Journal of Pathology
J Pathol 2008; **216**: 307–316
Published online 29 July 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/path.2413

Original Paper

Microdissection molecular copy-number counting (μMCC) — unlocking cancer archives with digital PCR

F McCaughan,[1,2] E Darai-Ramqvist,[6] AT Bankier,[2] BA Konfortov,[2] N Foster,[3] PJ George,[1,4] TH Rabbitts,[5] M Kost-Alimova,[6] PH Rabbitts[5] and PH Dear*[2]

[1]*Centre for Respiratory Research, Department of Medicine, Royal Free and University College Medical School, The Rayne Institute, London WC1E 6JJ, UK*
[2]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*
[3]*Clinical Cytogenetics, Human Genetics Unit, Ninewells Hospital, Dundee DD1 9SY, UK*
[4]*Department of Thoracic Medicine, University College London Hospitals, Grafton Way, London WC1E 6AU, UK*
[5]*Leeds Institute of Molecular Medicine, Section of Experimental Therapeutics, Wellcome Trust Brenner Building, St James's University Hospital, Leeds LS9 7TF, UK*
[6]*Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm SE-171 77, Sweden*

**Correspondence to:*
PH Dear, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.
E-mail: phd@mrc-lmb.cam.ac.uk

Conflict of interest. *A patent for molecular copy-number counting has been applied for by the UK Medical Research Council. Authors AT Bankier, PH Dear, BA Konfortov and TH Rabbitts are named on the patent application.*

Received: 8 May 2008
Revised: 24 June 2008
Accepted: 15 July 2008

Abstract

Most cancer genomes are characterized by the gain or loss of copies of some sequences through deletion, amplification or unbalanced translocations. Delineating and quantifying these changes is important in understanding the initiation and progression of cancer, in identifying novel therapeutic targets, and in the diagnosis and prognosis of individual patients. Conventional methods for measuring copy-number are limited in their ability to analyse large numbers of loci, in their dynamic range and accuracy, or in their ability to analyse small or degraded samples. This latter limitation makes it difficult to access the wealth of fixed, archived material present in clinical collections, and also impairs our ability to analyse small numbers of selected cells from biopsies. Molecular copy-number counting (MCC), a digital PCR technique, has been used to delineate a non-reciprocal translocation using good quality DNA from a renal carcinoma cell line. We now demonstrate μMCC, an adaptation of MCC which allows the precise assessment of copy number variation over a significant dynamic range, in template DNA extracted from formalin-fixed paraffin-embedded clinical biopsies. Further, μMCC can accurately measure copy number variation at multiple loci, even when applied to picogram quantities of grossly degraded DNA extracted after laser capture microdissection of fixed specimens. Finally, we demonstrate the power of μMCC to precisely interrogate cancer genomes, in a way not currently feasible with other methodologies, by defining the position of a junction between an amplified and non-amplified genomic segment in a bronchial carcinoma. This has tremendous potential for the exploitation of archived resources for high-resolution targeted cancer genomics and in the future for interrogating multiple loci in cancer diagnostics or prognostics.

Keywords: molecular copy-number counting; cancer genomics; formalin-fixed paraffin-embedded; laser capture microdissection; lung cancer

Introduction

It has long been appreciated that the genome is grossly abnormal in cancer, being characterized by major aberrations, such as deletions, amplifications and translocations [1,2]. Established cancers of the same organ and cell type in different patients often share similar genomic changes or signatures, a finding which has directed efforts towards defining the key 'driver' genes that are amplified, lost or altered [3,4].

The vast majority of cancer specimens that could be exploited to investigate these genomic events are formalin-fixed paraffin-embedded (FFPE) archived clinical biopsies. However, the assessment of archived FFPE specimens using existing methods presents a challenge. The quantity of DNA may be severely limited, due to the small size of some clinical biopsies. Further, many tumours are heterogeneous, necessitating microdissection to isolate a pure malignant cell population, further reducing the quantity of DNA available [5]. Moreover, the duration of both formalin fixation [6] and archiving [7] has a profound effect on the quality and integrity of DNA extracted from these specimens. This leads to a much reduced effective yield from biopsies and to the DNA often being inadequate for standard downstream molecular

biology applications, including the polymerase chain reaction (PCR) [6,8–10].

Many technologies are available to assess copy number alteration in cancer and their relative merits have recently been reviewed [11]. Quantitative PCR and fluorescence *in situ* hybridization (FISH) are limited by the number of loci that can be assessed, whilst FISH and metaphase comparative genomic hybridization (CGH) are limited in their resolution. Array experiments can assess copy number at high resolution across the entire genome, and have transformed our knowledge of structural variation in cancer genomes [5,12] (http://www.sanger.ac.uk/genetics/CGP). However, they are prone to experimental bias, such as variations in hybridization efficiency and specificity [5]. Raw array data, particularly from FFPE biopsies, is usually noisy, requiring normalization, smoothing and segmentation to reveal copy-number changes [5,13,14]. Critically, array-based methods often fail to produce results when a limited amount of DNA is available, particularly if it is of poor quality, as is often the case with fixed archival clinical biopsy material [5,8,9]. To overcome the problem of limited template DNA, array protocols often include a whole-genome amplification (WGA) step [9,15], but this may introduce representational bias, particularly of GC-rich regions [15]. Furthermore, the efficacy of WGA can be reduced by DNA cross-linking and degradation as a result of formalin fixation, causing a high proportion of such samples to fail [8,9].

A technical advance would therefore have the ability to assess copy-number changes over a wide dynamic range at multiple loci in a single experiment, starting from minute quantities of grossly degraded DNA and avoiding the use of a WGA step. Molecular copy-number counting (MCC; Figure 1) has previously been presented as a simple digital PCR technique for defining a non-reciprocal translocation by finding the associated two-fold transition in copy number, exemplified in a renal cell carcinoma cell-line [16]. In this paper, we demonstrate the utility of MCC and of a modified protocol (μMCC) to enable analysis of minute, degraded samples in a series of progressively more challenging contexts. First, we demonstrate the capacity of MCC to distinguish not only the 2 : 1 variation seen in non-reciprocal translocations, but also the wider copy number variation seen in a nasopharyngeal epithelial cell line. Second, we show that the modified μMCC protocol can derive reproducible data on regional genomic amplification, using grossly fragmented DNA extracted from formalin-fixed paraffin-embedded biopsy specimens. Third, we show that reproducible results can be obtained from picogram quantities of laser-captured material from 7 year-old archived FFPE specimens. Finally, we demonstrate the flexibility and potency of μMCC by using it to precisely resolve an amplicon border — the transition between low and high copy — and to resolve internal variation in copy number within the amplicon. Use of μMCC should allow researchers to readily interrogate somatic aberrations in cancer genomes in existing clinical and research archives.

Materials and methods

Patients and samples

All samples were from patients enrolled in the University College London Hospital Early Lung Cancer Project. Ethical approval was obtained for the study (UCLH MREC 01/0148) [17]. The protocols for DNA extraction, including from microdissected epithelium, are available in Appendix A (see Supporting Information). HONE-1 is a nasopharyngeal cell line [18]. DNA was extracted as described previously [19].

Markers and PCR primers (Table 1)

Repeat-masked *Homo sapiens* genomic sequence was downloaded from the Ensembl ftp site (ftp://ftp.ensembl.org/pub/current_homo_sapiens/data/fasta/dna/). NCBI Build 35 was used for primer sets 1 and 2; Build 36 for other primers. Hemi-nested primer sets were designed using custom software (PHD, unpublished). Primer parameters for MCC were as described previously [16], and external amplimer lengths (between the forward-external and reverse primers) were typically 100–300 bp. For μMCC, we only selected amplimers falling within a narrowly defined size range. Optimum internal amplimer length was 80–95 bp with external amplimer length of 100–120 bp. Typical oligonucleotide length was 18–23 bp and calculated melting temperature

Table 1. Characteristics of primer sets used. Further details of primer sets (specific loci and sequences) are in Supporting Information, Appendix B

Primer set	Characteristics
1	80 markers at regular intervals (average 2.3 Mb) along chromosome 3. Set also has five reference markers on chromosomes 1, 2, 6, 10 (multicopy) and 11. Phase 1 external amplimers vary in length from 113 to 271 bp
2	95 markers at regular intervals (average 189 kb) along 3q between 165.04 and 182.96 Mb. Phase 1 external amplimers vary in length from 104 to 277
1.1	81 markers at regular intervals (average 2.3 Mb) along chromosome 3. Set also has five reference markers on chromosomes 1, 2, 6 (×2), 10 (multicopy) and 11. All phase 1 external amplimers are in the range 100–120 bp
3	21 markers along chromosome 3 between 165.95 and 170.87 Mb. All phase 1 external amplimers are in the range 100–120 bp
4	21 markers along chromosome 3 between 166.35 and 170.87 Mb. All phase 1 external amplimers are in the range 100–120 bp

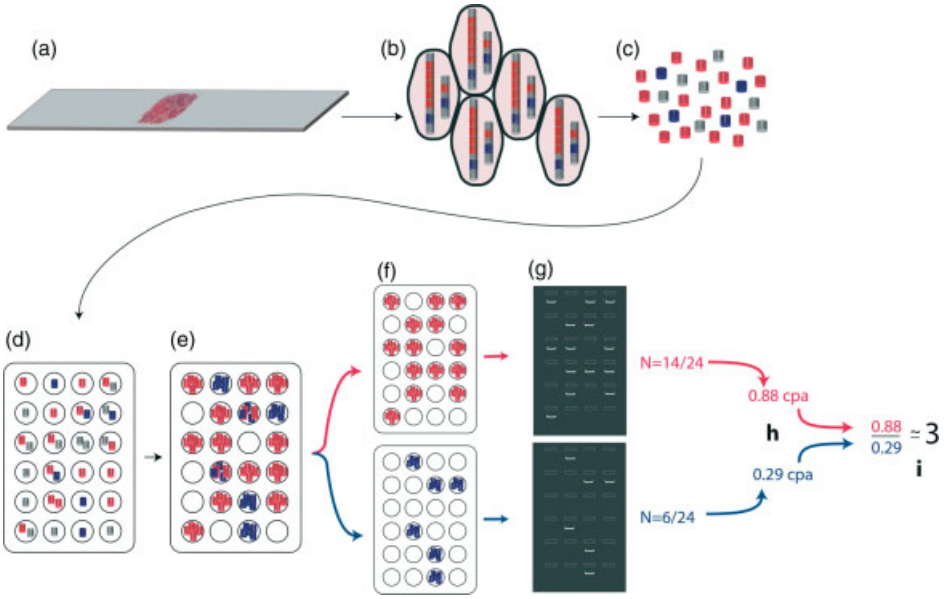

Figure 1. Principles of μMCC. Starting with a fixed, stained biopsy section (a), laser-capture microdissection is used to harvest cells (b). In this example, the blue band represents a reference locus present at two copies/cell; the red band represents a locus which has undergone amplification and is present at six copies/cell. Genomic DNA (c) is extracted and aliquoted into the wells of a microtitre plate (d), with about 2 pg DNA (about 0.7 haploid genomes)/well (for simplicity, only 24 wells are shown). A multiplex PCR is used to amplify both the red and blue sequences simultaneously (e). The products of this reaction are split into replica plates, one of which is amplified in a monoplex PCR for the red sequence (f, upper) and the other for the blue sequence (f, lower). PCR products are analysed by either gel (g) or melting-curve (not shown). In this example, the 'blue' (reference) locus is detected in 6/24 wells and the 'red' locus in 14/24 wells. A Poisson calculation converts these numbers into copies per aliquot (cpa; h), and the ratio between these gives the relative copy number of the amplified locus (in this case, three copies/haploid genome, or six copies/cell, or three times normal copy (i)

$[T_m; 2x(A+T)+4x(G+C)]$ was 55–62 °C. Primers were tested by *in silico* PCR, mainly using the National Centre for Biotechnology Information electronic PCR [20]. In general, primer sets where the forward-internal and reverse primers gave more than one predicted product (settings: two mismatches and two gaps) were discarded, with the exception of primer sets deliberately designed against known multicopy sequences. All primers were supplied by Operon Biotechnologies Gmbh.

MCC, μMCC and dilution assays

MCC (Figure 1) has been described in detail [16]; μMCC differs from MCC not only in the use of laser microdissection to obtain a pure population of cancer cells, but also in the selection parameters for PCR primers (above). Briefly, extracted DNA is diluted to approximately 0.4 pg/μl and 88 5 μl aliquots (each containing approximately 2 pg or 0.7 haploid genomes of DNA, plus eight 5 μl aliquots water) are dispensed into a microtitre plate. All aliquots are tested using PCR to score the presence or absence of each of the marker sequences. PCRs consist of an initial multiplex PCR, using the forward-external and reverse primers for all markers in a single reaction. Standard thermocycling conditions were: 93 °C for 9 min, followed by 28 cycles of 20 s at 94 °C, 30 s at 50 °C and 1 min at 72 °C. Products were diluted and used as template in monoplex PCRs with the forward-internal and reverse primer for each marker in turn. Standard thermocycling conditions for phase 2 were: 93 °C for 9 min, followed by 33 cycles of 20 s at 94 °C, 30 s at 56 °C and 1 min at 72 °C. Results were scored either by electrophoresis on 96-well 6% polyacrylamide gels, or by melting-curve analysis with SyBrGreenI (Invitrogen, UK) or EvaGreen (Biotium Inc, USA) dye included in the monoplex PCR reactions.

The correct template dilution can in principle be determined prior to μMCC by a dilution assay in which 16 aliquots at each of six serial two-fold DNA dilutions are tested for four markers, using the same protocol as for μMCC. The dilution returning approximately 8/16 positive aliquots for each of the markers (corresponding to about 0.7 haploid genomes/aliquot) is used.

Analysis of results

From the proportion (P) of aliquots (out of 88, excluding the eight negative controls) scoring positive for a given marker, the average number of copies per aliquot (cpa) is calculated using the Poisson equation ($cpa = -\ln(1-P)$; see reference 15). cpa values for each marker are normalized against the average of reference loci believed to be present at normal copy (ie, one copy per haploid genome) to give the absolute number of copies of each marker sequence per haploid genome. Reference loci used in these experiments were chosen on the basis of previous CGH experiments on the same clinical specimens [21]

and because these loci were within regions generally at normal copy in squamous lung cancer specimens assessed by CGH [22].

Fluorescence *in situ* hybridization (FISH)

Standard methods were used and the protocol is described in Appendix A.

Results

MCC can be used to accurately assess regional genomic copy number variation over a wide dynamic range

MCC uses a digital PCR approach to assess the relative copy number of genomic markers (loci) of interest. The results presented reflect the variation in copy number as assessed by the number of aliquots which were positive for each marker. In most of the presented experiments, DNA was diluted to give approximately 0.7 haploid genomes (about 2 pg) per aliquot. In this case, a 'normal' copy marker (present once per haploid genome) will score positive in about 50% of the aliquots (Poisson distribution). A marker lost from one chromosomal copy (present once per cell, or 0.5 copies per haploid genome) will score positive in about 30% of the aliquots, whereas a 'double copy' marker (2 copies per haploid genome) will be present in about 75% of the aliquots, a 'triple copy' marker in about 88% of the aliquots, etc. We first wanted to demonstrate that MCC can distinguish copy number variation over a wide dynamic range, in a nasopharyngeal epithelial cell line (HONE1) known to bear a regional amplification on chromosome 3 [19,23]. We designed hemi-nested primer sets at 88 loci along chromosome 3 at approximately 2.2 megabase (Mb) intervals. We also chose 'reference loci' on other chromosomes, and two multicopy sequences present at more than 10 copies in the reference genome sequence (see Materials and methods; see also Supporting Information, Appendix B).

Primer set 1 was used in a standard MCC protocol [16] to assess the copy number of these loci both on normal genomic DNA from peripheral blood lymphocytes and on *HONE-1* DNA. On normal DNA, all loci showed similar copy number, with the exception of the known multicopy markers (Figure 2a). MCC (like most methods) measures relative rather than absolute copy-numbers. The raw data give the average number of copies per aliquot in the MCC panel, and these values must be normalized against a value for 'normal' copy taken either from reference loci or (depending on the context) from the 'background' majority of markers in a region believed to be punctuated by local aberrations. Analysis of *HONE-1* DNA using the same primer set revealed a clear amplicon in distal 3q (Figure 2b).

We then designed a second primer set (primer set 2, Appendix B) to amplify loci at approximately 125 kb intervals within and adjacent to the amplicon delineated by primer set 1. MCC using this primer set revealed a complex internal structure to the amplicon, with 1.5–4-fold amplification relative to 'normal' copy (Figure 2c). Reproducibility is illustrated by comparing results from primer sets 1 and 2 over the same region. The MCC results are almost completely concordant with the integer copy number assignments obtained by FISH analysis (Figure 2c). This detailed assessment of intra-amplicon variation by MCC detected a second intra-amplicon peak, missed using FISH (Figure 2c, a four-fold amplification at approximately 180 Mb). These experiments also demonstrate that multiple loci can be interrogated accurately in one experiment, using minute quantities of DNA. Each experiment used approximately 60–70 haploid genome equivalents (180–210 pg) of DNA, equating to about 30–35 diploid cells or about 15–20 of the hypotetraploid HONE-1 cells [19,23].

Analysis of degraded DNA from fixed specimens using MCC and μMCC

We next evaluated the performance of MCC on the degraded DNA typically recovered from FFPE biopsy specimens. Two lung biopsies (lesions A and B) from the same lung-cancer patient were used, dating from 1998 and 1999, respectively — they had previously been analysed at low resolution using metaphase CGH to reveal amplification of part of 3q [21].

MCC analysis on archived DNA from lesion A showed significant marker-to-marker variation (Figure 3a) and was disappointing when compared to those obtained from lymphocyte or cell line DNA (Figure 2). Even the reference markers showed extreme variation in apparent copy-number, making the setting of the reference level arbitrary. Nevertheless, the results appeared to show a region of amplification on 3q and partial loss of 3p. Higher resolution analysis of the 3q amplicon using primer set 2 showed pronounced and chaotic marker-to-marker variation (Figure 3b), which we felt was biologically implausible and likely to be artefactual.

Given that the amplimer lengths of the markers varied considerably (103–273 bp), we postulated that the template DNA was so degraded that longer amplimers were likely to be underrepresented, since they would suffer more from instances of DNA breakage within the target amplicon. Comparison of the apparent copy number of each marker in primer set 2 with its amplimer length (Figure 3c) shows that this is indeed the case, and analysis of the results suggests that the DNA has either a break or a polymerase-blocking modification about every 50 bp. Gel electrophoresis of double-stranded DNA from another part of the same surgical specimen showed an average fragment size of about 300–400 bp (not shown), but this does not take into account single-stranded nicks, cross-links or other modifications that would block amplification [8]. Further experiments showed that, when loci in the

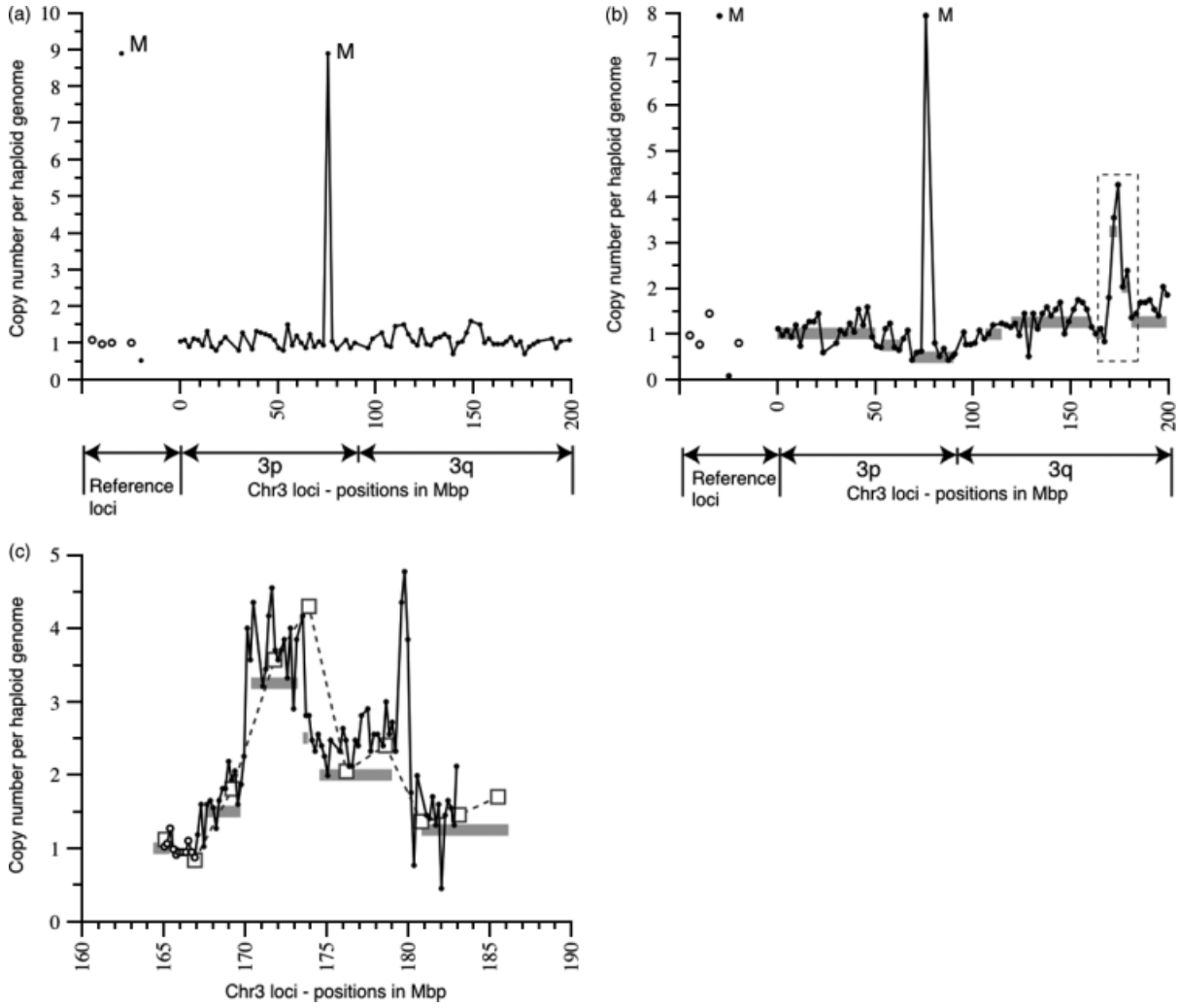

Figure 2. Assessment of regional genomic amplification. (a) MCC results from normal genomic DNA extracted from peripheral blood leukocytes using primer set 1, consisting of markers at approximately 2 Mb intervals across chromosome 3 and reference markers located on other chromosomes. Results were normalized against the average of four reference loci (open circles; there was one failed reference locus). M indicates known multicopy sequences. (b) MCC of DNA extracted from HONE1 cells. Normal copy is based on the average of four reference loci (open circles; one failed reference locus, and two reference loci from the X-chromosome were not used as references in this male-derived DNA). The dashed rectangle on 3q indicates a region of amplification. The grey bars indicate the copy number (per haploid genome) inferred from FISH analysis of the same cell line. (c) High-resolution MCC of the HONE1 3q amplicon using primer set 2, containing loci at approximately 189 kb intervals (circles and solid line), as well as selected loci from primer set 1 (squares and dashed line). In this case, normal copy is based on 3q loci known, from the results shown in (b), to lie to the left of the amplicon (open circles). The grey bars again represent copy number (per haploid genome) inferred from FISH data. The second intra-amplicon peak at approximately 180 Mb was not detected by FISH

lesion-derived DNA were queried in parallel experiments using either longer (106–120 bp) or shorter (65–73 bp) amplimers, the shorter amplimers gave a much higher apparent copy number; this was not true of the less damaged DNA extracted from fresh, frozen tissue or recently-fixed biopsies (data not shown).

We therefore designed a new set of primers (primer set 1.1) covering almost precisely the same chromosome 3-wide loci as primer set 1, but with shorter and more uniform (100–120 bp) amplimer length. The results from lesion A DNA using this primer set (Figure 3d) showed much less chaotic variation and more clearly revealed a biologically plausible amplicon on 3q and partial loss of 3p, consistent with previous CGH results [21], but affording far higher resolution. Primer set 1.1 was also applied to pooled normal DNA from 18 individuals and produced a flat profile, with each marker giving a very similar copy number except for the known multicopy markers (Figure 3e). Indeed, marker-to-marker variation was even less than that seen when the same DNA was analysed using primer set 1, suggesting that standardization of amplimer length may be advantageous even when analysing DNA of relatively good quality.

μMCC applied to microdissected FFPE biopsy samples

Having modified the MCC protocol to cope with degraded templates, we tested the μMCC protocol

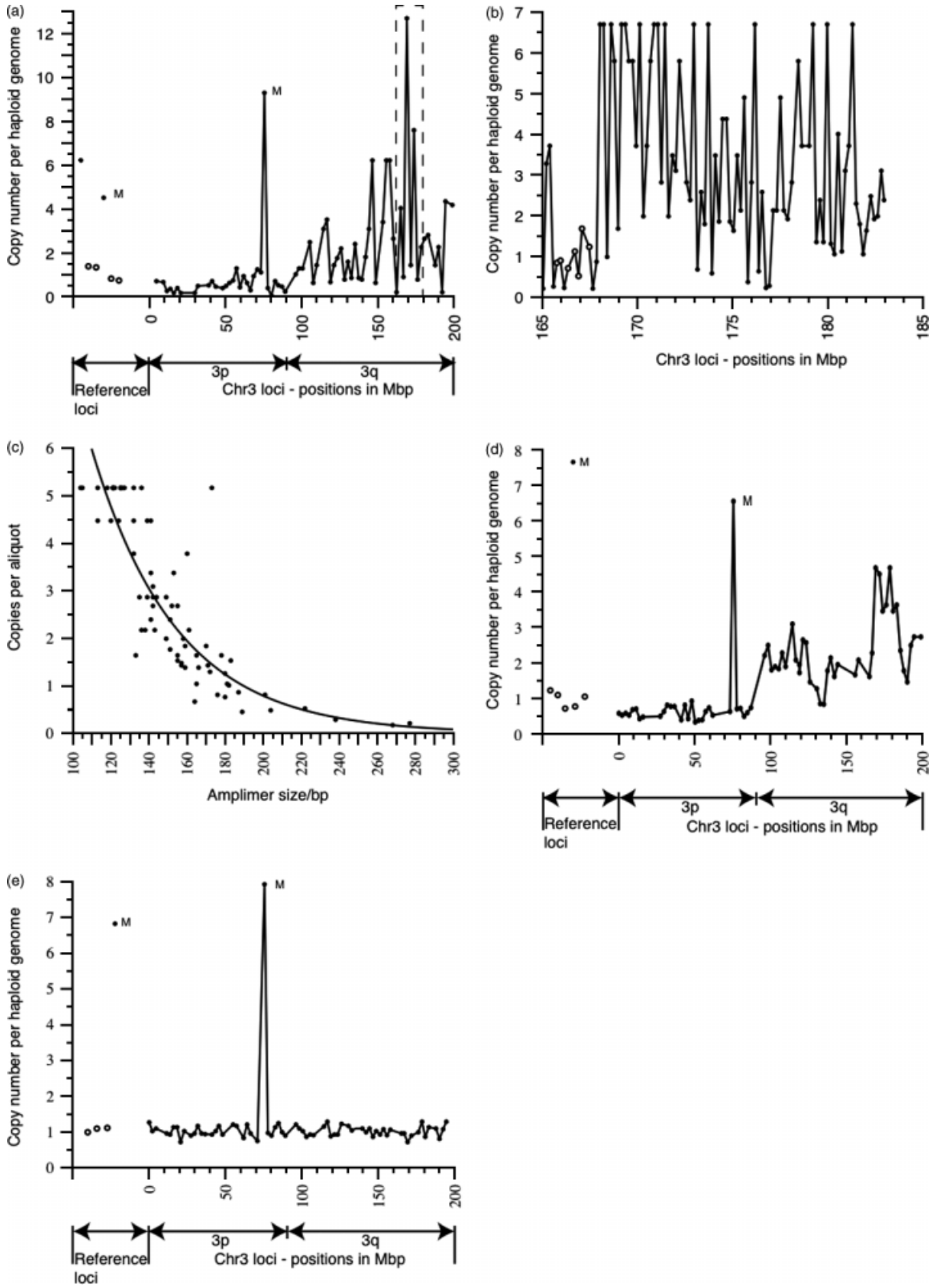

Figure 3. Effect of amplimer length variation. (a) MCC of lesion A biopsy-derived DNA using primer set 1, showing apparently random variation in copy number but also indicating regional amplification of part of 3q. Normal copy was taken from the average of four reference loci (open circles). (b) High-resolution analysis of the same DNA using primer set 2, covering the region bounded by the dashed rectangle in (a); results were normalized against markers believed to lie outside the amplicon (open circles). (c) Non-normalized MCC data (copies per aliquot) plotted against amplimer length, showing that the apparent copy number in this heavily-degraded DNA is strongly dependent on amplimer length. (d) μMCC of lesion A biopsy-derived DNA using primer set 1.1, in which amplimer length was standardized. Results were normalized to the average of five reference loci (white circles) (e) μMCC analysis of pooled normal genomic DNA using primer set 1.1. Results were normalized to the mean of three reference loci (open circles). In (a), (d) and (e), 'M' indicates a known multicopy marker

on the most challenging of templates: small numbers of cells microdissected from fixed biopsy material. For this, we used two DNA samples from a second FFPE biopsy (lesion B). The first (Bulk Sample 2001) had been extracted from epithelial cancer cells microdissected from multiple sections soon after surgical resection in 1999, and stored frozen since then. The second (Limited Sample 2007) was extracted in 2007 from laser-capture microdissected epithelium (Figure 4b) from two 6 μm sections of the same FFPE block after 6 years of storage at room temperature. This latter sample more accurately reflects a typical situation in which only small amounts of archive material may be available from a diagnostic biopsy.

The DNA from Limited Sample 2007 was first quantified using a dilution assay using amplimers of 100–120 bp. From these results (not shown) it was estimated that a total of about 160 haploid genome equivalents (480 pg, or 80 'normal cells'-worth) of effective genomic DNA (ie, DNA sufficiently intact to support amplification of 100–120 bp amplicons) had been recovered from the microdissected regions of the two 6 μm sections.

Despite the minute quantity of DNA available, analysis of the equivalent Limited Sample 2007 using μMCC with primer set 1.1 (Figure 4c) gave closely similar results to that of the better-preserved Bulk Sample 2001 (Figure 4a). There is more marker-to-marker variation in the Limited Sample results, possibly due to the greater level of degradation of the DNA, but the results illustrate that μMCC can be used to generate robust assessments of copy number variation at multiple loci, using picogram quantities of degraded DNA from fixed, microdissected archival tissue.

Fine junctional mapping of amplicons in fixed specimens by μMCC

Finally, we sought to illustrate how powerful a tool μMCC can be in precisely analysing cancer genomes in fixed material — in this case by delineating the boundaries of a copy-number variation (an amplification) in degraded DNA from archived FFPE specimens, using Bulk Sample 2001 (see above). The borders of the amplicon had previously been delineated to within 2 Mb using primer set 1.1 (above; Figure 4). In a similar iterative process to that employed previously using high quality DNA extracted from a cell line [16], repeated rounds of MCC were carried out using two further primer sets (primer sets 3 and 4, Appendix B) in order to finely map the copy number transition between normal and high copy at the centromeric border of the amplicon. We were able to pinpoint the copy-number transition to intron 17–18 of the gene *NP_078963.2*; DNA from histologically normal tissue from the same surgical resection specimen showed no apparent copy-number transition in this region (Figure 5).

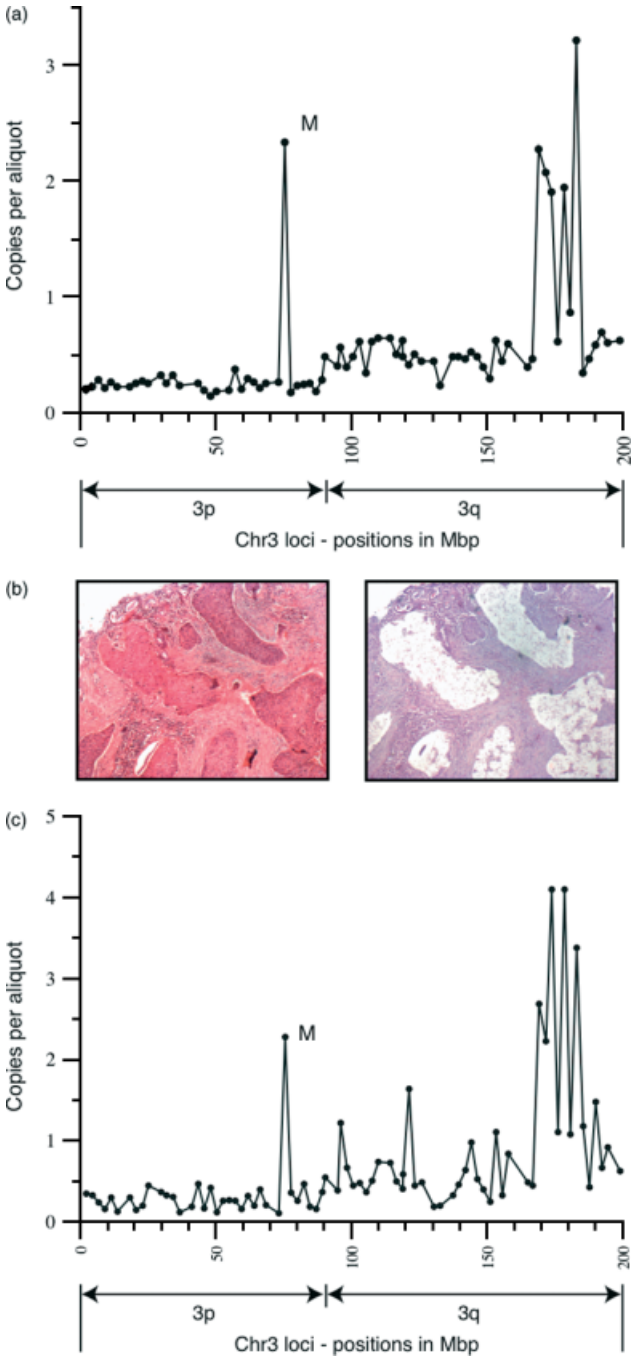

Figure 4. μMCC analysis is effective on microdissected archived specimens. (a) μMCC results using primer set 1.1 on DNA extracted in 2001 from a biopsy of a resected invasive lung cancer taken in 2000 (lesion B; Bulk Sample 2001). (b) One of two 6 μm sections cut in 2007 from the same biopsy from 2000. The section is shown before (left) and after (right) laser capture microdissection of cancerous cells. (c) μMCC results using primer set 1.1 on DNA recovered from the microdissected cells. These results were not normalized. In (a) and (c), 'M' indicates known multicopy loci

Discussion

Molecular copy-number counting (MCC) has been established as a technically simple and robust method for identifying copy-number changes of the order of 2 : 1 in good-quality genomic DNA [16] but, as we

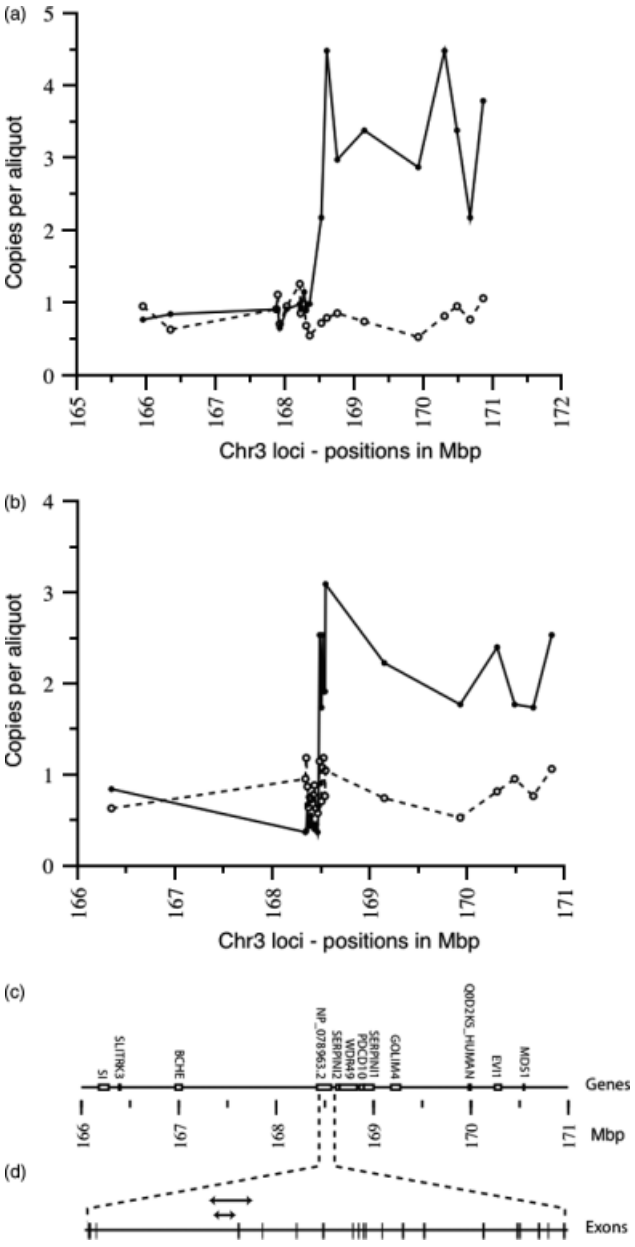

Figure 5. High-resolution analysis of amplicon junction. (a) μMCC results on DNA from Bulk Sample 2001 (filled circles, solid line; see text) and from Normal Lung 2001 (extracted from histologically normal lung tissue from the same surgical resection specimen as bulk sample 1 — open circles, dashed line) using sets of loci (primer sets 3 and 4) spaced more closely around the proximal border of the amplicon identified in Figure 4a; results are not normalized. (b) μMCC results on the same DNAs, using markers spaced more closely still around the amplicon border; results are not normalized. (c) Genes in the same region of chromosome 3 represented in (b). (d) Expanded view showing the exons of gene *NP_078963.2*; the results shown in (b) locate the proximal border of the amplicon to the region indicated by the upper double-ended arrow. A further round of μMCC with just two markers within this region (results not shown) defines the boundary to within a repeat element within intron 17–18 of *NP_078963.2* (lower double-headed arrow)

have shown, is less effective on badly degraded samples. μMCC is a modification of MCC which, by ensuring that all amplimers are of comparable (and, preferably, short) length, gives robust results even from heavily fragmented DNA. We have shown here that μMCC can operate over a wide range of copy numbers, accurately resolving intra-amplicon variation, and can yield good results even from the minute amounts of degraded DNA obtained by microdissection of archived formalin-fixed paraffin-embedded biopsy samples.

The wide dynamic range of μMCC is a key advantage. We have demonstrated a useful dynamic range of about eight-fold (0.5–4 copies/haploid genome), as seen in the cancer specimens which show both loss of 3p and amplification of parts of 3q. In the case of the HONE-1 amplicon (Figure 3c), we measured copy number changes in the range of 1.5 to the greatest amplification at 3.5–4-fold; this compares well with FISH data, which reveals 5, 6, 8, 10 and 13 copies of the corresponding loci per cell or (since HONE-1 is a hypotetraploid cell line [19,23]) approximately 1.5, 2 and 3–4 copies/ genome.

When DNA is loaded at approximately 0.7 haploid genomes/aliquot, the assay 'saturates' at >4 copies/ haploid genome. However, the range of μMCC can be extended considerably by choosing higher or lower amounts of DNA per aliquot. For example, by loading about 0.2 haploid genomes (0.7 pg) of DNA per aliquot, 'normal' copy reference markers give approximately 18% positives (enough to serve as a robust baseline in a panel of 80–90 aliquots), and the assay will not saturate until about 14-fold amplification is reached. The use of greater numbers of aliquots, or of sets of aliquots covering a range of DNA concentrations, can further extend the dynamic range of the assay. The ability to resolve extremes of amplification may be important in distinguishing 'driver genes', which one might expect to be the most highly amplified sequences in some cancers, from 'passenger genes', which may be co-amplified but whose amplification confers no selective advantage on the cell [4].

MCC assumes that all loci tested are amplified with approximately the same efficiency. Clearly this is not the case when badly degraded DNA is tested for amplicons of widely varying length: the longer targets in the template are more likely to have been broken than the shorter ones. However, by selecting short amplimers of approximately uniform length, μMCC largely overcomes this limitation. This in turn relies on the assumption that DNA degradation in fixed, embedded samples is approximately uniform throughout the genome — an assumption which we have not completely validated. Indeed, the residual 'noise' seen in the analysis of badly degraded DNA, even when using standardized amplimers (Figure 4c), may reflect non-uniform degradation or protection of the DNA at sporadic locations. As in any PCR assay, a proportion of markers do 'fail', consistently giving few or no (or very faint) positives; these usually occur sporadically, and are readily identified (eg because they suggest a copy-number of <0.5 copies/haploid genome, and behave similarly on all templates). In a few other cases, some markers give low but plausible

numbers of positives (data not shown); such apparent localized copy-number losses can be checked against a reference sample of normal genomic DNA. The standardized criteria developed for primer design keep such cases to a minimum.

Compared to other methods for analysing copy number variation [11], μMCC has relatively few technical requirements and, as it depends largely on PCR, is amenable to automation. It represents a step-change in the ability to assess multiple markers in one clinical biopsy when compared to other target-driven techniques such as FISH; we typically analyse up to 96 loci in a single experiment, but up to 1200 markers have been successfully amplified using a similar protocol (Justin Pachebat, personal communication). Although arrays are capable of analysing far larger numbers of loci in a single experiment, μMCC gives cleaner data (especially when compared to array performance on degraded DNA), requiring no smoothing or segmentation. It has also been reported that arrays tend to underestimate extremes of amplification [24], but this does not appear to be a problem in μMCC.

The key advantage of μMCC, however, is its applicability to formaldehyde-fixed and embedded archived clinical specimens. There has been great interest in recovering information from such archives [5–9,13–15,25,26]. Formalin causes cross-linking of DNA, both to proteins and between nucleotides [10], and there is evidence that the duration of both formalin fixation and storage can impact on subsequent analysis [6–8]. While there are reports of improved recovery of longer DNA molecules from such samples [8,15,25], and of successful array-based assessment of copy number variation using FFPE specimens [9,13–15,26], it remains the case that array platforms are often intolerant of the quality and quantity of template available from fixed material [5,8,9]. For example, in a recent series of 93 breast cancers, 42 failed to produce an acceptable array CGH profile despite an optimized protocol [8]. Furthermore, the noisier data generated on FFPE specimens means that more smoothing is required, leading to lower resolution mapping of genomic events [14]. In our extended experience with μMCC (not shown), we have successfully applied the protocol to microdissected epithelium from 23 preinvasive bronchoscopic biopsies, and to six lung cancer resection specimens. For a further five preinvasive lesions, we were unable to extract the equivalent of 160 haploid genomes and so had insufficient DNA to quantify and then perform an experiment. It is anticipated that sufficient DNA could be extracted from almost all invasive cancer specimens, whether bronchoscopic biopsy or surgical resection specimens.

μMCC is very flexible, and can be readily adapted to regions of particular interest, making it well suited to the screening of candidate regions, or of a set of scattered candidate loci. In this study we have demonstrated analysis of the 3q amplicon in an epithelial cell line and a squamous cell carcinoma of the lung. We have shown how specific questions, such as intra-amplicon variation and the precise localization of an amplicon junction, can be addressed by the simple design of new targeted primer sets. A similar approach could be used in a clinical context to examine, in small formalin-fixed diagnostic biopsies, multiple selected loci anywhere in the genome for diagnosis, prognosis and treatment stratification. A number of single loci are already assessed for amplification, and the results used to guide clinical decisions [27,28]. We believe that μMCC will also prove complementary to successful array-based analyses, allowing the rapid, accurate and ultra-fine resolution of structural variation detected by array platforms.

We have demonstrated clearly that μMCC, using amplimers of a consistent length, is tolerant of minute quantities of grossly degraded DNA from FFPE specimens, and even from cells microdissected from such specimens. This not only permits finer selection of the subpopulation of cells to be analysed, but also conserves valuable biopsy material for other analyses, such as immunohistochemistry or expression profiling. We expect this approach will be useful to biomedical researchers and clinicians hoping to unlock the secrets of the cancer genomes held in archives world-wide.

Acknowledgements

This work was supported by the UK Medical Research Council and the Rosetrees Trust.

Supporting information

Supporting information may be found in the online version of this article.

References

1. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, *et al*. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;**258:**818–821.
2. Rabbitts TH. Chromosomal translocations in human cancer. *Nature* 1994;**372:**143–149.
3. King CR, Kraus MH, Aaronson SA. Amplification of a novel *v-erbB*-related gene in a human mammary carcinoma. *Science* 1985;**229:**974–976.
4. Stallings RL. Are chromosomal imbalances important in cancer? *Trends Genet* 2007;**23:**278–283.
5. Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 2005;**37**(suppl):S11–17.
6. Miething F, Hering S, Hanschke B, Dressler J. Effect of fixation to the degradation of nuclear and mitochondrial DNA in different tissues. *J Histochem Cytochem* 2006;**54:**371–374.
7. Stoecklein NH, Erbersdobler A, Schmidt-Kittler O, Diebold J, Schardt JA, Izbicki JR, *et al*. SCOMP is superior to degenerated oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples. *Am J Pathol* 2002;**161:**43–51.
8. van Beers EH, Joosse SA, Ligtenberg MJ, Fles R, Hogervorst FB, Verhoef S, *et al*. A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer* 2006;**94:**333–337.

9. Little SE, Vuononvirta R, Reis-Filho JS, Natrajan R, Iravani M, Fenwick K, *et al*. Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. *Genomics* 2006;**87:**298–306.
10. Tang EP. *Path to Effective Recovering of DNA from Formalin-fixed Biological Samples in Natural History Collections.* Workshop Summary edn. National Academies Press: Washington, DC, 2006.
11. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006;**7:**85–97.
12. Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y, *et al*. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci USA* 2005;**102:**9625–9630.
13. Oosting J, Lips EH, van Eijk R, Eilers PH, Szuhai K, Wijmenga C, *et al*. High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. *Genome Res* 2007;**17:**368–376.
14. Jacobs S, Thompson ER, Nannya Y, Yamamoto G, Pillai R, Ogawa S, *et al*. Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. *Cancer Res* 2007;**67:**2544–2551.
15. Bredel M, Bredel C, Juric D, Kim Y, Vogel H, Harsh GR, *et al*. Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. *J Mol Diagn* 2005;**7:**171–182.
16. Daser A, Thangavelu M, Pannell R, Forster A, Sparrow L, Chung G, *et al*. Interrogation of genomes by molecular copy-number counting (MCC). *Nat Methods* 2006;**3:**447–453.
17. George PJ, Banerjee AK, Read CA, O'Sullivan C, Falzon M, Pezzella F, *et al*. Surveillance for the detection of early lung cancer in patients with bronchial dysplasia. *Thorax* 2007;**62:**43–50.
18. Glaser R, Zhang HY, Yao KT, Zhu HC, Wang FX, Li GY, *et al*. Two epithelial tumor cell lines (HNE-1 and HONE-1) latently infected with Epstein–Barr virus that were derived from nasopharyngeal carcinomas. *Proc Natl Acad Sci USA* 1989;**86:**9524–9528.
19. Darai-Ramqvist E, de Stahl TD, Sandlund A, Mantripragada K, Klein G, Dumanski J, *et al*. Array-CGH and multipoint FISH to decode complex chromosomal rearrangements. *BMC Genomics* 2006;**7:**330.
20. Schuler GD. Sequence mapping by electronic PCR. *Genome Res* 1997;**7:**541–550.
21. Foster NA, Banerjee AK, Xian J, Roberts I, Pezzella F, Coleman N, *et al*. Somatic genetic changes accompanying lung tumor development. *Genes Chromosomes Cancer* 2005;**44:**65–75.
22. Pei J, Balsara BR, Li W, Litwin S, Gabrielson E, Feder M, *et al*. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* 2001;**31:**282–287.
23. Kost-Alimova M, Darai-Ramqvist E, Yau WL, Sandlund A, Fedorova L, Yang Y, *et al*. Mandatory chromosomal segment balance in aneuploid tumor cells. *BMC Cancer* 2007;**7:**21.
24. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, *et al*. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004;**64:**3060–3071.
25. Sato Y, Sugie R, Tsuchiya B, Kameya T, Natori M, Mukai K. Comparison of the DNA extraction methods for polymerase chain reaction amplification from formalin-fixed and paraffin-embedded tissues. *Diagn Mol Pathol* 2001;**10:**265–271.
26. Johnson NA, Hamoudi RA, Ichimura K, Liu L, Pearson DM, Collins VP, *et al*. Application of array CGH on archival formalin-fixed paraffin-embedded tissues including small numbers of microdissected cells. *Lab Invest* 2006;**86:**968–978.
27. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science* 1987;**235:**177–182.
28. Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, *et al*. Erlotinib in lung cancer — molecular and clinical predictors of outcome. *N Engl J Med* 2005;**353:**133–144.