Journal of Pathology
J Pathol 2010; **220**: 297–306
Published online 19 November 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/path.2647

Invited Review

Single-molecule genomics

Frank McCaughan* and Paul H Dear
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

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

Conflict of interest: A patent for molecular copy-number counting has been granted to the UK Medical Research Council. Paul H Dear is named on the patent application.

Received: 27 September 2009
Revised: 5 October 2009
Accepted: 5 October 2009

Abstract

The term 'single-molecule genomics' (SMG) describes a group of molecular methods in which single molecules are detected or sequenced. The focus on the analysis of individual molecules distinguishes these techniques from more traditional methods, in which template DNA is cloned or PCR-amplified prior to analysis. Although technically challenging, the analysis of single molecules has the potential to play a major role in the delivery of truly personalized medicine. The two main subgroups of SMG methods are single-molecule digital PCR and single-molecule sequencing. Single-molecule PCR has a number of advantages over competing technologies, including improved detection of rare genetic variants and more precise analysis of copy-number variation, and is more easily adapted to the often small amount of material that is available in clinical samples. Single-molecule sequencing refers to a number of different methods that are mainly still in development but have the potential to make a huge impact on personalized medicine in the future.

Keywords: single molecule; PCR; gene; genomics; sequence; copy number variation; mutation detection; personalized medicine

Why single-molecule genomics?

The term 'single-molecule genomics' (SMG) refers to the study of genomes through the analysis of single, as opposed to pooled or cloned, DNA molecules [1–3]. SMG techniques can be applied either to the analysis of specified molecular targets or to genome resequencing [4], and have certain advantages that may be particularly attractive to modern diagnostic pathology and pharmacogenomics. For example, single-molecule digital PCR is robust, sensitive and quantitative, as well as being tolerant of the minute quantities of DNA available in clinical samples. With respect to sequencing, much of the current effort in DNA sequencing goes into the assembly of short reads of DNA sequences into a final sequence result [5]. One important aim of single molecule sequencing is to produce long reads from single molecules, removing any need for cloning/amplification and dramatically reducing the demands on sequence assembly.

The translational application of genomic techniques currently lags far behind the technical capabilities. It is of critical importance that, in the desire to deliver personalized medicine, genomic information that is accurate, reproducible and has direct clinical relevance is obtained. It is our view that single-molecule techniques can play a major role in delivering personalized medicine. As this is a vast and ever-expanding field, we here offer an assessment of the current status of single-molecule genomics, with a focus on current clinical applications, and we speculate on likely future developments. We concentrate on two broad areas: first, techniques relying on the detection and/or counting of single molecules; and second, single-molecule DNA sequencing.

Detecting and counting single molecules — digital PCR

A critical distinction between standard and single-molecule PCR (smPCR) is that in the latter, template DNA undergoes limiting dilution, to on average, less than 1 target molecule per aliquot [6,7], so that each aliquot either does or does not contain the sequence of interest. Detection relies on the fact that PCR is sufficiently sensitive to amplify single target molecules if they are present. In this scheme individual aliquots can only be positive or negative for the target sequence — a binary situation leading to the term 'digital PCR' [8]. Digital PCR with multiple aliquots allows the relative quantitation of separate target molecules, or the detection and quantitation of rare variants by the simple process of counting the number of aliquots that are positive for a target molecule (Figure 1). This differs from standard PCR, in which the signal generated from the amplification of multiple copies of the same locus template is measured, a process that is less sensitive to the presence of rare variants in a template pool and less accurate if relative

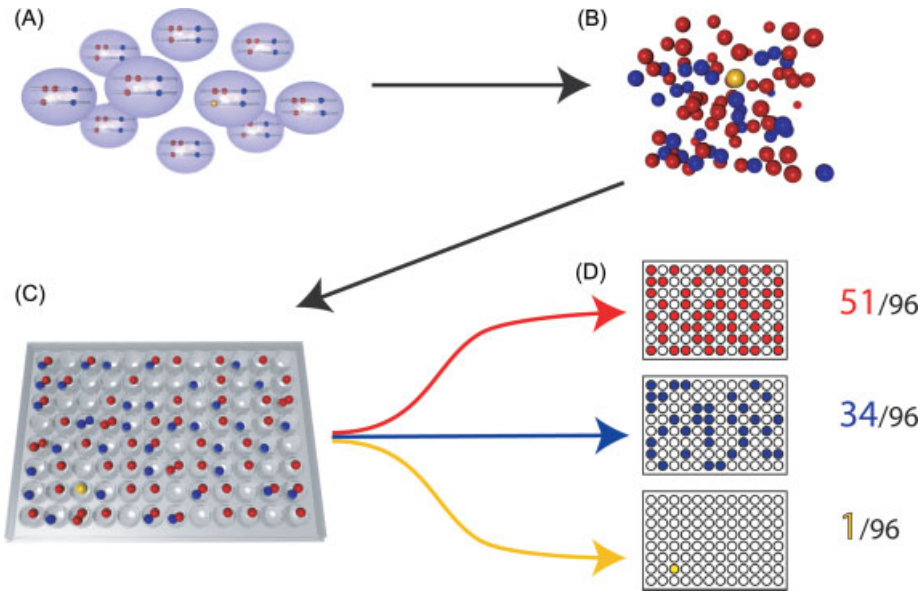

Figure 1. Basis of digital PCR. In this example, a population of cells (A) each carry two copies of one locus (blue) and three of another locus (red); in addition, a minority of the cells carry a rare variant sequence (yellow). When DNA is prepared *en masse* from these cells (B), conventional quantitative methods may struggle to measure accurately the ratio of red to blue loci, and may fail to detect the rare variant. In digital PCR, the DNA is divided at limiting dilution into a set of aliquots (C; shown here in a 96-well microtitre plate). Each aliquot is screened for each of the loci of interest (D), allowing accurate calculation of the relative abundance of different loci and the robust detection of rare variants

Table 1. Single-molecule PCR (smPCR) strategies

Objective	Strategy	References
Target detection and/or quantification		
Rare variant (mutation, SNPs, methylation status)	Limiting dilution digital PCR with locus-specific probes	8,21,23,42
Forensics	Limiting dilution PCR to analyse microsatellite alleles	6
Establishing linkage		
Happy mapping, molecular haplotyping	Multiplex digital PCR and further analysis to establish linkage; direct imaging of polymorphic sites on individual DNA molecules; polony haplotyping	7,53–55,60,61
Structural genomic variation		
Copy-number variation, including detection of aneuploidy	Relative locus quantitation using digital PCR	8–10,36,37,42

Table 2. Single-molecule PCR (smPCR) platforms

Strategies	Description
Microtitre plates	SmPCR in a standard plate format
Polonies	Key description of solid-phase localized single-molecule PCR—effectively creating concentrated clones of a single DNA template molecule [58])
Emulsion PCR	As above, except that PCR of a single template molecule was performed in droplets [56]
BEAMing*	Involves single-molecule emulsion PCR, in which the template is tethered to a magnetic bead that facilitates the subsequent separation of allelic variants [57]
Microfluidics (static digital array)	Individual PCR reactions take place in nanolitre reaction chambers on a microfluidics chip. In the Fluidigm BioMark system (www.fluidigm.com) the chip fits in a proprietary apparatus consisting of both a thermocycler for 'on-chip' PCR and a detection apparatus (CCD) to read the result from each chamber. The number of reactions is finite, depending on the chip design
Microfluidics (dynamic)	Individual PCR reactions take place in droplets that undergo on-chip PCR as they migrate via an oil stream through a microfluidic chip that is exposed in sequence to the temperatures required for the polymerase chain reaction. The number of individual PCR reactions is in principle unrestricted [42,59]

* BEAMing (beads, emulsion, amplification, magnetics).

quantitation is attempted. Using the basic digital PCR system, experiments can be readily designed to answer a specific research or clinical question. A number of applications and platforms have been proposed and are described below and summarized in Tables 1 and 2.

Applications of digital PCR

In principle, any DNA sequence/variant that can be tested for and detected with a standard PCR strategy is suitable for digital PCR, with the added advantage that by counting the number of aliquots positive for a target sequence, digital PCR can be quantitative. Therefore, not only can it be used to detect rare variants in a population of DNA molecules [8] but it can also estimate the frequency of a variant sequence or, indeed, the

relative copy-number of separate sequences in template DNA [9,10]. A number of proposed applications are discussed in more detail.

Detection of (rare) variants of clinical significance

In clinical medicine there is great interest in the ability to detect rare genetic variants within a large pool of 'normal' DNA. A common example is the detection, in oncological practice or screening, of pathological mutations in known oncogenes. This was the context in which Vogelstein and Kinzler first coined the term 'digital PCR' [8]. The detection of critical mutations, even at a low frequency, suggests that, within the mass of normal cells in a given clinical sample, there are some that have acquired a molecular biomarker that may be directly associated with cancer or may indicate the presence of preneoplastic disease with a high risk of progression to cancer. Point mutations can be readily detected by standard PCR when there is a mixed population of wild-type and mutated sequences. However, when the mutated sequence is relatively rare (previously estimated as $<20\%$ of the total [11,12]), the efficiency of detection falls. Using a digital PCR approach, in each positive aliquot the individual molecule is either wild-type or mutated and the detection strategy can be tailored to detect both. Therefore, depending on the number of aliquots analysed, a mutation can be detected even when it accounts for only a very small percentage of the total alleles present, and the frequency of that mutation in the sample can be reliably estimated.

This strategy was first described in the detection of *KRAS* mutations in stool samples [8], but a similar approach has now been applied in multiple scenarios, including the analysis of *ABL* tyrosine kinase in chronic myeloid leukaemia [13], *KRAS* mutations in ovarian cancer [14], the detection of multiple mutations (*TP53, PIK3CA, KRAS* and *APC*) in plasma and stool samples of patients with colorectal cancer [15], and the analysis of *EGFR* mutations in tissue and plasma samples from patients with non-small cell lung cancer [16]. The ability to perform these assays on non-invasive specimens such as stool or peripheral blood is particularly important, with the implication that disease can potentially be both diagnosed and monitored in this way.

Single molecule detection of methylation status

The methylation of CpG dinucleotides is a critical epigenomic control mechanism that has been implicated in the pathogenesis of cancer [17] and multigenic non-neoplastic diseases [18], and there is much interest in exploiting specific gene methylation profiles as molecular biomarkers in oncology [19,20]. Standard methodology for the detection of methylation at specific loci is based on either bisulphite sequencing and/or methylation-specific probes. The potential benefits that single molecule approaches bring to these techniques has been demonstrated in a recent series of papers relying on limiting dilution of bisulphite-converted template DNA [21–23]. Bisulphite sequencing traditionally necessitated bacterial cloning and, although many modern protocols have sought to overcome this time-consuming step, it remains the gold standard [21]. Since PCR amplification of individual molecules effectively clones the original single molecule, there is scope to omit bacterial cloning, thus reducing assay complexity and cost without compromising the methylation read accuracy [21]. In one published protocol, following template dilution, aliquots are tested using primers specific for bisulphite-converted DNA and the PCR products in positive wells are then sequenced, allowing a precise analysis of the methylation status at multiple specific CpG dinucleotides in individual DNA molecules [23]. In a related protocol, methylation-specific probes can be used to detect rare methylation events with a sensitivity that greatly exceeds that of a standard protocol [23].

Analysing relative copy-number by digital PCR

Large tracts of the genome are duplicated or deleted in phenotypically 'normal' individuals [24–26]. Although the vast majority of inherited copy-number variants (CNVs) are likely to represent benign variation, a number have been associated with a clinical phenotype [27,28]. Somatic copy-number alterations also occur; some years ago regional genomic amplifications were described in human cancers [29,30] and it is clear that in some cancers these events are critical and may predict the response of an individual's tumour to specific biological therapies [31,32]. The potential for knowledge of specific CNVs to directly influence clinical decision-making has led to the routine integration of copy-number analysis in patient care algorithms, an example being the analysis of *HER2* copy-number in the stratification of breast cancer patients to trastuzumab [33].

There are many methods described for measuring copy-number variation, each with potential benefits and problems, and these have recently been reviewed in detail [34]. Digital PCR has since emerged as a viable alternative to more traditional methods, and there are situations in which it undoubtedly offers considerable benefits. Simply by counting the number of aliquots that are positive for one sequence and comparing that to the number of aliquots positive for a second sequence, an estimate of the relative abundance, and therefore copy-number, of two sequences in the template DNA can be made [8,9]. The potential benefits lie in the relative precision of results and in the ready application to targeted loci in scarce clinical specimens.

With respect to accuracy, it is recognized that array-based analysis of CNVs underestimates the amplitude of CNV compared to real-time or quantitative PCR (qPCR) approaches [35]. There is evidence that digital

PCR approaches may be more accurate than qPCR [36–38] and be more readily able to discriminate integer copies of target sequences in a way that is not reproducibly the case for qPCR [36].

A key part of the strategy in digital PCR is limiting dilution of template DNA. Since in many clinical situations, for example diagnostic biopsies, there may only be a tiny quantity of DNA available, the ability to tolerate small quantities of template is an advantage. This is particularly true of small, heterogeneous biopsies consisting of both tumour and stromal cell populations necessitating microdissection of the tumour cells. In addition, the standard fixatives used to preserve histological appearance for diagnostic purposes significantly compromise the quality DNA available for downstream analysis [39]. We recently described molecular copy-number counting (MCC) [9] and a modified protocol microdissection MCC (μMCC) [10]. MCC/μMCC incorporates a two-phase nested PCR protocol in which multiple primer pairs can be multiplexed in phase 1, allowing for the simultaneous and accurate analysis of the relative copy-number of hundreds of loci on the same template DNA and in the same experiment. Through the design of external primers of consistent length, we demonstrated that this type of analysis can be applied to grossly fragmented DNA from archived formalin-fixed specimens [10].

MCC/μMCC has been used to define the breakpoint of a non-reciprocal translocation in cell line DNA and to define regional copy-number variation in cell line and archived material [9,10]. The benefits of this system are that multiple sequences are targeted simultaneously and cheaply, using basic primer design strategies, and the techniques required are already available in most molecular biology laboratories.

Aside from cancer diagnostics and treatment decision-making, digital PCR has been most successfully applied to the prenatal diagnosis of fetal aneuploidy. In one study the rapid discrimination of aneuploidy was demonstrated in amniotic fluid or in tissue from chorionic villus sampling [40].

Genome mapping using digital PCR

HAPPY mapping (mapping based on the analysis of approximately haploid DNA samples using the polymerase chain reaction) uses limiting dilution and single molecule PCR to examine the physical relationship between markers on high molecular weight DNA. By establishing how close markers are to each other, a linkage map of the genome can be built up. Following a nested or hemi-nested protocol similar to that described above for MCC, DNA that has undergone limiting dilution to less than a haploid genome per well is simultaneously tested for the presence or absence of a series of markers. As the DNA is of very high molecular weight, adjacent genomic markers will be more likely to be present in the same aliquots, whereas more distant markers will be less likely to segregate to the same aliquots — the greater the physical difference between markers, the more likely a single DNA molecule will have fragmented in the interval. Up to 1200 markers can be multiplexed simultaneously and their relative positions established, so that a detailed linkage map can be established. This technique has already been used for a number of genomes [7,41] and is currently being applied to the study of cancer genomes, with the aim of mapping across reciprocal translocations and cloning the translocation junctions (Pole J, McCaughan F, Dear P and Edwards P, personal communication). In other work, next-generation sequencing strategies and HAPPY mapping are being combined to facilitate the more rapid assembly of genomes sequenced *de novo* (Dear P, personal communication).

Practical issues related to digital PCR

Choice of platform

The ideal digital PCR system would have infinite numbers of negligible volume single-molecule reactions, with each PCR reaction proceeding with 100% efficiency and being detected with fail-safe methodology. We are not there yet. The reality is that the available systems are evolving rapidly and that a user's choice of platform will depend on the specific application, the budget and reagent costs. In practice, the choice lies between the more traditional format of 96-, 384- or 1526-well microtitres plates, readily adaptable to modern robotics, and newer static [36] and dynamic microfluidic systems [42].

Numbers of aliquots

There are two main advantages to having a large number of separate aliquots. One is that the detection threshold for rare variants can be lowered, leading to improved sensitivity. For example, in microtitre plates the theoretical limit of sensitivity would be approximately 1% in a standard 96-well plate, decreasing to <0.1% in 1536-well plates [8]. However, in high-throughput microfluidics systems with the potential to rapidly analyse thousands of aliquots, there is the potential for even greater sensitivity [42]. A second is that the accuracy and dynamic range of relative copy-number analysis increases, assuming optimal loading of DNA to less than a haploid genome per aliquot [38,42,43]. Microfluidics platforms with many thousands of individual reactions therefore have a distinct advantage over microtitre plates, although using the latter it has been repeatedly shown that accurate results are readily obtained. When determining the relative copy-number of individual loci based on the number of aliquots that are positive, the Poisson equation is used. This assumes that the distribution of DNA molecules is random and anticipates that some positive wells will contain more than one molecule. Further details of the

mathematical background and a method for determining the statistical significance of copy-number results are available [38,43].

Reaction volumes

With respect to reaction volume, the smaller the volume of reaction, the lower the cost per assay, as reagents (polymerase, primers, etc.) contribute significantly to costs. The microtitre plate format has the advantage that most laboratories already have the necessary facilities — access to a 'clean room' and thermocyclers. For many users this robust system will remain attractive for low-throughput small-scale applications and it ensures that PCR products are available for downstream analysis. The microfluidic platforms [36,42] have undoubted benefits in terms of scale and reagent costs. A further advantage of smaller volumes is the theoretically reduced risk of aliquot contamination with non-template DNA.

PCR efficiency

In standard quantitative PCR strategies, ensuring that amplication efficiency is close to 100% is very important and, if the relative copy of two different loci are being compared, preliminary assays are performed to demonstrate that the amplification efficiency for each markers is almost equivalent. Even with these measures, amplification efficiency remains an important source of bias and a reason why very accurate measures of copy-number are difficult with qPCR. In principle, amplification efficiency for different targets in a digital PCR assay does not have to be 100%; efficiency merely needs to be sufficient that a threshold for detection will be met over the course of a given number of amplification cycles. In practice, even this goal may require significant primer optimization prior to 'passing' each assay. One strategy is to include, as in published protocols [9,10,36], a two-phase PCR strategy to enrich the concentration of target molecules prior to detection. However, for clinical diagnostics, a streamlined single-phase PCR would be preferable. Optimizing amplification efficiency in digital systems is a subject of ongoing research and development [38].

Detection strategy

A full discussion of detection strategies is beyond the scope of this article and readers are directed to the primary publications. In brief, amplicons can readily be detected using standard nucleic acid stains, such as SYBR-green. With respect to the detection of rare variant sequences, fluorescently-labelled molecular probes can be designed to discriminate wild-type from mutants [8,16]. Again, it is only the presence or absence of a product that is being measured in each aliquot, so detection is relatively straightforward. As with standard PCR, the situation becomes more complex if attempting to assay multiple specific gene mutations with a probe-based assay. For example, in a recent study detecting the two most common hot spots for mutations in *EGFR* in lung adenocarcinoma, a total of five fluorescent markers were required [16]. For microfluidic devices, if separate laser detection systems are necessary, the complexity of the engineering increases.

Single-molecule DNA sequencing

Background

DNA sequencing is the basic technology underpinning the field of genomics, with the most notable triumph being the first draft of the complete human genome sequence in 2001 [44], accomplished entirely in sequencing farms by standard Sanger sequencing. Since its original description [45], there have been refinements of the Sanger sequencing protocol to improve throughput and sequence accuracy. It continues to be the 'gold standard' but is restricted to a maximum read or sequence length of approximately one kilobase.

'Next-generation' DNA sequencing

The speed and potential of genome sequencing has been transformed by the so-called 'next-generation sequencing' (NGS) platforms [5]. In NGS, individual short (<1 kb) DNA template molecules are isolated and then amplified through PCR. This process is performed in parallel on many millions of DNA molecules (massively parallel). In this way huge numbers of individual DNA molecules are amplified or 'molecularly cloned' prior to sequencing. Once DNA molecules have been cloned, sequencing of each clone is accomplished in different ways; Solexa/Illumina (http://www.illumina.com/) and 454/Roche (http://www.454.com/) use 'sequencing-by-synthesis', in which the incorporation of fluorophore-labelled nucleotides is measured through the phasing of nucleotide delivery or differential labelling of the four nucleotides. The Applied Biosystems/SOLiD (http://www3.appliedbiosystems.com) technology is slightly different, in that DNA clones are sequenced by repeated hybridization of differentially-labelled degenerate octamer probes. In each case the protocol is complicated by the need for repeated alternate reagent and washing cycles. Complete genomics (http://www.completegenomicsinc.com/) have produced a modified hybridization-based protocol in which the read length has been increased, facilitating more rapid sequence analysis.

There is no doubt that NGS platforms are dramatically enhancing the scale and ambition of sequencing projects, and they have also been applied successfully to expression studies, the assessment of CNVs and, through template-enrichment protocols, rare variant analyses [46]. The speed and versatility of these

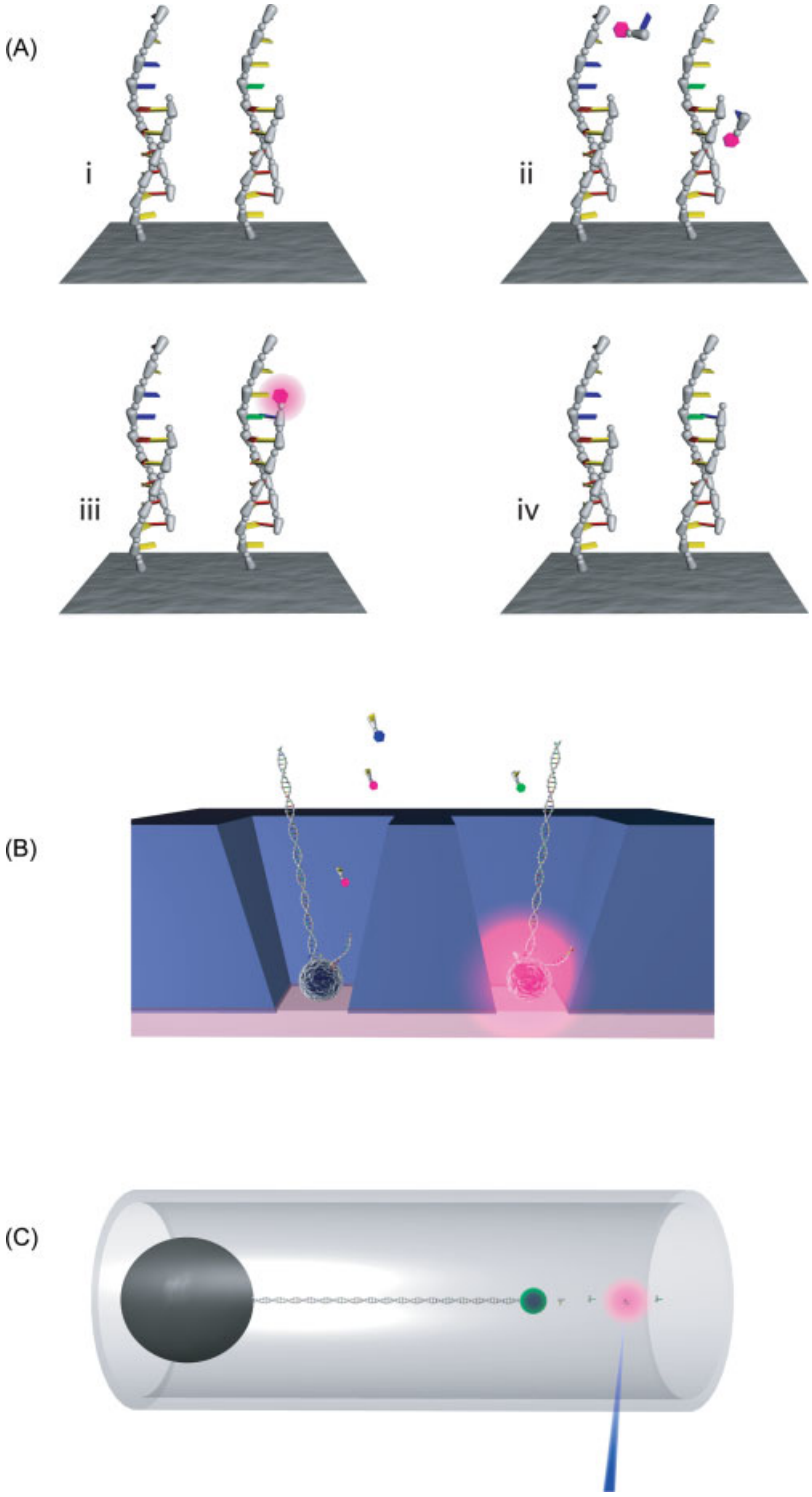

Figure 2. Schematic representation of some current and proposed single-molecule sequencing strategies. In the Helicos system (A), poly(dT) molecules are tethered to a glass slide and capture template DNA molecules that have been primed with poly(dA) tails (i). A DNA polymerase and one of the four bases (A, C, G or T) are added (ii), and the polymerase will incorporate the base into any template molecules whose extension is awaiting that base (iii). The added base carries a fluorescent tag, which not only prevents further extension of the template but also allows the added base to be detected by fluorescence microscopy. Once this has been done, the fluorescent tag is removed (iv) and the cycle repeats with next base, similarly tagged, being added. In Pacific Bioscience's SMRT system (B), DNA polymerase molecules (spheres) are fixed to the transparent base of each tiny chamber and bind DNA template. All four nucleotides, carrying different fluorescent tags on their terminal phosphates, are added. As a polymerase molecule incorporates a nucleotide, the fluorescence in the chamber (shown as a pink glow here) is detected. The fluorescently-tagged terminal phosphate group is released by the polymerase and diffuses out of the chamber. In exonuclease sequencing (C), a DNA molecule is tethered to a bead (sphere on left) in a microfluidic channel; liquid flows past the bead, from left to right. An exonuclease (green) cleaves successive bases from the end of the DNA. These are carried by the flow stream past a detector and identified (right). To date, this method has not been implemented successfully

J Pathol 2010; **220**: 297–306 DOI: 10.1002/path

platforms has meant they have quickly become standard and extremely productive tools in the research environment. However, there are a number of issues that may restrict the routine application of NGS in the clinic.

Limitations of next-generation sequencing

There are a number of limitations associated with the widespread clinical use of NGS platforms. First, the absolute template requirements are generally very high and, although a strategy involving digital PCR has been proposed to mitigate this [47], the implication is that, for many clinical samples, there may be too little DNA available. There are no published data on the performance of archived material in NGS protocols. Second, the protocols are neither straightforward nor cheap and, although improvements in both parameters are anticipated, the costs would currently be difficult to justify for routine clinical analysis. Third, each read length is relatively short and huge numbers of reads are peformed in a single analysis, with a consequent massive dependence on major information technology infrastructure and bioinformatics expertise. This issue can be mitigated by enrichment strategies aimed at the analysis of specific loci, but will continue to be a significant issue. Fourth, there is a small but significant base-call error rate, particularly associated with homopolymer repeats in the 454 protocol and with so-called 'dephasing' on other platforms. The latter refers to the fact that the synchronicity of sequencing by synthesis cycles may not be maintained, so that reads of the same part of the genome may give slightly varying results. The coverage (the number of times the same section of the genome is read and compared) needs to be high, perhaps over $\times 15$, to ensure acceptable sequence accuracy. Finally, as there is a PCR amplification step, there is a risk of amplification bias that can be difficult to measure. This has implications for some clinical applications, particularly the measurement of CNVs, as the variation in relative amplification efficiency of two loci will rapidly bias the estimation of their relative copy-number. Although this seems a long list of criticisms, it is important to stress that all sequencing technology has limitations, that NGS is a young technology that is evolving and being refined, and to restate that it is currently the most powerful genomics tool available and the ideal technology for the research environment. However, it does not yet seem to be appropriate for routine clinical application.

Single molecule sequencing — potential benefits and limitations

True single molecule sequencing is somewhat different from NGS; it refers to sequence analysis of individual molecules without prior cloning and, although full of promise, is not yet ready for 'prime-time' in the clinic. The reason that there is so much interest in developing SMS, despite the enormous achievements of NGS, is not merely technological bravado. The real prize is the ability to carry out accurate sequence analysis of individual long DNA molecules of up to 100 kb and beyond. Although this has not yet been achieved, it is the focus of a number of well-funded biotechnology companies. The reason that precise base-calling of much longer molecules is attractive is that such read-lengths, at a stroke, would remove many of the limitations discussed above in the context of NGS — first, the need for complicated and computer-hungry sequence assembly would be much reduced; second, the absence of PCR from the protocol would immediately remove any amplification-related bias; third, the amount of template required would theoretically be minimized, enabling clinical specimens to be more readily assessed; fourth, protocols would be vastly simplified, as the repeated reagent/washing cycles should be unnecessary. To date the major issues are that it is a largely unproven technology and therefore there are many uncertainties, including with respect to sequence accuracy and costs. If realized, SMS will lead to an explosion in sequencing which, in turn, will raise significant information technology (IT) issues about raw data management and mining.

There are a number of SMS protocols and platforms at various stages. We will now briefly summarize the proposed approaches — more details are provided in Figure 2, in company websites and in the quoted references.

Sequencing by synthesis (SMS)

Three of the proposed SMS technologies use a modified sequencing by synthesis technology. Helicos BioScience (http://www.helicosbio.com/) have already brought an SMS platform to market — True Single Molecule Sequencing (tSMS), utilizing similar technology to some of the NGS platforms. In their protocol, single molecules of DNA are tethered to a flow cell, then sequencing by synthesis is performed, with sequential incorporation of labelled nucleotides (Figure 2). As single molecules are analysed, the 'dephasing' problem discussed above is overcome and the protocol is simplified, although reagent cycling is still required. Critically, the reads remain short, at an average of 32 bps, meaning that sequence assembly continues to be a computer-intensive issue, and the incremental benefits over NGS for this application may therefore be limited. Helicos have successfully published complete viral [48] and human [49] sequences and are now directly competing with the NGS platforms. Pacific Biosciences (http://www.pacificbiosciences.com/) have introduced a separate method of sequencing by synthesis. They have described the use of a number of innovative solutions, including the design of tiny reaction chambers with even smaller detection volumes (20×10^{-21} l) and the use of fluorescently labelled phosphate groups,

rather than the more commonly used base labelling. In each nanochamber a single DNA polymerase molecule is tethered and, as it incorporates a labelled nucleotide complementary to the template DNA, the signal is read and the sequence deduced prior to the labelled phosphate moiety being cleaved by the polymerase (Figure 2). Pacific Biosciences plan a market release date of 2010. Visigen (http://visigenbio.com/) have used a different strategy involving the combination of sequencing by synthesis and Förster (fluorescence) resonance energy transfer (FRET), in which both the polymerase (donor) and the nucleotide (acceptor) are labelled with a fluorophore. On incorporation of one of the four nucleotides, a FRET signal specific for that nucleotide is produced and sequence information extracted.

Future technologies

Some of the other proposed methods of SMS are further from being realized but are technically fascinating and have significant potential. The idea that bases could be sequentially cleaved from a DNA molecule and then detected (exonuclease single-molecule sequencing), thus providing a sequence 'read', was first proposed many years ago [50]. However, despite much effort no viable technology has emerged. New solutions for this approach are actively under development (http://www.mrc-lmb.cam.ac.uk/happy/HappyGroup/seq.html), including a method that incorporates nanopore technology.

Sequencing using nanopore technology has already made significant progress and is attracting much investment [3,51]. The basis of nanopore sequencing is that single-stranded DNA is negatively charged and will move along an electric gradient towards a positive charge. If that gradient is across a membrane with nanometre pores, single DNA molecules will get caught in the pores and sequence information can be gleaned from the change in ionic current across individual nanopores, as the base composition influences the amplitude of current flow. There are many technical hurdles before this could be developed into a feasible SMS technology [51]. However, a number of groups are working on this type of platform in very innovative ways, eg one proposal combines exonuclease sequencing and nanopore technology for base detection [52]. The reader is referred to the following company websites for further information on other proposed strategies (http://nabsys.com/; http://www.lingvitae.com; http://www.bionanomatrix.com). Finally, ZS genetics (http://www.zsgenetics.com/) are devising an SMS method based on incorporating heavy atoms (such as iodine or bromine) as nucleotide labels. These molecules are of sufficient molecular weight to be 'read' by a high-resolution electron microscope.

Although it is too early to know how well SMS will perform in comparison to competing sequencing technologies, it is certain that technological advances will be made and it will be very interesting to see how the field develops.

Single-molecule genomics and personalized medicine

One of the aims of these heroic efforts in the field of single molecule genomics is to deliver truly personalized medicine. This is a complex issue, with huge ethical and economical as well as therapeutic implications. It is therefore important to approach the application of technology to decisions about patient care with due caution. With respect to delivering personalized medicine, we feel that physicians have relatively straightforward needs. They want a test that will enable them to make an informed clinical decision for the benefit of their patients. Currently, there is no benefit to clinicians in having the gigabytes of data on each biopsy/blood sample that next-generation or single molecule sequencing could deliver. In fact, this would be a distinctly negative factor, both in terms of data management and with respect to the provocation of a number of uncomfortable and ill-informed discussions on, for example, the link between a specific SNP and an ill-defined excess risk of a life-threatening illness.

Our prejudice is that the current potential for clinical application of single-molecule digital PCR techniques is very different to that of the sequencing technologies. PCR is an exceptionally reliable and now mature technique that has been readily adapted to the robust detection of single DNA molecules. We have discussed a range of potential clinical applications for which smPCR is appropriate, likely to be more accurate than competing technologies and more easily adapted to the often small amount of clinical material that is available. In addition, specific targets are assayed, so that redundant genomic information of spurious clinical worth is not generated. There is real and current potential for these types of analysis to be integrated into personalized medicine protocols.

We hope the reader is also persuaded of the potential importance of single-molecule sequencing — that it may revolutionize the field of genomics and, further, that it may significantly impact on the diagnosis and treatment of disease in the future. At present both NGS and SMS are research tools that can aid our understanding of disease and susceptibility to disease. Indeed, they may generate new molecular targets for the specific assays for which smPCR could then be employed. In this way the two broad areas discussed in this article could dovetail to influence clinical decision-making. In short, the future appears bright for single-molecule genomics but any foray into personalized medicine needs to be with a large dose of common sense.

Acknowledgement

The authors are employed by the UK Medical Research Council.

Teaching materials

PowerPoint slides of the figures from this review are supplied as supporting information in the online version of this article.

References

1. Dear PH. One by one: single molecule tools for genomics. *Brief Funct Genom Proteom* 2003;**1:**397–416.
2. Pohl G, Shih Ie M. Principle and applications of digital PCR. *Expert Rev Mol Diagn* 2004;**4:**41–47.
3. Bayley H. Sequencing single molecules of DNA. *Curr Opin Chem Biol* 2006;**10:**628–637.
4. Gupta PK. Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol* 2008;**26:**602–611.
5. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008;**26:**1135–1145.
6. Monckton DG, Jeffreys AJ. Minisatellite 'isoallele' discrimination in pseudohomozygotes by single molecule PCR and variant repeat mapping. *Genomics* 1991;**11:**465–467.
7. Dear PH, Cook PR. Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Res* 1993;**21:**13–20.
8. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci USA* 1999;**96:**9236–9241.
9. 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.
10. McCaughan F, Darai-Ramqvist E, Bankier AT, Konfortov BA, Foster N, George PJ, *et al*. Microdissection molecular copy-number counting (microMCC) — unlocking cancer archives with digital PCR. *J Pathol* 2008;**216:**307–316.
11. Bar-Eli M, Ahuja H, Gonzalez-Cadavid N, Foti A, Cline MJ. Analysis of N-RAS exon-1 mutations in myelodysplastic syndromes by polymerase chain reaction and direct sequencing. *Blood* 1989;**73:**281–283.
12. Fan X, Furnari FB, Cavenee WK, Castresana JS. Non-isotopic silver-stained SSCP is more sensitive than automated direct sequencing for the detection of *PTEN* mutations in a mixture of DNA extracted from normal and tumor cells. *Int J Oncol* 2001;**18:**1023–1026.
13. Oehler VG, Qin J, Ramakrishnan R, Facer G, Ananthnarayan S, Cummings C, *et al*. Absolute quantitative detection of ABL tyrosine kinase domain point mutations in chronic myeloid leukemia using a novel nanofluidic platform and mutation-specific PCR. *Leukemia* 2009;**23:**396–399.
14. Singer G, Oldt R III, Cohen Y, Wang BG, Sidransky D, Kurman RJ, *et al*. Mutations in *BRAF* and *KRAS* characterize the development of low-grade ovarian serous carcinoma. *J Natl Cancer Inst* 2003;**95:**484–486.
15. Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, *et al*. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* 2008;**135:**489–498.
16. Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res* 2009;**15:**2076–2084.
17. Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007;**16:**(spec. no. 1): R50–59.
18. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* 2009;**5:**401–408.
19. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer* 2004;**4:**707–717.
20. Kim WJ, Kim YJ. Epigenetic biomarkers in urothelial bladder cancer. *Expert Rev Mol Diagn* 2009;**9:**259–269.
21. Chhibber A, Schroeder BG. Single-molecule polymerase chain reaction reduces bias: application to DNA methylation analysis by bisulfite sequencing. *Anal Biochem* 2008;**377:**46–54.
22. Snell C, Krypuy M, Wong EM, Loughrey MB, Dobrovic A. BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res* 2008;**10:**R12.
23. Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, *et al*. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res* 2008;**36:**4689–4698.
24. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, *et al*. Detection of large-scale variation in the human genome. *Nat Genet* 2004;**36:**949–951.
25. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, *et al*. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005;**77:**78–88.
26. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, *et al*. Global variation in copy number in the human genome. *Nature* 2006;**444:**444–454.
27. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, *et al*. DNA duplication associated with Charcot–Marie–Tooth disease type 1A. *Cell* 1991;**66:**219–232.
28. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;**39:**S37–42.
29. King CR, Kraus MH, Aaronson SA. Amplification of a novel *v-erbB*-related gene in a human mammary carcinoma. *Science* 1985;**229:**974–976.
30. Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, *et al*. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 1983;**305:**245–248.
31. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, *et al*. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;**131:**18.
32. Hirsch FR, Varella-Garcia M, Cappuzzo F. Predictive value of *EGFR* and *HER2* overexpression in advanced non-small-cell lung cancer. *Oncogene* 2009;**28**(suppl 1):S32–37.
33. NICE. Early and locally advanced breast cancer, 2009; http://www.nice.org.uk/nicemedia/pdf/CG80QuickRefGuide.pdf [accessed 24 September 2009].
34. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006;**7:**85–97.
35. Kendall J, Liu Q, Bakleh A, Krasnitz A, Nguyen KC, Lakshmi B, *et al*. Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc Natl Acad Sci USA* 2007;**104:**16663–16668.
36. Qin J, Jones RC, Ramakrishnan R. Studying copy number variations using a nanofluidic platform. *Nucleic Acids Res* 2008;**36:**e116.
37. Lun FM, Chiu RW, Allen Chan KC, Yeung Leung T, Kin Lau T, Dennis Lo YM. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. *Clin Chem* 2008;**54:**1664–1672.
38. Bhat S, Herrmann J, Armishaw P, Corbisier P, Emslie KR. Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number. *Anal Bioanal Chem* 2009;**394:**457–467.
39. 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.
40. Fan HC, Blumenfeld YJ, El-Sayed YY, Chueh J, Quake Sr. Microfluidic digital PCR enables rapid prenatal diagnosis of fetal aneuploidy. *Am J Obstet Gynecol* 2009;**200:**543 e541–547.

41. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, *et al*. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 2005;**435:**43–57.
42. Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW, *et al*. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal Chem* 2008;**80:**8975–8981.
43. Dube S, Qin J, Ramakrishnan R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PLoS One* 2008;**3:**e2876.
44. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001;**409:**860–921.
45. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;**74:**5463–5467.
46. Ansorge WJ. Next-generation DNA sequencing techniques. *N Biotechnol* 2009;**25:**195–203.
47. White RA III, Blainey PC, Fan HC, Quake Sr. Digital PCR provides sensitive and absolute calibration for high-throughput sequencing. *BMC Genom* 2009;**10:**116.
48. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, *et al*. Single-molecule DNA sequencing of a viral genome. *Science* 2008;**320:**106–109.
49. Pushkarev D, Neff NF, Quake Sr. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 2009;**27:**847–852.
50. Jett JH, Keller RA, Martin JC, Marrone BL, Moyzis RK, Ratliff RL, *et al*. High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *J Biomol Struct Dyn* 1989;**7:**301–309.
51. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, *et al*. The potential and challenges of nanopore sequencing. *Nat Biotechnol* 2008;**26:**1146–1153.
52. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 2009;**4:**265–270.
53. Ding C, Cantor CR. Direct molecular haplotyping of long-range genomic DNA with M1–PCR. *Proc Natl Acad Sci USA* 2003;**100:**7449–7453.
54. Konfortov BA, Bankier AT, Dear PH. An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res* 2007;**35:**e6.
55. Xiao M, Wan E, Chu C, Hsueh WC, Cao Y, Kwok PY. Direct determination of haplotypes from single DNA molecules. *Nat Methods* 2009;**6:**199–201.
56. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 1998;**16:**652–656.
57. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci USA* 2003;**100:**8817–8822.
58. Mitra RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* 1999;**27:**e34.
59. Beer NR, Hindson BJ, Wheeler EK, Hall SB, Rose KA, Kennedy IM *et al*. On-chip, real-time, single-copy polymerase chain reaction in picoliter droplets. *Anal Chem* 2007;**79:**8471–8475.
60. Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci USA* 2003;**100:**5926–5931.
61. Zhang K, Zhu J, Shendure J, Porreca GJ, Aach JD, Mitra RD *et al*. Long-range polony haplotyping of individual human chromosome molecules. *Nat Genet* 2006;**38:**382–387.