Optimization of genomic DNA shearing by sonication for next-generation sequencing library preparation

Le Jie Lee¹,² and Maha Abdullah*¹

¹Immunology Unit, Department of Pathology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Malaysia
²Malaysia Genome Institute, National Institutes of Biotechnology Malaysia, Jalan Bangi, 43000 Kajang, Selangor Malaysia

Received 2nd October 2014/ Accepted 18th November 2014

Abstract. Next-generation sequencing (NGS) technologies, capable of sequencing genomic DNA and RNA at high throughput with unprecedented speed, have revolutionized genomic research as well as clinical diagnosis. DNA fragmentation is a critical step in library preparation in all NGS platforms, and determines the quality and diversity of the final library. DNA shearing by acoustic sonication is one of the ways to randomly break DNA into small fragments, however many variables affect the outcome. Here, we describe an optimized procedure to shear genomic DNA into fragments of 150 bp to 120 bp using a focused-ultrasonicator. Parameters that were assessed included DNA quantity, the effect of repeat shearing, treatment time, peak incident power and shearing reproducibility. This input of pure and optimum quality DNA samples is an essential starting point to the NGS system. We identified peak incident power as being the key determining factor in obtaining small target fragments. By increasing the peak incident power to 75W, a peak size within the 150 bp to 200 bp range was achievable, a result which was reproducible in multiple samples. Repeat shearing and increased treatment time were less successful in producing optimally sized DNA fragments. The proposed method may be used as a guide for NGS users involved in library construction, particularly when small fragment sizes are required.

Keywords: DNA shearing, Next-generation sequencing, NGS library preparation, Sonication.

INTRODUCTION

The Human Genome Project instigated a revolution in DNA sequencing methods and led a drive to overcome old limitations. New technologies were able to process hundreds of thousands to millions of DNA templates in parallel, resulting in low cost and throughput on the gigabase scale (Lander et al., 2001; McPherson et al., 2001; Sachidanandam et al., 2001). Newer methods in sequencing technology are referred to as next-generation sequencing (NGS) as compared to conventional Sanger sequencing (Mamanova et al., 2010). With the high-throughput parallel NGS sequencing platforms, complete sequencing of a whole human genome can be accomplished in less than three days at a cost of USD $1000 (Anon, 2014).

The various available NGS technologies rely on different strategies to combine template preparation, sequencing, imaging, genome alignment and assembly methods. NGS platforms provided by different service providers may be based on different combinations of sequencing chemistry such as Illumina (sequencing by synthesis), Roche 454 (pyrosequencing), SOLiD (sequencing by ligation), Ion Torrent (ion semiconduc-
carry over chemicals or enzymes that will affect the read qualities. Thus, comparison between libraries constructed by sonication and enzymatic shearing was made, since the former would have the advantage of avoiding carry-over; both were found to be effective (Quail et al., 2012b). However, the enzymatic method produces a greater number of artifact insertions/deletions in raw sequence reads compared with samples prepared by physical sonication (Knierim et al., 2011).

Genomic DNA shearing by acoustic sonication is thus one of the preferred methods for preparation of high quality, random and size-appropriate NGS libraries. In this study, we describe various parameters to consider during DNA shearing and the optimization of a procedure for shearing genomic DNA intro fragments in the range of 150 bp to 200 bp using a focused-ultrasonicator.

MATERIALS AND METHODS

DNA samples. DNA samples were isolated from peripheral blood mononuclear cells obtained from healthy volunteers recruited for this study, which was approved by the Institutional Medical Research Ethics Committee, Universiti Putra Malaysia. Informed consents were obtained.

Isolation of mononuclear cells and DNA extraction. Peripheral blood mononuclear cells were isolated from whole blood samples using density gradient centrifugation on the Ficoll-Paque Plus platform (GE Healthcare, Sweden) according to the manufacturer’s protocol. Isolated mononuclear cells were washed with 1× saline phosphate buffer and then suspended in 3 ml of nucleic acid lysis buffer (10 mM Tris-HCl, 400 mM NaCl and 2 mM Na2 EDTA, pH 8.2). The salting out procedure for extracting genomic DNA described by Miller et al. (1988) was performed after lysis of mononuclear cells. Extracted DNA was dissolved in TE buffer (10 mM Tris-HCL, 0.2 mM Na2 EDTA, pH 7.5).

Pre-fragmentation analysis of DNA quality. DNA quality was determined by measuring absorbance at 260 nm (A260), 280 nm (A280) and 230 nm (A230) with a Nanodrop 1000 Spectrophotometer (Thermo Scientific, USA). The A260/A280 and A260/A230 ratio values were used to estimate DNA purity. DNA integrity was confirmed by running an aliquot of each DNA sample on a 2% agarose gel stained with ethidium bromide.

DNA quantity was also determined using the Qubit dsDNA assay kit (Life Technologies, USA) and analysis using the Qubit Fluorometer (Invitrogen, USA).

DNA shearing parameters. The Covaris M220 Focused-ultrasonicator (Covaris, USA) was used for DNA shearing, set to produce DNA fragments with a target peak size range of 150 bp to 200 bp, that which is required for subsequent steps of SureSelect target enrichment (Agilent Technologies, USA) library preparation. The parameters provided in the user guide to obtain a target base pair peak of 150 bp were as follows: 50W of peak incident power, 20% of duty factor, 200 cycles per burst, and 350 seconds of treatment time at 20 °C. However, other parameters require optimization to the individual lab setting and user. In this study Peak incident powers of 50W, 70W and 75W were tested, an extra 20 seconds was added to treatment time to examine the shearing effect in an extended treatment time, the effect of repeat shearing on sheared DNA was examined, and two different DNA quantities, 3.5 µg and 5.0 µg, were prepared for shearing. The recovered DNA yield was compared after the shearing, and the reproducibility of the optimized method was tested using ten biological replicates.

Purification of fragmented DNA and post-fragmentation DNA quality analysis. The library of sheared DNA was purified using Agencourt AMPure XP beads (Beckman Coulter, USA) according to the manufacturer’s instruction. The purified DNA was eluted in 50 µl nuclease-free water (Qiagen, USA).

The quality of DNA, including the distribution of fragment length and peak height, was assessed using a 2100 Bioanalyzer (Agilent Technologies, USA). An Agilent DNA 1000 chip and a High Sensitivity DNA chip (Agilent Technologies, USA) were used to run the sheared DNA library data, and the electropherogram was analyzed by Agilent 2100 Expert Software version B.02.02.

RESULTS AND DISCUSSION

Pre-fragmentation DNA quality. A260/A280 and A260/A230 ratios and concentrations of extracted DNA were measured by both Nanodrop 1000 Spectrophotometer and Qubit Fluorometer are shown in Table 1. The A260/A280 ratio is used to indicate protein contamination of DNA; generally, a ratio in the range of 1.8-2.0 indicates good quality (Sambrook et al., 2001). The A260/A230 ratio is used as a secondary measure of DNA purity; absorption at 230 nm can be caused by contaminants other than protein such as phenolate ions, thiocyanate, organic compounds or polysaccharides. Expected A260/A230 ratios in pure DNA are commonly in the range of 2.0-2.2 (Turner et al., 2005).

Only samples producing values within these ranges were selected to proceed with DNA shearing. It was noticed that concentrations measured by the Qubit
Table 1. Quality of DNA (A_{260/280} and A_{260/230} ratios and concentrations measured by Nanodrop 1000 Spectrophotometer and Qubit Fluorometer) from ten genomic DNA samples extracted from mononuclear cells of human blood.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Ratio A_{260/280}</th>
<th>Ratio A_{260/230}</th>
<th>Concentration measured by Nanodrop (ng/µl)</th>
<th>Concentration measured by Qubit (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.85</td>
<td>2.29</td>
<td>596.5</td>
<td>137.3</td>
</tr>
<tr>
<td>2</td>
<td>1.80</td>
<td>2.28</td>
<td>458.1</td>
<td>314.7</td>
</tr>
<tr>
<td>3</td>
<td>1.84</td>
<td>2.18</td>
<td>310.0</td>
<td>148.0</td>
</tr>
<tr>
<td>4</td>
<td>1.88</td>
<td>2.20</td>
<td>224.7</td>
<td>130.0</td>
</tr>
<tr>
<td>5</td>
<td>1.84</td>
<td>2.25</td>
<td>387.5</td>
<td>133.0</td>
</tr>
<tr>
<td>6</td>
<td>1.85</td>
<td>1.97</td>
<td>208.6</td>
<td>121.3</td>
</tr>
<tr>
<td>7</td>
<td>1.85</td>
<td>2.28</td>
<td>1046.5</td>
<td>330.7</td>
</tr>
<tr>
<td>8</td>
<td>1.86</td>
<td>2.24</td>
<td>954.1</td>
<td>513.7</td>
</tr>
<tr>
<td>9</td>
<td>1.82</td>
<td>2.15</td>
<td>343.1</td>
<td>185.0</td>
</tr>
<tr>
<td>10</td>
<td>1.84</td>
<td>2.11</td>
<td>244.3</td>
<td>108.7</td>
</tr>
</tbody>
</table>

Fluorometer were 1.5 to 4 times lower in all samples compared to the concentrations measured using the Nanodrop 1000 Spectrophotometer. Based on other studies, it is common for spectrophotometry to overestimate the concentration, giving values ranging from 2 up to 30 times greater than their actual concentration (Mee et al., 2011; Sironen et al., 2011; Mathot et al., 2013). Although spectrophotometry is the most common method used to quantify nucleic acid, it can be unreliable and inaccurate (Clasel, 1995). This is because DNA is not the only molecule that can absorb UV light at 260 nm. Contaminants that will affect the purity of DNA such as RNA are highly absorbant at 260 nm, aromatic amino acids absorbing at 280 nm will contribute to the final measurement at these wavelengths, and the nucleotide guanidine gives an even higher value for 260 nm absorbance (Manchester, 1996). So, the major disadvantages of spectrophotometry are insensitivity and the tendency to overestimate the DNA quantity. As fluorometric technology utilizes specific probe dyes that will only bind to double-stranded DNA, the specificity and sensitivity are higher than the conventional spectrophotometry method in which the absorbance reading is affected by the background noise of contaminants (Georgiou and Papapostolou, 2006; Shokere et al., 2009; Singer et al., 1997). Thus, Qubit dsDNA assay is suggested as a more accurate method to measure the concentration of DNA for NGS library preparation.

Apart from the quality, the integrity of extracted DNA is also important for DNA shearing. Figure 1 shows the image of extracted DNA samples visualized by 2% agarose gel electrophoresis to assess the degree of degradation. Although extracted DNA is meant to be fragmented for downstream construction of NGS library, degraded or partially degraded DNA may produce smaller fragments that will affect average distribution of the desired target fragment length (Simbole et al., 2013).

Parameters affecting DNA fragmentation. The initial quantity of DNA affects the outcome of shearing. The average recovery of DNA after shearing of 3.5 µg DNA was 70.2% (±7.3) while with 5.0 µg the average recovery increased to 77% (±18.2). DNA recovery after shearing however might be sample dependent. Therefore, it is recommended that a higher starting amount of DNA is used to increase the yield since the maximum DNA input for Covaris M220 is 10 µg. However, too great a DNA input will also result in incomplete shearing causing tailing of the electropherogram where the majority of the DNA remains in larger sizes.

Figure 1. Gel image of 10 genomic DNA samples extracted from mononuclear cells of human blood (lanes labelled 1 to 10) run on a 2% agarose gel. Single bands indicated that the extracted genomic DNA samples were intact and not degraded.
The quality of DNA shearing is assessed as peak heights on electropherogram from the Bioanalyzer DNA 1000 chip (Table 2). The average peak size of ten fragmented specimens using the manufacturer’s recommended parameters was 205.3 bp (±15.4). The peaks produced by most of the samples were higher than the target of 150 bp to 200 bp and the electropherogram of the sheared DNA samples showed a very broad distribution of fragment sizes (Figure 2). Broad distribution of fragments indicates that there were a lot of fragments present that were smaller or larger than the target range. As seen in the overlaid electropherograms in Figure 2, the majority of the samples produced distribution of post-shearing fragment sizes that skewed to the right, meaning that the average size of the sheared DNA fragments was still larger than 150 bp to 200 bp. Indeed most of the sheared DNA samples had fragment lengths above 200 bp, and so an optimization of parameters was conducted.

Repeat shearing of DNA might be an option to avoid wastage of DNA. Samples from rare species, clinical samples and histological fixed sample are precious as the quantity of nucleic acid extracted is low. As the results of repeat shearing have not been previously reported, three previously sheared DNA samples (no. 4, 6 and 7) were selected for repeat shearing for an additional 1 minute. Samples no. 4 and 7 (Figure 3a) showed a shift of the peak and size distribution, from 206 bp to 180 bp and from 205 bp to 199 bp, respectively. Although improvement was observed the peak size was still skewed to the right, slightly out of the target range of 150 bp to 200 bp. The fragment size distribution remained broad. Sample no. 6 on the other hand, resulted in a peak size shift from 240 bp to 178 bp. The distribution curve of fragment sizes for this sample was sharp and fitted nicely into the target range of 150 bp to 200 bp (Figure 3b). Although good results were observed in sample no. 6, repeat shearing of sheared DNA is still not recommended as it seems to create a lot of off-target fragments that will broaden the fragment size distribution (shown in Figure 3a, graph region from 200 bp to 700 bp).

Increasing treatment time on the other hand may help resolve this problem. As the DNA shearing process is rate-limiting, fragment size generation, defined by mean peak base pair size, can be affected by treatment duration. A longer duration will produce smaller fragments (Dibya et al., 2012; Schoppee and Wämhoff, 2011). Thus, treatment time was extended by 20 seconds from the recommended time in processing five samples of intact DNA. The results however, did not show much alteration from original average peak base pair size

Table 2. Peak sizes of ten genomic DNA samples extracted from mononuclear cells of human blood, determined from Bioanalyzer electropherograms.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Peak size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>186</td>
</tr>
<tr>
<td>2</td>
<td>194</td>
</tr>
<tr>
<td>3</td>
<td>199</td>
</tr>
<tr>
<td>4</td>
<td>206</td>
</tr>
<tr>
<td>5</td>
<td>223</td>
</tr>
<tr>
<td>6</td>
<td>240</td>
</tr>
<tr>
<td>7</td>
<td>205</td>
</tr>
<tr>
<td>8</td>
<td>199</td>
</tr>
<tr>
<td>9</td>
<td>199</td>
</tr>
<tr>
<td>10</td>
<td>202</td>
</tr>
</tbody>
</table>

Average 205.3±15.4 bp

Figure 2. Electropherogram of sheared DNA run on a DNA 1000 Bioanalyzer. Fragmentation procedures were conducted as recommended by the manufacturer. The two small peaks on the far left and right of the graph were the upper and lower markers of DNA ladder. Six overlaid samples (indicated by different colored linear plots) show the distributions were outside the 150 bp to 200 bp range. Fluorescent unit (FU) on the y-axis is indicative of DNA concentration.
Optimization of DNA shearing by sonication

Figure 3. Electropherograms after repeat shearing of DNA sample. a) Repeat shearing improved fragment size from 206 bp and 205 bp of sample no. 4 and no. 7 to 180 bp and 199 bp. However, off-target fragments were still present (region from 200 bp to 700 bp). b) Repeat shearing improved fragment size of sample no. 6 from 240 bp to 178 bp. The region in green box indicates the target peak size range 150 bp to 200 bp. Fluorescent unit (FU) on the y-axis is indicative of DNA concentration.

(201.4±17.2 bp vs 205.3±15.4 bp). The overlaid electropherograms of five sheared DNA samples are shown in Figure 4. Longer treatment times may help to achieve the target peak size, but this is not recommended as it may cause a broader peak size distribution and accumulation of small fragments by breaking down more DNA. Moreover, a prolonged acoustic sonication treatment time may increase the temperature and affect the stability of the DNA sample, and a treatment time of more than 370 seconds (recommended time is 350 seconds) is very time consuming especially when preparing a pooled NGS library that consists of a large number of samples.

Peak incident power represents the power (Watts) emitted from the transducer during each cycle (Covaris, 2011). The Covaris shearing treatment consists of very high frequency acoustic signals emitted in a series of bursts with each burst being followed by a zero power state. Cycles per burst in the setting indicates the number of acoustic oscillations contained in each burst. Figure 5 shows the electropherograms of two intact samples after increasing shearing power to 70W and 75W. Increase in peak incident power enhanced the shearing effect thus achieving fragments in the target range of 150 bp to 200 bp from both samples.

Lastly, the ability to reproduce target peak sizes using a peak incident power of 75W was confirmed using ten new intact DNA samples with the achievement of an average peak size of 175±4.6 bp. Figure 6 shows the electropherogram of seven overlaid samples and all the peak sizes were within the range of 150 bp to 200 bp. Off-target fragments (those appearing to be larger than 200 bp on the electropherogram) produced as a result of incomplete shearing were comparatively less common than from the previous setting. Peak incident power and treatment time were described as key factors in shearing bacterial genomic DNA by Jeannotte et al. (2014). A 175W peak incident power and 180 seconds of treatment time were identified to be optimal values for shearing bacterial genomic DNA by Covaris E220, though this differs somewhat from the more cost-effective M220 model that we used.

As the length of adaptor sequence is constant in NGS library preparation, the size of the DNA fragment from the library which is inserted between the adapter sequences, determined by the shearing step, becomes very important. Optimal insert size is determined by the limitation of the NGS instrumentation and by the specific sequencing application. Sequencing fragments that do not fall in
Figure 4. Overlaid electropherograms of five DNA samples extracted from mononuclear cells of human blood with an extended 20 second treatment time. The extended time did not produce a significant difference in the peak size. Fluorescent unit (FU) on the y-axis is indicative of DNA concentration.

Figure 5. Electropherograms of two DNA samples, extracted from mononuclear cells of human blood, produced using peak incident powers of 70W and 75W. Sample 1 is represented in a) and b), and sample 2 in c) and d). a) 70W, peak size: 186 bp, b) 75W, peak size: 180 bp, c) 70W, peak size: 185 bp, and d) 75W, peak size: 178 bp.

the recommended size distribution may cause low read depth or decrease in the coverage of specific regions of the sequence (Head et al., 2014). For the Illumina NGS platform, insert size impacted the process of cluster generation (library denaturation, dilution and distribution on the two-dimensional surface of the flow-cell followed by amplification of the library) by means of amplification efficacy i.e. shorter products amplify more efficiently than longer products (Sakharkar et al., 2004). In the case of SureSelect target enrichment or other target enrichment systems, fragment size distribution can affect the final enrichment efficiency and the percentage of on-target capture by the probes which will affect the final results and quality of sequencing.

Fragmentation to the size range to suit the requirements of different NGS platforms is also crucial to avoid sequencing bias, as when the fragmented DNA are too variable in size, the smaller off-target fragments tend to
be amplified more easily, resulting in over-representation of these fragments. Thus, the optimization of the range of insert sizes is important to minimize amplification bias (Aird et al., 2011). Apart from the instrumental considerations, the structure and GC content of a DNA sample may also affect the fragmentation process which eventually will lead to an uneven coverage of reads due to bias in GC-poor or GC-rich sequences (Kozarewa et al., 2009; Poptsova et al., 2014). Breaking of double-stranded DNA by sonication occurs preferentially in 5′-CpG-3′ dinucleotides (Grokhovsky, 2006). Differential cleavage rates (CpG > CpA/T > CpC) mean that the efficacy of sonication for DNA fragmentation is sequence dependent, determined by the CG content (Grokhovsky et al., 2011; Dohm et al., 2008). However, the precise effect of GC content on the quality of NGS data needs to be further examined.

CONCLUSION

To initiate the preparation of an NGS library, the production of good quality DNA (with regards to $A_{260/280}$, $A_{260/230}$ ratios and integrity) is crucial, and several parameters and techniques should be used to obtain the best results. Fluorometry is a more accurate method to quantify DNA than spectrophotometry. A higher input of DNA is recommended in order to counter the loss of DNA that occurs during shearing. Repeating or increasing the length of DNA shearing treatment is not recommended as there are disadvantages to these methods. We conclude that peak incident power is the key factor in achieving a small target peak without accumulating too many smaller fragments. By increasing the peak incident power to 75W, a greater proportion of fragments in the target 150 bp to 200 bp range can be produced, results which were reproducible across multiple samples. This optimized method can therefore be useful when small fragments are required in NGS library preparation.

ACKNOWLEDGEMENTS

This study was supported by the Malaysia Genome Institute, Ministry of Science, Technology and Innovation. We would like to acknowledge the contribution of Irni Suhayu Sapian and Norfarazieda Hassan. We would also like to thank all volunteers for their participation.

REFERENCES


Optimization of DNA shearing by sonication


Mee, B.C., Carroll, P. and Donatello, S. 2011. Maintaining breast cancer specimen integrity and individual or simultaneous extraction of quality DNA, RNA and proteins from all protect-stabilized and nonstabilized tissue samples. Biopreservation and Biobanking 9: 389-398.


