5 (2008) 1005–1010.
[12] I. Kozarewa, Z. Ning, M.A. Quail, M.J. Sanders, M. Berriman, D.J. Turner, Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes, Nat. Methods 6 (2009) 291–295.
[13] L. Mamanova, R.M. Andrews, K.D. James, E.M. Sheridan, P.D. Ellis, C.F. Langford, T.W. Ost, J.E. Collins, D.J. Turner, FRT-seq: amplification-free, strand-specific transcriptome sequencing, Nat. Methods 7 (2010) 130–132.
[14] C.A. Hutchison III, J.C. Venter, Single-cell genomics, Nat. Biotechnol. 24 (2006) 657–658.
[15] A. Walker, J. Parkhill, Single-cell genomics, Nat. Rev. Microbiol. 6 (2008) 176–177.
[16] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B.B. Tuch, A. Siddiqui, K. Lao, M.A. Surani, mRNA-Seq whole-transcriptome analysis of a single cell, Nat. Methods 6 (2009) 377–382.
[17] M. Meyer, A.W. Briggs, T. Maricic, B. Hober, B. Hoffner, J. Krause, A. Weihmann, S. Paabo, M. Hofreiter, From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing, Nucleic Acids Res. 36 (2008) e5.
[18] R.A. White III, P.C. Blainey, H.C. Fan, S.R. Quake, Digital PCR provides sensitive and absolute calibration for high throughput sequencing, BMC Genomics 10 (2009) 116.
[19] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, C.B. Burge, Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476.
[20] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (2008) 621–628.
[21] R. Yelin, D. Dahary, R. Sorek, E.Y. Levanon, O. Goldstein, A. Shoshan, A. Diber, S. Biton, Y. Tamir, R. Khosravi, S. Nemzer, E. Pinner, S. Walach, J. Bernstein, K. Savitsky, G. Rotman, Widespread occurrence of antisense transcription in the human genome, Nat. Biotechnol. 21 (2003) 379–386.
[22] 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, Stem cell transcriptome profiling via massive-scale mRNA sequencing, Nat. Methods 5 (2008) 613–619.
[23] R. Lister, R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, J.R. Ecker, Highly integrated single-base resolution maps of the epigenome in Arabidopsis, Cell 133 (2008) 523–536.
[24] N.J. Croucher, M.C. Fookes, T.T. Perkins, D.J. Turner, S.B. Marguerat, T. Keane, M.A. Quail, M. He, S. Assefa, J. Bahler, R.A. Kingsley, J. Parkhill, S.D. Bentley, G. Dougan, N.R. Thomson, A simple method for directional transcriptome sequencing using Illumina technology, Nucleic Acids Res. 37 (2009) e148.
[25] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, A. Soldatov, Transcriptome analysis by strand-specific sequencing of complementary DNA, Nucleic Acids Res. 37 (2009) e123.
[26] F. Ozsolak, A. Goren, M.A. Gymrek, M. Guttman, A. Regev, B.E. Bernstein, P.M. Milos, Digital transcriptome profiling from attomole-level RNA samples. Genome Res. (in press), doi:10.1101/gr.102129.109.
[27] 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, Direct RNA sequencing, Nature 461 (2009) 814–818.
[28] V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, Serial analysis of gene expression, Science 270 (1995) 484–487.
[29] L.J. Core, J.J. Waterfall, J.T. Lis, Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters, Science 322 (5909) (2008) 1845–1848.
[30] J.S. Marcus, W.F. Anderson, S.R. Quake, Microfluidic single-cell mRNA isolation and analysis, Anal. Chem. 78 (2006) 3084–3089.
[31] D. Pushkarev, N.F. Neff, S.R. Quake, Single-molecule sequencing of an individual human genome, Nat. Biotechnol. 27 (2009) 847–852.