Biochemical and Biophysical Research Communications 425 (2012) 379–383

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

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

Validation of whole genome amplification for analysis of the p53 tumor suppressor gene in limited amounts of tumor samples

Johanna Hasmats [a], Henrik Green [a], Beata Werne Solnestam [a], Pawel Zajac [b], Mikael Huss [c], Cedric Orear [d], Pierre Validire [e], Magnus Bjursell [f], Joakim Lundeberg [a,*]

[a] Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, Royal Institute of Technology, Stockholm, Sweden
[b] Laboratory for Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[c] Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden
[d] Genomics Unit, Institut Gustave Roussy, Villejuif, France
[e] Department of Pathology, Institut Mutualiste Montsouris, Paris, France
[f] AstraZeneca R&D, Mölndal, Sweden

ARTICLE INFO

Article history:
Received 18 July 2012
Available online 27 July 2012

Keywords:
Whole genome amplification
TP53
Mutations
Validation

ABSTRACT

Personalized cancer treatment requires molecular characterization of individual tumor biopsies. These samples are frequently only available in limited quantities hampering genomic analysis. Several whole genome amplification (WGA) protocols have been developed with reported varying representation of genomic regions post amplification. In this study we investigate region dropout using a φ29 polymerase based WGA approach. DNA from 123 lung cancers specimens and corresponding normal tissue were used and evaluated by Sanger sequencing of the p53 exons 5–8. To enable comparative analysis of this scarce material, WGA samples were compared with unamplified material using a pooling strategy of the 123 samples. In addition, a more detailed analysis of exon 7 amplicons were performed followed by extensive cloning and Sanger sequencing. Interestingly, by comparing data from the pooled samples to the individually sequenced exon 7, we demonstrate that mutations are more easily recovered from WGA pools and this was also supported by simulations of different sequencing coverage. Overall this data indicate a limited random loss of genomic regions supporting the use of whole genome amplification for genomic analysis.

1. Introduction

DNA based diagnostics are becoming increasingly important in cancer treatment and prevention, and the increasing knowledge is leading to more tailored analysis of biomarkers. Ever since the discovery of the tumor suppressor gene p53 in 1979 by Arnold Levine [1], there have been numerous approaches to investigating this particularly interesting gene in various cancers [2]. It is activated by DNA damage and acts as a master regulator for gene networks involved in DNA repair, cell cycle arrest and apoptosis. There are several studies reporting the clinical associations between mutations in p53 and various forms of cancers (reviewed in [3,4]), including lung cancer. For example, the prevalence of somatic mutation association in the p53 gene ranges from 50% in non-small cell lung cancer (NSCLC) to 80% in small cell lung cancer (SCLC) [5,6]. The role of the high frequency of p53 mutations in lung cancer is still under investigation, in particular its relation the overall genome mutation patterns have been presented after whole genome sequencing of a small cell lung cancer cell line [7].

To facilitate the application and use of both broad genomic tools, such as massive parallel sequencing, and more focused studies of biomarker genes on small biopsies and tumor samples, increased amount of genomic DNA is commonly required. Robust methods are therefore needed to amplify the genomic material from scarce samples without compromising the quality or introducing significant bias. PCR-based methods for whole genome amplification (WGA) include primer extension pre-amplification (PEP), improved-PEP (I-PEP), degenerate oligonucleotide primed PCR (DOP-PCR) and ligation mediated PCR [8–10]. Additionally, there are several non-PCR based methods such as strand displacement amplification (SDA), T7-based linear amplification of DNA (TLAD) and isothermal multiple displacement amplification (MDA) (reviewed in [11]). MDA is based on the φ29 polymerase enzyme, derived from *Bacillus subtilis* bacteriophage.

Amplification bias and allelic drop-out are issues in current whole genome amplification methods when starting from low input amounts of DNA. Different laboratories have reported their

* Corresponding author. Address: Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, Royal Institute of Technology, Stockholm, Sweden P.O. Box 1031, SE-171 21 Solna, Sweden.
E-mail address: joakim.lundeberg@scilifelab.se (J. Lundeberg).

0006-291X/
http://dx.doi.org/10.1016/j.bbrc.2012.07.101

preferred approach in terms of performance [12–14]. The short product length of the PCR based methods is a limitation for some applications and this together with the ease of use of the MDA approach makes several research groups focus on MDA [15–17].

In this study, we investigate the use of a φ29 polymerase based WGA strategy by Sanger sequencing of amplicons, as an approach to analyze scarce lung cancer tumor samples. In particular we focused on the exons 5–8 in p53 tumor suppressor gene in individual normal and tumor samples. The results can serve as a valuable guide for projects requiring whole genome amplification starting from minute amounts of DNA.

2. Materials and methods

2.1. Samples

A total of 123 patients diagnosed with non-small cell lung cancer or small cell lung cancer were used in this study obtained from Institut de Gustave Roussy and Institut Mutualiste Montsouris (IMM), Paris. The homogeneity of tumor cells was according to microscopical examination >70%. Tumor and corresponding normal genomic DNA was extracted by conventional means. Genomic DNA concentration spanned from 2 to 41.6 ng/ul. Samples were used in accordance with the ethical guidelines of CHEMORES consortium (http://www.chemores.org) with the purpose of generating a systems biology database to study resistance to chemotherapy.

2.2. Whole genome amplification

The Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, Waukesha, Wisconsin) with random hexamer primers was used to amplifying the DNA according to manufacturer's instructions [18,19]. To confirm that the WGA reaction was successful samples was subjected to gel electrophoresis (data not shown).

2.3. PCR of individual samples

Four separate PCR reactions were performed for each WGA template to amplify the p53 exons 5–8. The AmpliTaq Gold kit (Applied Biosystems, Foster City, California) and FastStart HiFi PCR system (Roche, Basel, Switzerland) were used, following the manufacturer's protocol. The PCR cycle conditions were as follows: 40 cycles; 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 10 min. For the latter: 40 cycles; 94 °C for 30 s, 60 °C for 45 s, 72 °C for 75 s and a final extension at 72 °C for 5 min. The quality of the PCR was assessed by running a arbitrary set of PCR reactions on a gel (data not shown).

2.4. Sequencing and data analysis

For each PCR reaction, a cycle sequencing reaction was carried out; all PCR products were sequenced using conventional Sanger capillary sequencing technology [20]. The sequencing reactions were performed using BigDye Terminator v3.1 Cycle Sequencing Kit, (Applied Biosystems, Foster City, California), using the following conditions: 35 cycles; 94 °C for 40 s, 59 °C for 40 s, 72 °C for 100 s and a final extension at 72 °C for 8 min. The obtained sequences (tumor and normal) were analyzed and aligned to the following; exon 5: position 625–810 in NM_000546.3, exon 6: position 811–928 in NM_000546.3, exon 7: position 924–1033 in NM_000546.3, exon 8: position 1033–1170 in NM_000546.3. This alignment was used to infer SNP and mutation information. We defined a mutation to be present when the following criteria were met:>30% signal from an alternative base in both forward and reverse direction. Each position where a plausible mutation was present was examined in dbSNP [21] in order to annotate it as an already known SNP.

2.5. Validation - pooling and cloning

In order to validate if the WGA would introduce amplification bias and/or allele drop-out, a pooling strategy was used to overcome the lack of sufficient amounts of sample DNA. Two pools were generated using the cancer samples; one containing equal amounts of unamplified genomic DNA from all 123 cancer samples and a corresponding pool using the WGA DNA from the same set of samples. The purpose of this was to determine how many of the mutations detected from the amplicon sequencing of individual samples that are represented in each pool. We focused our analysis on exon 7 since it displayed the highest frequency of mutations among the four analyzed exons.

Exon 7 of the two pools were amplified using AmpliTaq Gold kit (Applied Biosystems, Foster City, California), using the following PCR cycling conditions: 40 cycles; 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 10 min. After amplification, the PCR products were cloned into *Escherichia coli* (pJET1.2/blunt Cloning Vector, CloneJET™ PCR Cloning Kit, Fermentas International Inc., Burlington, Canada) to allow quantitative determination of the mutation frequency.

To assess what level of redundancy was necessary to recover all mutations found in the individually sequenced exon 7, colonies were picked from each pool to reach a coverage of $\geqslant$10x. A total of 1800 individual colonies were picked and sequenced for each pool. The sequencing reactions were performed as described above.

2.6. Simulations

Since it is not feasible to practically detect all mutations by cloning and sequencing PCR fragments amplified from a pool even at very high coverage, we simulated re-sampling of the clones and then calculated a theoretical value of minimum redundancy. The average number of recovered mutations is calculated as a function of coverage for unamplified and amplified samples, respectively. Non-linear curve fitting (the "nls" function in R [22]) was used to fit saturating curves to the data points. For the unamplified case, the curve is of the form:

$N = 20\text{-}(a/(\text{coverage-}b))$
where N is the number of recovered mutations, $a = 120$ and $b = -5.95$.
For the WGA case, the curve is of the form
$N = 20^{*}(1 - \exp(-a^{*}(\text{coverage-}b))$
where $a = 0.111$ and $b = 0.0528$.

3. Results

3.1. Mutation detection in p53 from individual cancer biopsies

To enable a comparative analysis of the whole genome amplification (WGA) procedure we selected the tumor suppressor gene p53, since it has been reported to be frequently mutated in lung cancer [23]. Exons 5–8 of the p53 tumor suppressor gene were investigated to identify the most informative region in the current material for the comparison. Each genomic sample (tumor and normal) was first whole genome amplified followed by PCR amplification of the selected exons. The PCR products were sequenced using the Sanger methodology and the data was aligned to the p53 reference sequence, followed by manual mutation detection identifying gene variants and mutations (as outlined in Fig. 1A). A mutation

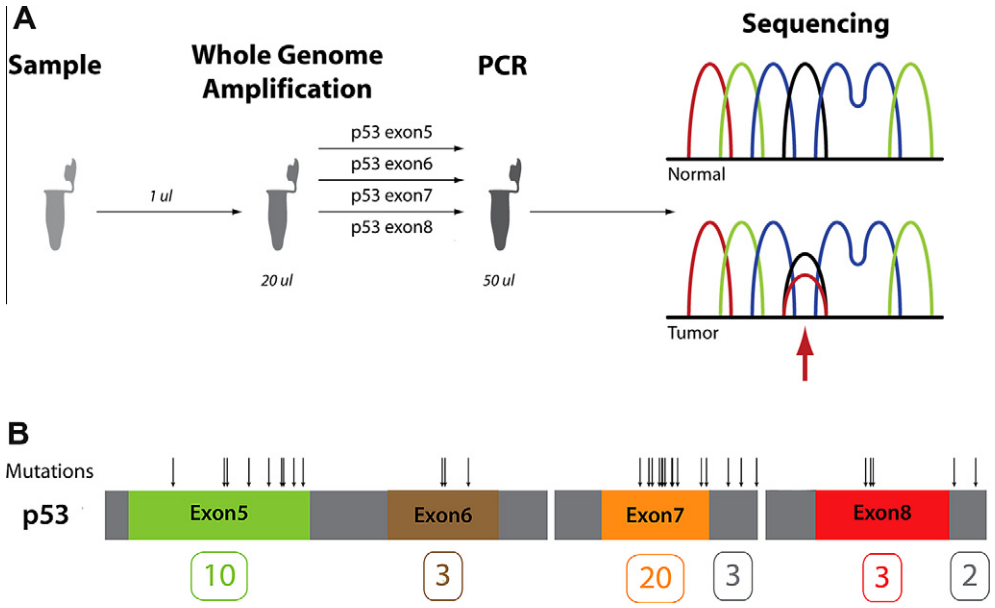

Fig. 1. Experimental design and mutational spectra. (A) Experimental design. The 123 tumor and normal samples in concentrations 2–41.6 ng/ul were used for whole genome amplification. This was followed by PCR targeting exon 5–8 in p53. All amplicons were Sanger sequenced in a forward and reverse direction and aligned against the reference sequence. (B) Mutation spectra for exons 5–8. The four exons targeted are schematically visualized. Mutations are shown with black arrows. The number of mutations is shown with corresponding color, and mutations outside exons are indicated in grey.

was called when the mutated signal was >30% over wild type allele.

98% of the PCR products yielded usable sequence data resulting in discovery of a total of 40 p53 mutations among the 123 samples, i.e., 33% mutation frequency. The mutations were found in all exons but the majority of mutations were found in 7 (20 of 40 mutations). Five mutations were identified outside the exon boundaries (Fig. 1B). The identified mutations were compared to known lung cancer variants in the IARC database (http://www-p53.iarc.fr/). Mutation prevalence and distribution was found to be similar to the 28.8% for exons 5–8, reported in the database (data not shown).

3.2. Mutation recovery from whole genome amplified and unamplified sample pools

To investigate if our whole genome amplification introduces bias or leads to random loss of mutations, a pooling strategy was devised to circumvent the low amounts of available material. Two equimolar pools of all 123 cancer specimens were created; one using unamplified genomic DNA and one containing a pool of whole genome amplified genomic DNA. Exon 7 was chosen as target region due to its relatively high mutation frequency. 14 of the 20 mutations identified in exon 7 were non-synonymous. To allow a quantitative estimation of the mutation frequency a cloning procedure was used. PCR amplification of exon 7 from these pools were cloned into *E.coli* to facilitate sequencing of individual samples. Over 1200 clones were analyzed corresponding to an approximately 10x coverage for the unamplified pool and over 1600 clones were analyzed from the whole genome amplified pool corresponding to an approximately 13x coverage. Mutations were manually detected and compared to the mutations found during the individual sample sequencing. The number of mutations recovered was 13/20 and 16/20 from the unamplified and WGA pools, respectively.

Plotting the number of times each mutation was detected in the pools reveals similarities in the mutation detection distribution. For the most abundant mutations, similar patterns were observed in the unamplified and WGA pools, however, mutations with a low frequency were exclusively detected in the WGA pool. Of the six synonymous mutations identified in the individual sequencing, 3 were observed in both unamplified and WGA pools, and one of the 3 was heterozygous in the WGA pools. Two additional mutations were found homozygous, exclusively observed in the WGA pools. (Supplementary Table S2). This suggests that WGA bias is limited for these selected regions, given such low amounts of starting material (Fig. 2).

3.3. Investigation of the effects of sequence coverage on mutation recovery

To determine how the different levels of sequence coverage affect mutation detection in the two different DNA pools, we simulated how many mutations would be recovered at different levels of sequence coverage. For a more reliable estimation, we randomly selected the mutation results from one PCR plate (96 sequences) at a time in an additive manner to make a cumulative curve of mutations detected as a function of coverage (by counting number of plates analyzed). This procedure was repeated 100 times both for the unamplified and WGA samples. The obtained data points for the average of 100 simulated runs were plotted with confidence intervals and a saturating curve was fitted to them (Fig. 3A). Given these data points and the corresponding fitted curves, we extrapolated to estimate the average number of recovered mutations as a function of sequencing coverage (Fig. 3B).

Since the functions of the curves are saturated and approaches the limit value of 20 mutations asymptotically, it is per definition impossible to reach the corresponding coverage. To circumvent this issue, values for recovering 80% (16 mutations) or 90% (18 mutations) of the 20 total mutations were estimated. We found that a theoretical coverage of at least 11.4x was needed to recover 80% of the mutations (detected in the first part of this study) when looking at the WGA material, as opposed to approximately 18.2x in the unamplified DNA. The corresponding values for 90% mutation recovery are 16.2x and 41.0x coverage for the WGA and unamplified pools, respectively. See Fig. 1(A and B) in the supplementary

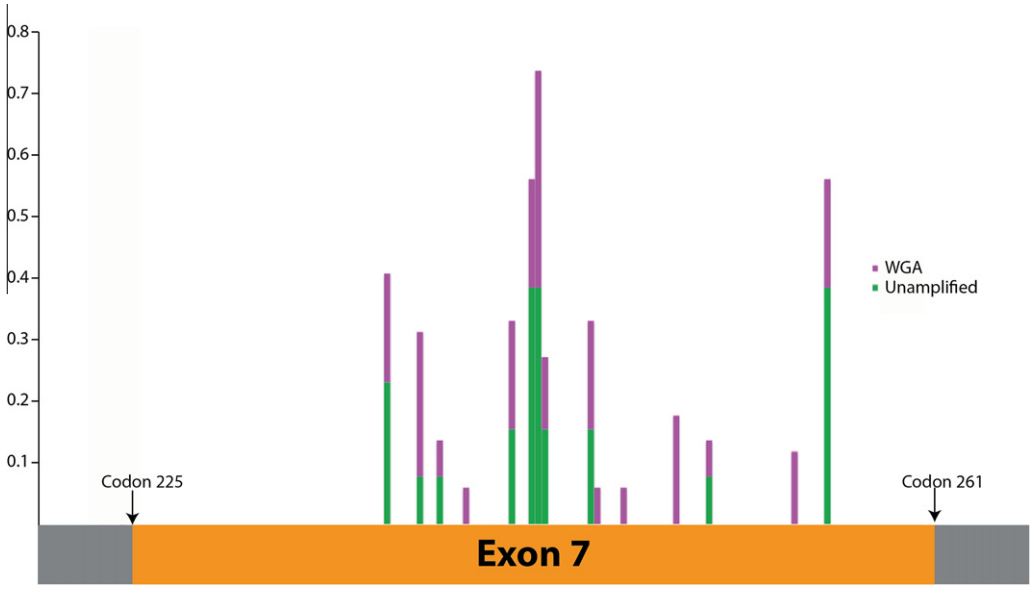

Fig. 2. Exon 7 mutations recovered from unamplified and WGA pools. A summary of the mutations found in exon 7 is visualized by the frequency of each mutation, normalized with respect to coverage. Mutation frequency for unamplified pool is shown in dark green and WGA pool in purple. It can be observed that the mutation frequencies follow a similar pattern for the unamplified and WGA pools. This suggests that WGA is working well on this material, even with very low amounts of starting material.

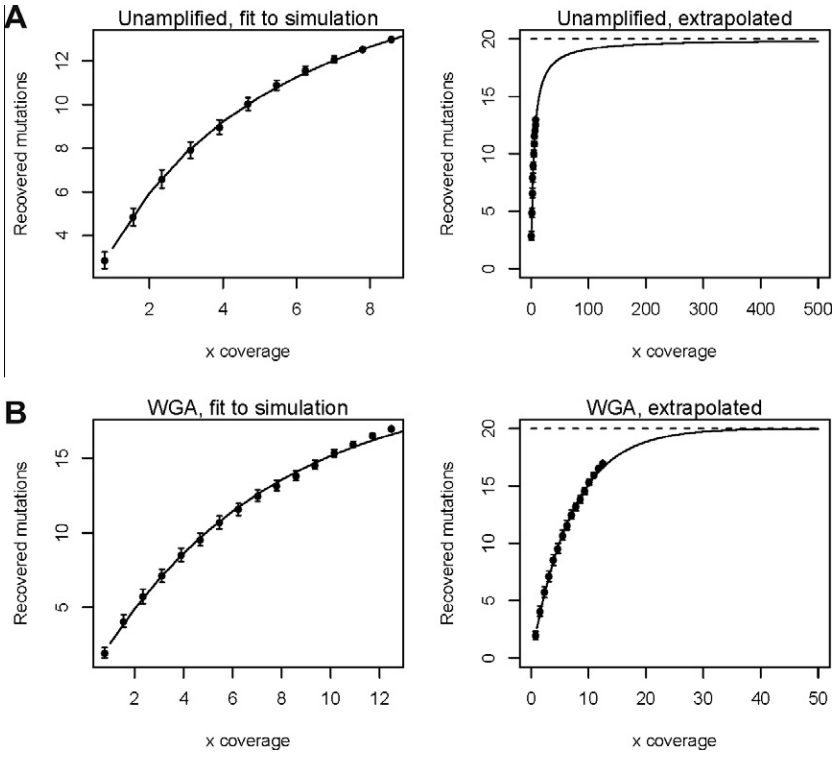

Fig. 3. Simulations and fitted curves of resampling. (A) Average number of recovered mutations are plotted as a function of coverage for unamplified and WGA samples after 100 rounds of random shuffling of the original samples. Mean values and 95% confidence intervals are indicated with bars. (B) Extrapolations of the observed data in (A) where theoretical coverage for a specific number of mutations can be estimated.

section for details that give further support and validation of our findings.

4. Discussion

Tumor biopsies are usually available in very limited quantities; however, DNA consuming genomic analysis platforms are today adding vital information to cancer research and are starting to be used in cancer diagnostics. Furthermore, samples are often distributed to multiple sites that have specialized in different analysis tools. Thus amplification of genomic DNA from very limited quantities is of critical importance to enable thorough genetic investigations and sample distribution for independent site-specific confirmation of findings.

In this study, we identify a significant number of mutations in the p53 exons using a WGA approach based on multiple displacement amplification of 123 tumor normal paired samples. Our data supports the common view that p53 is a frequent target for mutations in lung cancer and that most mutations identified are only present in one or a few individuals. Moreover, the mutation prevalence in exons 5–8 found in this study is comparable to mutations reported in the database (http://www-p53.iarc.fr/).

Whole genome amplification has been used in a number of studies, but there has been a controversy related to bias and allele drop out resulting in alternative interpretations. Here we suggest that WGA amplifies genomic regions without significant bias or genome loss substantiated by a direct overlap between mutation distribution found in unamplified material and corresponding WGA material. Interestingly, when exploring bias from the whole genome amplification, it was noticed that a larger portion of the mutations at low frequency in the pool were found in WGA material as compared to the original unamplified samples. A plausible explanation for this is that it is difficult to achieve an even representation of all samples in the pool when using minuscule amounts of starting material. After whole genome amplification the relative concentration distribution is more even among the samples and WGA can therefore serve as an approach to normalize between samples if a pooling strategy is employed such as in screening of mutations in larger population cohort.

Based on our simulations, we did not observe significant sample dropout in our WGA pool. This is of high importance when approaching analysis on a single cell level, where there is always a risk of DNA loss, and thus genomic regional loss, during sample handling. A theoretical coverage of approximately 41x is needed to recover 90% of the mutations in the original samples, whereas only about 16x is sufficient for the WGA pool. This is in accordance with previously published results [24].

With the advent of genome sequencing as a clinical diagnostics method, the importance of unbiased WGA is becoming more important. It will be valuable in the field of cancer treatment with the possibility to perform several important quality controls, as a result of accurate DNA amplification. We believe this study sheds light on the performance of this method.

The tumor suppressor gene p53 has been extensively studied since its discovery, and plays an important role in several cancers, including lung cancer. We present the mutation spectra of exons 5–8 in p53 from lung cancer biopsies and show that applying WGA on the samples does not affect the composition of variations, validated by comparison with the p53 database from IARC and comparison with unamplified material. Not surprisingly, our results show that the method described here aids in the detection of individual mutations in pools of patients. There is no definite number of minimum coverage for mutation detection, however, it is mentioned that at least 10x is required for some software [25] and 25x for a 50% false positive error rate degradation in performance metrics (454 sequencing) [26]. In conclusion, whole genome amplification is a valuable tool when dealing with small samples, and can be used without negotiating genetic composition.

Acknowledgments

The authors thank Bahram Amini for technical help and assistance during the Sanger sequencing. We also thank Lili Gong for her help with the sequencing preparations.

This work was financially supported by Grants from the European Commission (CHEMORES LSHC-CT-2007-037665). We would like to acknowledge Science for Life Laboratory (SciLifeLab Stockholm), SNISS, and Uppmax for providing sequencing and computational infrastructure.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2012.07.101.

References

[1] D.I. Linzer, A.J. Levine, Characterization of a 54 K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells, Cell 17 (1) (1979) 43–52.
[2] J.M. Nigro, Mutations in the p53 gene occur in diverse human tumour types, Nature 342 (6250) (1989) 705–708.
[3] M.M. Maslon, T.R. Hupp, Drug discovery and mutant p53, Trends in Cell Biology 20 (9) (2010) 542–555.
[4] A.I. Robles, C.C. Harris, Clinical outcomes and correlates of TP53 mutations and cancer, Cold Spring Harbor Perspectives Biology 2 (3) (2010) a001016.
[5] Y. Li, A meta-analysis of TP53 codon 72 polymorphism and lung cancer risk: evidence from 15,857 subjects, Lung cancer 66 (1) (2009) 15–21.
[6] S. Dai, P53 polymorphism and lung cancer susceptibility: a pooled analysis of 32 case-control studies, Human Genetics 125 (5–6) (2009) 633–638.
[7] E.D. Pleasance, A small-cell lung cancer genome with complex signatures of tobacco exposure, Nature 463 (7278) (2010) 184–190.
[8] W. van Elmpt, 3D dose delivery verification using repeated cone-beam imaging and EPID dosimetry for stereotactic body radiotherapy of non-small cell lung cancer, Radiotheraphy and Oncology 94 (2) (2010) 188–194.
[9] L. Lovmar, Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA, Nucleic Acids Research 31 (21) (2003) e129.
[10] O. Alsmadi, Specific and complete human genome amplification with improved yield achieved by phi29 DNA polymerase and a novel primer at elevated temperature, BMC Research Notes 2 (2009) 48.
[11] S. Hughes, The use of whole genome amplification in the study of human disease, Progress in Biophysics and Molecular Biology 88 (1) (2005) 173–189.
[12] F.B. Dean, Comprehensive human genome amplification using multiple displacement amplification, Proceedings of the National Academy of Sciences of the United States of America 99 (8) (2002) 5261–5266.
[13] M. Hockner, Whole genome amplification from microdissected chromosomes, Cytogenetic and Genome Research 125 (2) (2009) 98–102.
[14] D.L. Barker, Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel, Genome Research 14 (5) (2004) 901–907.
[15] Y. Hou, Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm, Cell 148 (5) (2012) 873–885.
[16] X. Xu, Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor, Cell 148 (5) (2012) 886–895.
[17] R.S. Lasken, Genomic DNA amplification by the multiple displacement amplification (MDA) method, Biochemical Society Transactions 37 (Pt 2) (2009) 450–453.
[18] F.B. Dean, Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification, Genome Research 11 (6) (2001) 1095–1099.
[19] P.M. Lizardi, Mutation detection and single-molecule counting using isothermal rolling-circle amplification, Nature Genetics 19 (3) (1998) 225–232.
[20] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proceedings of the National Academy Sciences USA 74 (12) (1977) 5463–5467.
[21] S.T. Sherry, DbSNP: the NCBI database of genetic variation, Nucleic Acids Research 29 (1) (2001) 308–311.
[22] R: A Language and Environment for Statistical Computing. 2010, R Foundation for Statistical Computing.
[23] L. Ding, Somatic mutations affect key pathways in lung adenocarcinoma, Nature 455 (7216) (2008) 1069–1075.
[24] R. Pinard, Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing, BMC Genomics 7 (2006) 216.
[25] V. Bansal, Accurate detection and genotyping of SNPs utilizing population sequencing data, Genome Research 20 (4) (2010) 537–545.
[26] O. Harismendy, Evaluation of next generation sequencing platforms for population targeted sequencing studies, Genome Biology 10 (3) (2009) R32.