

PCR-free method detects high frequency of genomic instability in prostate cancer

Nick M. Makridakis^{1,*}, Troy Phipps², Sudesh Srivastav³ and Juergen K. V. Reichardt⁴

¹Department of Epidemiology, Tulane University, New Orleans, LA 70112, ²Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA 90033, ³Department of Biostatistics, Tulane University, New Orleans, LA 70112, USA and ⁴Plunkett Chair of Molecular Biology (Medicine), University of Sydney, Camperdown, NSW 2006, Australia

Received May 5, 2009; Revised August 28, 2009; Accepted August 31, 2009

ABSTRACT

Most studies of tumor instability are PCR-based. PCR-based methods may underestimate mutation frequencies of heterogeneous tumor genomes. Using a novel PCR-free random cloning/sequencing method, we analyzed 100 kb of total genomic DNA from blood lymphocytes, normal prostate and tumor prostate taken from six individuals. Variations were identified by comparison of the sequence of the cloned fragments with the nr-database in Genbank. After excluding known polymorphisms (by comparison to the NCBI dbSNP), we report a significant over-representation of variants in the tumors: 0.66 variations per kilobase of sequence, compared with the corresponding normal prostates (0.14 variations/kb) or blood (0.09 variations/kb). Extrapolating the observed difference between tumor and normal prostate DNA, we estimate 1.8 million somatic (*de novo*) alterations per tumor cell genome, a much higher frequency than previous measurements obtained by mostly PCR-based methods in other tumor types. Moreover, unlike the normal prostate and blood, most of the tumor variations occur in a specific motif ($P = 0.046$), suggesting common etiology. We further report high tumor cell-to-cell heterogeneity. These data have important implications for selecting appropriate technologies for cancer genome projects as well as for understanding prostate cancer progression.

INTRODUCTION

Prostate cancer is the most common malignancy in men, with over 300 000 diagnoses and 30 000 deaths in the United States alone (1). Underlying factors implicated in disease etiology include diet, ethnicity, genetic predisposition, hormones and other environmental contributors (2).

Most experts agree that genomic instability plays a role during cancer evolution, but its exact significance is debated (3). Two types of genomic instability are evident in most cancers (3): large-scale alterations (e.g. aneuploidy, translocations) and small-scale alterations (e.g. single nucleotide substitutions, microsatellite instability). The extent of small-scale somatic alterations has been mostly probed with PCR-based methods: the majority of tumors display low frequencies of genomic instability (4–6) although a screen of the complete protein kinase gene family (which contains genes that are commonly mutated in tumors) revealed the equivalent of 142 000 alterations per tumor cell (7). PCR-based methods however are bound to underestimate the total number of variations present in heterogeneous genomes because they only record the most common genotypes in a pool of heterogeneous molecules (8). In order to account for tumor heterogeneity one needs to employ cell-specific or single molecule analyses, such as sequencing individual clones.

Single molecule PCR analysis has resulted in the discovery of significant instability in various types of tumors (9), but the target sequence analyzed is a 4-bp fragment in a p53 intron, making it difficult to extrapolate those mutation frequencies for the whole cancer genome. Single molecule methods based on next generation sequencing have identified somatic tumor mutations not detected by Sanger sequencing (10), but the cost per sequencing run is significant, and the relatively high background makes it impossible to identify mutations present in <0.2% of the tumor genome (10). Furthermore, both of the above methods are based on the presumed reduction of heterogeneous tumor samples to single molecules by dilution (before amplification). It is technically much harder to guarantee generation of single DNA molecules by dilution, compared to, e.g. cloning. Accurate measurement of the extent of genomic instability in tumors is important for understanding cancer etiology, disease prognosis and for successful treatment.

We recently reported a high frequency of somatic mutations in several genes in prostate cancer tissue, and

*To whom correspondence should be addressed. Tel: 504-9885418; Fax: 504-9885516; Email: nmakrida@tulane.edu

the presence of a degenerate DNA sequence motif (THEMIS) in the mutation sites (11). The THEMIS motif {WKVnRRRnVWK: W = A/T, K = G/T, V = G/A/C, R = purine (A/G); *n* = any nucleotide; total number of *n* = 0–2 nucleotides; the underline indicates the position of the mutated base}, was found significantly more often than expected at the somatic mutation sites of both prostate and breast cancer tissue, when one mismatched was allowed (11). Here we utilize a PCR-free cloning/sequencing approach to report much higher somatic mutation frequencies than previously measured in cancer tissue, mostly from intergenic regions across the tumor genome.

MATERIALS AND METHODS

Samples

Prostatic adenocarcinoma patients used for the assessment of variation frequencies were of African American or Hispanic American background and were obtained from the Multiethnic Cohort (12). Patient age at diagnosis varied from 55 to 66 years and grading (Gleason score) varied from 6 to 7. One of the patients refused to provide blood DNA. For the assessment of heterogeneity we used Caucasian prostatic adenocarcinoma patients taken from a different dataset (13). Patient age at diagnosis varied from 56- to 72-years old and stage varied from pT2a to pT3c (using the TNM staging system (14)).

DNA isolation

Prostate tissue and blood were collected from each patient. Prostate tissue was microdissected manually. DNA was extracted from the prostate as in ref. (13) and from the blood with standard methods (15).

Shotgun cloning and sequencing

A 100–250 ng genomic DNA from matched lymphocytes, normal prostate and prostate tumor were digested with AluI (NEB, Beverly, MA, USA) and ligated into (SmaI) linearized, dephosphorylated pUC18. The plasmids were transformed into One Shot[®] TOP10 Ultracompetent *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA) and recombinants were identified via blue-white screening. Plasmids were spin-column prepared (Qiagen, Valencia, CA, USA) and DNA sequenced (Applied Biosystems BigDye v3.1, Foster City, CA, USA) using M13F or M13R universal primers (IDT, Coralville, IA, USA). Sequencing reactions were aliquoted in 0.2% SDS, heated to 98°C for 5 min, prior to purification by AutoSeq50 size exclusion chromatography (Amersham Biosciences, Pittsburg, PA, USA). Purified samples were then run on an ABI 3100 automated DNA sequencer.

Genomic heterogeneity assessment

A 50–100 ng of matched prostate tumor and blood DNAs were PCR amplified on a Robocycler thermal cycler (Stratagene, La Jolla, CA) using 2.5 U AmpliTaq Gold DNA polymerase (Roche), 1.5 mM MgCl₂, 0.5 mM (CA)₈RY oligo and 5% v/v DMSO, using the following

PCR protocol: 15 min at 95°C for 1 cycle, followed by 40 cycles of 2 min 20 s at 95°C, 1 min 20 s at 48°C and 2 min 20 s at 72°C. PCR products were AluI-digested and then shotgun cloned and sequenced as described above.

Variant identification

We sequenced 60–100 clones from each blood, tumor and normal prostate sample, bi-directionally. Vector sequences were removed, and each clone pair (forward and reverse sequence) was aligned using BLAST 2 SEQUENCES (www.ncbi.nlm.nih.gov/BLAST/). Mismatches were manually edited based on the sequence information from both strands (to remove obvious base-calling errors). This final edited sequence output was then compared to the nr-database in Genbank using BLASTN (www.ncbi.nlm.nih.gov/BLAST/). Variations with the best match were confirmed by blasting against the genomic reference sequence. Confirmed variations were scored, unless they represented >5% of the total sequence, in which case the clone was rejected from the analysis (for further explanation see the Discussion section).

All confirmed variations were blasted against dbSNP build 129 (http://www.ncbi.nlm.nih.gov/SNP/snp_blast ByOrg.cgi), to identify known polymorphisms.

Statistics

Variation frequency. All statistical analyses comparing variation frequency for the three groups (blood, normal prostate and prostate tumor) were performed using SAS statistical software, version 9.1 (SAS Institute, Cary, NC, USA), and all results were tested at a 5% level of significance. Data were summarized using descriptive statistics such as mean and standard deviations. The analysis of variance (ANOVA) method and Bonferroni multiple comparison test as a *post hoc* analysis were used for comparing means for these three frequency groups.

THEMIS motif. Two-sided *P*-values were calculated for the prostate tumor and blood/normal prostate groups using the chi-squared test.

RESULTS

We measured the frequency of small-scale somatic alterations in sporadic prostate cancer by using a PCR-free cloning and sequencing method (Supplementary Figure S1). This technique was used to identify the number of variants present in randomly cloned genomic pieces of prostate tumor, normal prostate and blood DNAs from the same individual by first sequencing each DNA fragment and then comparing the sequence to the nr (nonredundant)- and refseq_genomic-databases in Genbank (Supplementary Figure S1). Our prostate tumor DNA samples were obtained by microdissecting formalin fixed tissue slides under the microscope, a technique that results in minimum normal DNA contamination of the tumor material, but which may cause an artificial increase of the mutation frequencies (18).

In order to circumvent this potential problem, we also analyzed formalin fixed normal tissue from the same slides, in addition to non-fixed blood DNA from the same patients. We purposely chose to analyze relatively advanced prostate tumors (Gleason score 6–7), reasoning that failure to identify an increased alteration frequency by this point in tumor evolution may indicate that genomic instability does not play a major role in prostate cancer progression.

After sequencing 100 kb of total DNA from blood lymphocytes, normal prostates and tumor prostates taken from six men with prostate cancer, we identified 46 variations in the tumors, 30 in the normal prostates and 26 in the blood DNAs. Each of these variations was confirmed in independent sequencing reactions using both forward and reverse primers. Excluding variations found in dbSNP build 129 (www.ncbi.nlm.nih.gov/SNP/; presumed to be polymorphisms) resulted in 20 variations remaining in the tumors, 5 in normal prostates and 4 in the blood DNA. The only statistically significant differences were between the mean tumor and normal prostate DNA variation frequency, and the mean tumor and blood DNA variation frequency (Figure 1A; see 'Materials and Methods' section). Although our analysis is based on distinct sequences from each sample type (normal prostate, tumor and blood), the randomness of the cloning methodology and of the analyzed data (evidenced by e.g. G/C content and chromosomal distribution; Supplementary Figure S2) makes it possible to measure variation frequencies across each type of genome. Subtracting genomic variation frequencies between tumor and normal prostate can thus estimate the frequency of small-scale somatic events in the tumor. The difference between tumor and normal prostate variation frequencies shown in Figure 1A corresponds to 1.8 (± 1.0) million somatic alterations per prostate tumor cell genome. The relatively high standard deviation can be attributed to the existence of both tumors with higher (2.6 ± 0.6 million) and lower (1.0 ± 0.36 million) genomic instability frequencies (Supplementary Figure S3), a finding evident by the distribution of the variation frequency per sample (Figure 1B). The absolute number of variations and base pairs sequenced per sample is presented in Supplementary Table S1, along with the sequence surrounding each variation.

Distinct etiologies may result in specific types of variations in the tumor or in a specific DNA sequence context. Thus, we decided to test these variations for the presence of a specific motif we have previously found to be overrepresented at the sites of tumor DNA mutations in several genes (THEMIS motif; 11). We report that variations in the tumor DNA occur in the context of the THEMIS motif (with one mismatch allowed) significantly more often (78%) than in the normal prostate and blood DNA (33%; two-sided $P = 0.046$) (Figure 2A), suggesting distinct molecular etiology in the tumors. In contrast, the distribution of the variations by substitution type did not vary significantly between tumor and controls (Figure 2B).

The methodology we utilized for estimating variation frequencies was based on the premise that prostate

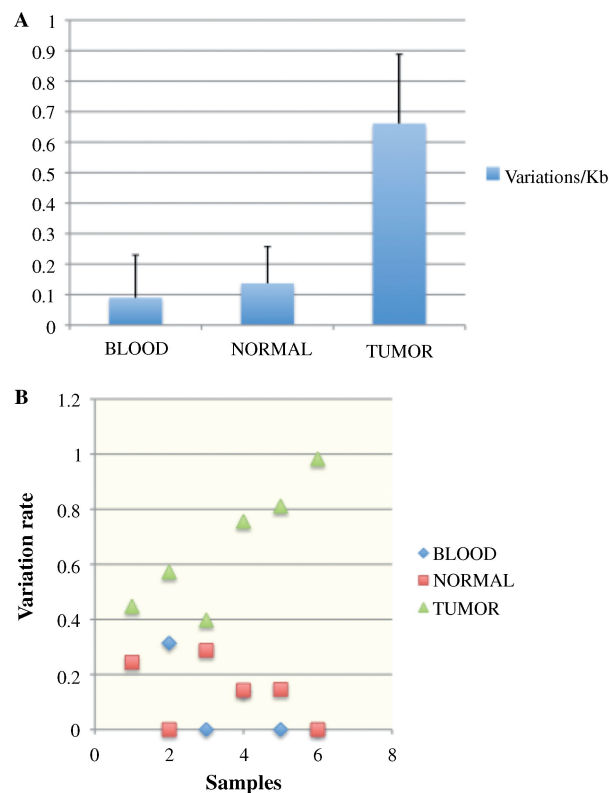


Figure 1. Genomic variation rate is significantly higher in prostate tumors. Variation rates were calculated by our cloning/sequencing protocol (see 'Materials and Methods' section). Variations found in dbSNP build 129 were excluded. (A) Mean genomic variation rate. Mean variation rate = mean number of variations/Kb sequenced. The mean genomic variation rate is significantly different between prostate tumor and the other two tissue types (normal prostate and blood; $P < 0.05$). (B) Individual genomic variation rate in six prostate cancer patients. Variation rate = number of variations/Kb sequenced. Normal, normal prostate; Tumor, prostate tumor. When not visible, the blood rate is identical to the normal prostate rate, except for sample 1, for which no blood was available.

tumors may be genetically heterogeneous. Although the results presented in Figure 1A suggest a significant frequency of genomic instability in prostate tumors, they do not address the issue of cancer heterogeneity. In order to probe the existence of heterogeneity in prostate cancer, we employed the same methodology that we used to measure variation frequencies, but with the addition of a degenerate PCR step before cloning (see 'Materials and Methods' section). This modification resulted in the generation of multiple individual sequences from a small number (10–15) of the same genomic regions for both tumor and control DNA, after sequencing 60–100 clones per sample. The first prostate tumor sample we analyzed showed extensive heterogeneity in one of those genomic regions, specifically in the tumor (Figure 3). A different tumor sample showed a lower degree of heterogeneity (two variations in each of three genomic regions, one of these regions being the same region shown in Figure 3). Thus, we conclude that there is significant but variable degree of genetic heterogeneity in prostate tumor genomes. Likewise, prostate tumors show a variable degree of genomic instability (Figure 1B).

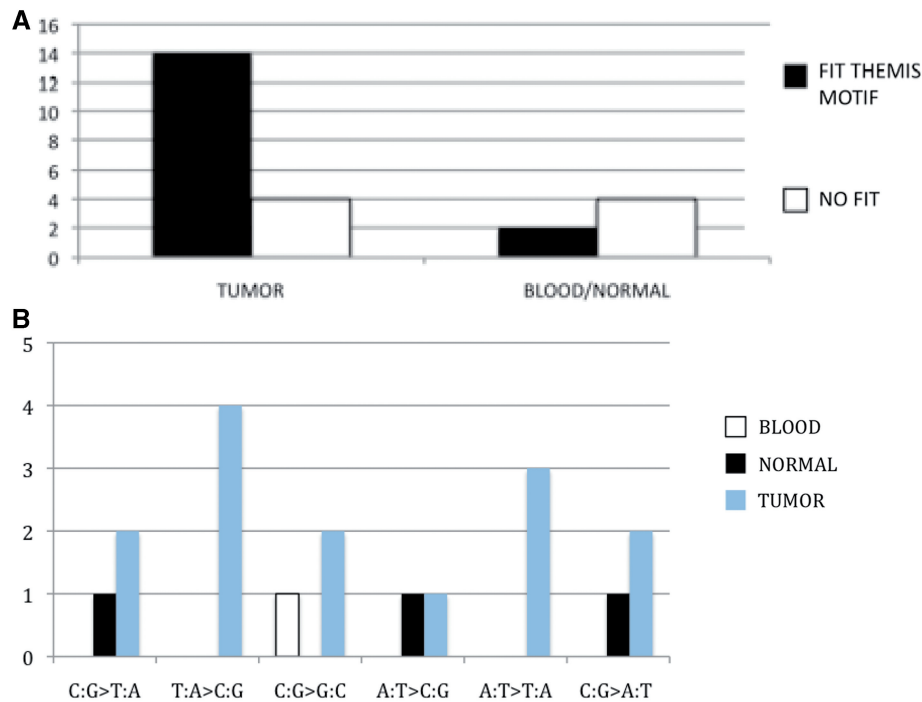


Figure 2. Distribution of the variations found by (A) the presence of the THEMIS motif and (B) substitution type. Variations found in dbSNP were excluded. The data presented in (A) show the number of variations that occur (or not) in the context of the THEMIS motif with one mismatch allowed. For a more complete explanation of the THEMIS motif, see text. Only the differences in (A) between the distribution of the variations in the tumor and the other two types of tissue are significant ($P = 0.046$).

T2 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T51 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T41 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T38 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T46 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T4 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T3 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

T37, T47, T48 tttacatggagtggttattagtggaagacaatggtgataaattgcttttaacacac
 tgctcccgggtgctgctctaagcactcccctcctccggatctctgtgaggatgacctcacttc

Figure 3. Genetic heterogeneity in prostate cancer. Data were obtained using the same cloning/sequencing protocol utilized for measuring genomic variation rates, with the inclusion of a degenerate PCR step prior to cloning (see 'Materials and Methods' section). All clones come from the same tumor sample. Numbers indicate clone number (e.g. T2 = tumor clone #2). The three clones grouped together in the bottom contain wild-type sequence. Variations are capitalized and shown in highlight. Only the minimum sequence needed to show the maximum degree of heterogeneity is shown. These variations are not present in dbSNP build 129.

DISCUSSION

The average number of somatic events reported here in six prostate cancers is orders of magnitude higher than previous reports (4–9). Some of those studies analyzed coding sequences (4–7), while others noncoding (8,9).

Thus, selection of coding region mutations during tumor evolution cannot be responsible for this difference in variation frequencies. In contrast, either the tumor type (prostate) or the PCR-free methodology we used may be responsible for our finding. PCR-based analysis of various tumors has revealed less extensive but significant tumor-to-tumor mutation and mutated 'driver' gene heterogeneity (4–7). Most of the mutations reported here are in intergenic regions and thus presumably 'passengers' (i.e. not tumor 'drivers'; 16). If distinct cells of the same tumor contain different passengers, then they may also contain different drivers. This possibility can only be assessed by single cell/molecule methods such as the one we employed here. Before embarking on full-scale cancer genome projects with techniques that may miss most mutations in heterogeneous tumors it may be worthwhile first to use single cell/molecule approaches to classify tumors based on their degree of heterogeneity. Then the appropriate method can be employed to the full project based on tumor type (heterogeneous or not).

We report here an average tumor mutation frequency that significantly exceeds the current estimates (4–9). We believe that this finding has important biological consequences for prostate tumorigenesis. In addition to finding significantly higher mutation frequency in tumor DNA, we observed that prostate tumors contain distinct types of mutations: unlike normal prostate and blood from the same patient, tumor mutations fit the THEMIS motif (11), suggesting a common etiology. We do not know the exact molecular mechanism generating these

mutations in the prostate tumor genome, but based on our previous analysis of the THEMIS motif mutations in several genes in prostate cancer tissue (11), this genomic instability in prostate cancer may be the result of aberrant DNA repair in the prostate.

In performing our genomic variation analysis, all of the variations found from the best aligned match in Genbank were scored, unless the number of variations represented >5% of the total insert length, in which case the clone sequence was rejected from the analysis (the level of genetic diversity among different normal individuals rarely approaches 1% of the total sequence (17), so we reasoned that more than 5-fold higher would be unlikely even for the tumor DNA, and it probably represents a sequencing artifact (either on our side or the Genbank reference sequence). Using this criterion, five clones were rejected from the analysis of the tumor DNA, six from the normal prostate and seven from the blood DNA. All but one of the rejected clones represented non-unique sequence or included repeats (commonly centromeric DNA). The best alignment for the remaining clone is shown in Supplementary Figure 4, and may represent template-independent DNA synthesis performed by the tumor (18), or a fragment not yet sequenced by the various human genome projects. We favor the former explanation, since we were unable to PCR amplify this clone specific fragment from all normal unrelated DNA samples we examined (data not shown).

Variants in repeats are more likely to be uncharacterized polymorphisms (though not yet reported in dbSNP) and thus may increase the total count for both control and tumor DNAs. However, Supplementary Figure S5 demonstrates that the excess of tumor variations (especially substitutions) can not be merely explained by repeat variations. Thus we decided to include the repeat variants in our analysis. Subtracting the repeat variations from the total count changes the frequency estimate to 1.1 (± 0.7) million somatic alterations per prostate tumor cell genome.

The distribution of the variations by type (Supplementary Figure S5) shows that unlike normal prostate and blood, prostate tumors contain mostly substitutions, not deletions or insertions. However, the significance of this finding is limited to small-scale tumor alterations, because the utilization of a four base cutter (AluI) in our cloning methodology does not allow us to adequately examine large-scale genomic instability (translocations, aneuploidy, etc.).

The genomic analysis reported here demonstrates that normal prostate DNA does not show increased genomic instability compared to blood (Figure 1A), suggesting that fixation artifacts (previously reported to artificially increase mutation rates in formalin fixed samples; 19) do not play a major role in the data that we report here. The utilization of a PCR-free approach may be responsible for this finding, since *Taq* DNA polymerase is thought to be the cause of these artifacts (19). Most of the observed variations in normal tissue (both prostate and blood) are probably yet-unidentified polymorphisms. Some of the tumor variations are also likely to be yet-unidentified polymorphisms. This is a major reason we subtracted the

variation frequency of the normal from the tumor tissue, in order to measure the somatic variation frequency in tumors. Moreover, Figure 1B demonstrates that all patients show an increase in the variation frequency specifically in the tumor genome (this trend is not significant for individual patients, probably due to the low number of variations per patient).

In summary, a PCR-free methodology that measures genomic variation rate reveals substantial somatic instability in prostate cancer. This finding suggests at least a partial role for the mutator phenotype hypothesis (8) in prostate tumor progression.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Fabricio Rojas, Chris Haiman and Susan Roberts for technical support. We thank Laurence Kolonel and Brian Henderson for providing the tissue samples.

FUNDING

National Cancer Institute (P01 CA108964 project 1 to J.K.V.R.); the National Center for Research Resources (P20 RR020152-01 to N.M.M.) J.K.V.R. is a Medical Foundation Fellow at the University of Sydney. Funding for open access charge: National Institutes of Health grant.

Conflict of interest statement. None declared.

REFERENCES

- Jemal,A., Siegel,R., Ward,E., Hao,Y., Xu,J., Murray,T. and Thun,M.J. (2008) Cancer statistics, 2008. *CA Cancer J. Clin.*, **58**, 71–96.
- Bostwick,D.G., Burke,H.B., Djakiew,D., Euling,S., Ho,S.M., Landolph,J., Morrison,H., Sonawane,B., Shifflett,T., Waters,D.J. *et al.* (2004) Human prostate cancer risk factors. *Cancer*, **101**, 2371–2490.
- Marx,J. (2002) Debate surges over the origins of genomic defects in cancer. *Science*, **297**, 544.
- Greenman,C., Stephens,P., Smith,R., Dalgliesh,G.L., Hunter,C., Bignell,G., Davies,H., Teague,J., Butler,A., Stevens,C. *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153–158.
- Wood,L.D., Parsons,D.W., Jones,S., Lin,J., Sjöblom,T., Leary,R.J., Shen,D., Boca,S.M., Barber,T., Ptak,J. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.
- The Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
- Stephens,P., Edkins,S., Davies,H., Greenman,C., Cox,C., Hunter,C., Bignell,G., Teague,J., Smith,R., Stevens,C. *et al.* (2005) A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat. Genet.*, **37**, 590–592.
- Loeb,L.A., Loeb,K.R. and Anderson,J.P. (2003) Multiple mutations and cancer. *Proc. Natl Acad. Sci. USA*, **100**, 776–781.
- Bielas,J.H., Loeb,K.R., Rubin,B.P., True,L.D. and Loeb,L.A. (2006) Human cancers express a mutator phenotype. *Proc. Natl Acad. Sci. USA*, **103**, 18238–18242.

10. Thomas, R.K., Nickerson, E., Simons, J.F., Jänne, P.A., Tengs, T., Yuza, Y., Garraway, L.A., LaFramboise, T., Lee, J.C. and Shah, K. (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat. Med.*, **12**, 852–855.
11. Makridakis, N.M., Ferraz, L. and Reichardt, J.K.V. (2009) Genomic analysis of cancer tissue reveals that somatic mutations commonly occur in a specific motif. *Hum. Mutat.*, **30**, 39–48.
12. Kolonel, L., Henderson, B.E., Hankin, J.H., Nomura, A.M., Wilkens, L.R., Pike, M.C., Stram, D.O., Monroe, K.R., Earle, M.E. and Nagamine, F.S. (2000) A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am. J. Epidemiol.*, **151**, 346–357.
13. Akalu, A., Dlmajian, D.E., Highshaw, R.A., Nichols, P.W. and Reichardt, J.K.V. (1999) Somatic mutations at the SRD5A2 locus encoding prostatic steroid 5 α -reductase during prostate cancer progression. *J. Urol.*, **161**, 1355–1358.
14. Schroder, F.H., Hermanek, P., Denis, L., Fair, D.R., Gospodarowicz, M.K. and Pavone-Maculoso, M. (1992) The TNM classification of prostate cancer. *Prostate Suppl.*, **4**, 129–138.
15. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1995), *Current Protocols in Molecular Biology*. Wiley & Sons, New York, pp. 2.2.1–2.2.3.
16. Parmigiani, G., Boca, S., Lin, J., Kinzler, K.W., Velculescu, V. and Vogelstein, B. (2009) Design and analysis issues in genome-wide somatic mutation studies of cancer. *Genomics*, **93**, 17–21.
17. Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L. and Lander, E.S. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.
18. Bignell, G.R., Santarius, T., Pole, J.C., Butler, A.P., Perry, J., Pleasance, E., Greenman, C., Menzies, A., Taylor, S., Edkins, S. *et al.* (2007) Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res.*, **17**, 1296–1303.
19. Quach, N., Goodman, M.F. and Shibata, D. (2004) In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clin. Pathol.*, **4**, 1.