J Fluoresc (2008) 18:1021–1026
DOI 10.1007/s10895-008-0386-2

SHORT COMMUNICATION

Single Molecule Genotyping by TIRF Microscopy

Steffen Rüttinger · Baptiste Lamarre · Alex E. Knight

Received: 11 February 2008 / Accepted: 9 May 2008 / Published online: 14 June 2008

Abstract As part of a programme to develop a metrological framework for single molecule measurements in biology, we have investigated the applications of single molecule imaging to genomics. Specifically, we have developed a technique for measuring the frequencies of single nucleotide polymorphisms (SNPs) in complex or pooled samples of DNA. We believe that this technique has applications to statistical genotyping—the identification of correlations between SNP frequencies and particular phenotypes—and other areas where it is desirable to track the frequencies of SNPs in complex DNA populations.

Keywords Single molecule imaging · Total internal reflection fluorescence microscopy · Statistical genotyping · Single nucleotide polymorphism

Introduction

It is estimated that one in every 1,000 bases in the human genome exhibits a single nucleotide polymorphism (SNP); i.e., within the human population there is a variability in the base present at that position [1]. Such variations may be present in the coding sequence of a gene, in which case they may be associated with a disease, or the susceptibility to a disease; or they may affect drug metabolism. Alternatively, similar effects may be caused when SNPs are present in the regulatory regions of genes. Detecting such polymorphisms is important in forensics, diagnostics, pharmacogenomics and disease risk assessment. In addition, the many SNPs located in non-coding regions of the genome may be used in linkage disequilibrium studies to identify genes involved in particular phenotypes, again including those responsible for disease.

There exists a wide variety of techniques for the detection of single nucleotide polymorphisms, comprehensively reviewed in [2]. Given the vast numbers of polymorphisms that are present in the human genome, and the numbers of samples that need to be characterised, there is an increasing requirement for very high-throughput genotyping techniques.

In most instances, where genotyping is applied to a DNA sample from an individual patient, there are usually three possible outcomes: the patient is homozygous for one or the other allele, or they are heterozygous. Therefore, while methods are required to be reliable enough to provide accurate calling of genotypes, they do not need to be truly quantitative. However, in some cases it is of interest to determine the frequencies of polymorphisms more precisely. For example, where one is looking for an association between a SNP and a phenotype, one may compare pooled samples of DNA from many individuals. The frequencies of alleles are then compared with the occurrence of the phenotype(s) of interest to identify any correlations. Such an investigation may be done for many different SNPs to identify those that are of interest [3].

Another example might be investigating mutants of a virus. For example, where a single base mutation in a viral genome is associated with a drug resistance phenotype, it

S. Rüttinger
Biomedical Optics Department,
Physikalisch-Technische Bundesanstalt,
Berlin, Germany

B. Lamarre · A. E. Knight (✉)
Biotechnology Group, National Physical Laboratory,
Hampton Road,
Teddington, Middlesex TW11 0LW, UK
e-mail: alex.knight@npl.co.uk

might be desirable to monitor the frequency of such an allele in a patient as treatment progresses.

We have developed a method for quantitative genotyping of SNPs, which counts the number of molecules of each allele in a sample. This method has a number of advantages over conventional "bulk" methodologies. Firstly, because it operates directly on the sample, no amplification step is required. This means that the process is simpler and no bias is introduced by the amplification step. Secondly, most fluorescence-based methods use two different probes, one for each allele of the polymorphism. Because each probe will have different characteristics, such as quantum yield and extinction coefficient; and because the instrument will respond differently to each probe (e.g. because of different excitation intensities, filter transmittances and detector sensitivities) some form of correction or normalisation of the measured fluorescence intensities is required to obtain accurate ratios of the two alleles. This normalisation process introduces additional uncertainties into the measurement. However, by counting the numbers of molecules labelled with each probe, these difficulties can be avoided. Thirdly, any fluorescence-based approach is susceptible to contamination from background signals. These could include the detector itself; contamination of the fluorescence with excitation light (due to the limitations of the optical system used); or unwanted fluorescence from materials (or their contaminants) used in the method. Because the fluorescence is measured as a single quantity, it is difficult to determine the contribution of the background signal. Furthermore, such a signal will contain a contribution of random noise from a variety of sources. These are not problems for the single molecule approach.

Single molecule measurements have many advantages and have been used to investigate many biological systems, such as molecular motors [4] and enzymes [5, 6]. Fluorescence imaging offers a versatile route to single molecule detection, and is particularly useful in this context, where one wishes to count the numbers of molecules in a sample.

The approach we describe consists of the following steps:

1. A labelling reaction is performed so that each allele in the sample has a different fluorescence label. In this case, we used a so-called minisequencing approach [2] where a primer is used to direct the incorporation of a single, chain terminating, fluorescently-labelled nucleotide at the SNP position.
2. The labelled DNA molecules are immobilised on the surface of a microscope slide to facilitate visualisation. The surface density must be such that the distance between adjacent molecules typically exceeds the resolution of the imaging system.
3. The labelled molecules are visualised using total internal reflection fluorescence (TIRF) microscopy. By using two lasers and two emission filters in sequence, the molecules labelled with each probe can separately be imaged.
4. The two images are processed and the numbers of molecules labelled with each probe are extracted. The frequency of each allele can then be calculated from the equation $f_A=n_A/(n_A+n_B)$, where f_A is the frequency of allele A; n_A is the number of molecules of allele A detected; n_B is similarly the number of molecules of allele B detected; and only these two alleles of the polymorphism are present.

We present an example of the above approach and then describe how it could be modified to form a high-throughput quantitative genotyping technique.

Methods and materials

Materials

Plasmid DNA was purified by using the Fastplasmid® Mini purification kit (Eppendorf, Hamburg, Germany) and concentration and purity determined by absorbance spectroscopy. The plasmids used were pBluescript® II KS(+) and pBluescript® II SK(+) (Stratagene, La Jolla, CA, USA).

The sequence of the primer used is given in Table 1. Primers were synthesised and purified by MWG (Ebersberg, Germany).

Labelling ("minisequencing") reactions were performed using a Perkin-Elmer (Waltham, MA, USA) Acyclo-Prime™-II SNP detection kit with a G/A terminator mix. This mix contains R110-acyGTP, TAMRA-acyATP, and two unlabelled terminators (where "acy" indicates the "acyclo" nucleotide analogues that are used as terminators). The kit also includes a mutated thermostable polymerase (AcycloPol™) and a reaction buffer.

Methods

Minisequencing reactions

Mixtures of 50 ng of plasmid DNA composed of known amounts of KS(+) and SK(+) were used as template for the minisequencing reactions. Reactions were performed as recommended by the manufacturer, except where otherwise stated. Reactions were performed in Thermo-Fast® 96 well plates (Abgene, Epsom, UK). Plates were sealed with adhesive plate seals (Abgene, Epsom, UK) and the minisequencing reactions were run on a GeneAmp PCR system 2700 (Applied Biosystems, Foster City, USA). The

Table 1 Plasmid and primer sequences

Plasmid	Sequence	Allele	Probe
pBluescript® II KS(+)	5′ TTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGG GCGAATTGG**A***GCTCCACCGCGGTGGCGGCCGCTCT 3′*	A	TAMRA
pBluescript® II SK(+)	5′ TTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGA ATTGG**G***TACCGGGCCCCCCCTCGAGGTCGAC 3′*	G	R110

Shown are the sequences at the boundary between the plasmid backbone (left) and the multiple cloning site (MCS, right). The sequence of the primer is indicated by a single underscore. The first base of the multiple cloning site is the model polymorphism in bold. The remaining MCS sequence is italicized. The primer anneals to the complementary sequence. Immediately 3′ to the primer-binding site is either a T (KS(+), polymerase incorporates an A) or a C (SK(+), polymerase incorporates a G).

programme used was as follows: 120 s at 95 °C, followed by 30 cycles of 15 s at 95 °C and 30 s at 55 °C, and a final 120 s at 4 °C. After the completion of the reactions, the samples were stored at −20 °C until required.

Preparation of samples for visualisation

To minimise photobleaching during observation, an "oxygen scavenger" solution was prepared. At working concentration, this consisted of 0.02 mg/ml catalase, 0.1 mg/ml glucose oxidase, 20 mM dithiothreitol (DTT) and 3 mg/ml glucose in 1× Tris–EDTA (TE) buffer pH 8. Sample reactions were diluted typically by a factor of 1:10 in TE followed by a further 1:10 dilution into the oxygen scavenger solution.

To immobilise the molecules for visualisation purposes, polylysine-coated microscope slides were used (Menzel Glaser, Braunschweig, Germany). Visualisation chambers were constructed from a slide, a 22×22 mm cover slip, and spacers made from strips of 10 μm thick PTFE film (Goodfellow, Cambridge, UK). After the addition of a 10 μl aliquot of sample, the cell was sealed using clear domestic nail varnish. After 5–10 min for the nail varnish to dry and DNA molecules to bind to the surface, the slide was placed on the microscope for visualisation. No washing step was performed.

Instrumentation

The system used for TIRF imaging of the labelled DNA molecules is shown in Fig. 1. The system is based on an Olympus (Tokyo, Japan) IX71 inverted microscope, equipped with an Andor (Belfast, UK) iXon DV887-BV EMCCD camera. A filter wheel on the emission port of the microscope contained interference filters (Semrock, Rochester, NY, USA) to select the wavelength range appropriate to the fluorescent probe. Shutters (Prior Scientific, Rockland, MA, USA) were used to select the appropriate laser. For visualisation of R110, excitation was with a 488 nm, 30 mW Sapphire laser (Coherent, Santa Clara, CA, USA) and the emission filter had a band centre of 536 nm with a 40 nm bandwidth. For TAMRA visualisation, a Coherent Compass 532 nm, 50 mW laser was used for excitation, with a 593 nm band pass filter with a 40 nm bandwidth.

Image sequences of 100 or 200 frames were collected with an exposure time of one frame per second. For genotyping experiments, the identical area of the specimen was imaged with both laser/filter combinations to enable both dyes to be counted.

Data analysis

Data acquisition and analysis were performed with software developed using MATLAB and the Image Processing and Signal Processing Toolboxes (The MathWorks, Natick, MA, USA). A number of algorithms are possible to detect and count the molecules present in the images. For demonstration purposes, we selected a relatively simple image-processing algorithm, as follows:

1. An average image was calculated for each image sequence.
2. A background image was calculated from the mean image, using a morphological opening operation with a "rolling ball" structuring element with a diameter larger than the expected size of a single molecule image.
3. The background image was subtracted from the mean image.
4. The image was thresholded to identify objects (groups of connected pixels). The threshold level was set manually to maximise the discrimination between objects and background.
5. The objects were filtered to remove objects that were too small or too large. Also, objects that overlapped in the R110 and TAMRA fluorescence images were removed.
6. The number of remaining objects in each image was counted.

For display purposes, the background subtracted images from step 3 above were contrast-enhanced (see Fig. 2).

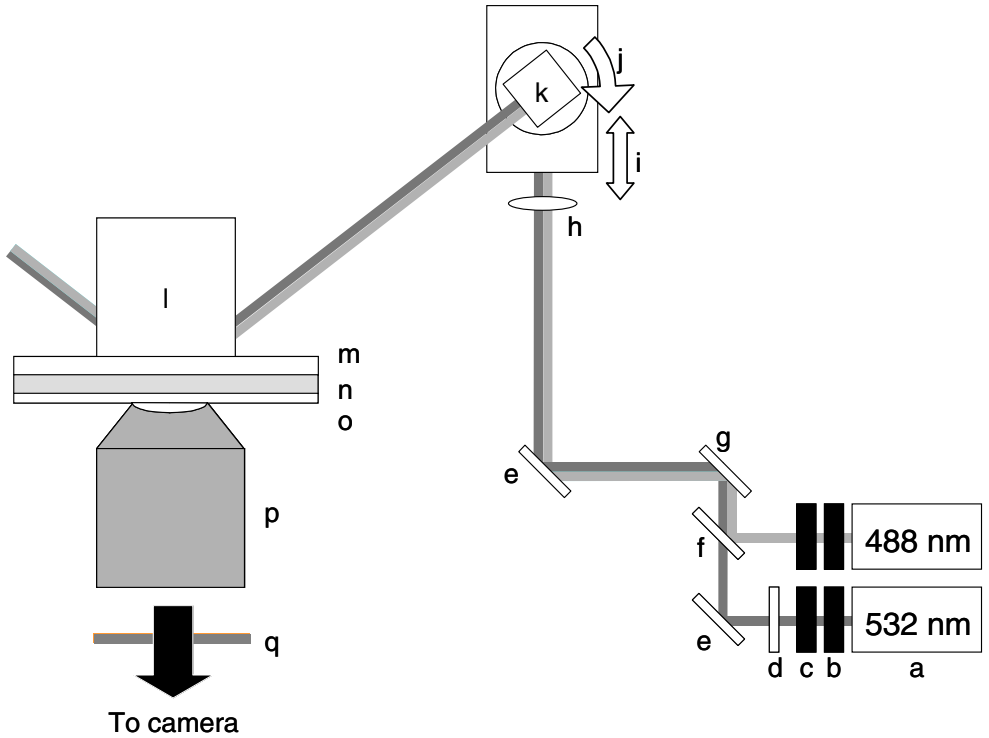

Fig. 1 Schematic of the TIRF apparatus. This is a schematic of the apparatus that was used in the experiments described. Fluorescence is excited using one of two lasers (*a*), the beams of which are combined using a series of dichroic mirrors (*f*, *g*) and a polarisation rotator (*d*). The beam is focussed to a spot at the viewing position using a lens (*h*). The angle of incidence of the beam is controlled using the translation and rotation stages (*i*, *j*) and pentaprism (*k*). A cubic quartz prism (*l*) is used to couple the laser beam into the microscope slide (*m*). Total internal reflection occurs at the interface between the slide and the sample medium (*n*) and generates an evanescent wave. Any fluorescence is captured using a high numerical aperture 100× 1.4 NA oil immersion objective lens (*p*) and the emission wavelength selected using an interference filter (*q*) before being imaged by an electron-multiplication CCD camera. For clarity, many beam-steering mirrors and other components have been omitted. (Other components shown include safety interlock shutters (*b*) and excitation selection shutters (*c*); mirrors (*e*), and cover slip (*o*)). See main text for more details

Results

To demonstrate the practicality of the method, we used test samples consisting of a mixture of known amounts of two plasmid DNA molecules, with a different base adjacent to the primer binding site. The results of analysing an equimolar mixture of the two templates (SK(+) and KS(+), corresponding to the G and A genotypes respectively) are shown in Fig. 2a–d. From these results, we can see that 41 R110-labelled (G allele) and 43 TAMRA-labelled (A allele) objects were counted. Therefore the relative frequency of the G allele is estimated as 49%. For a control experiment containing 100% of the KS(+), (A allele), the counts were 12 R110-labelled objects and 156 TAMRA-labelled objects. In this case, the frequency of the A allele is estimated as 93%. Similar results were obtained for other ratios of the alleles (not shown).

Discussion

As expected, for the equimolar mixture of alleles, approximately equal numbers of spots of each colour are visible in the two images (Fig. 2a,b) and identified by the image analysis (Fig. 2c,d). In the control experiment, there are many more spots corresponding to the A allele. This is in line with expectations, although there is a significant background level (7%) of the G allele. This could be due to misincorporation in the minisequencing reaction, or to background fluorescence being misinterpreted as R110 molecules.

The number of molecules counted would be expected to follow a Poisson distribution, where the standard deviation is equal to the root of the number of objects. Therefore, by increasing the number of objects counted, the confidence in the value obtained can be increased. This could be achieved by collecting larger images, or collecting more images from other fields of view; alternatively, the surface density of molecules could be increased, although this would increase the frequency of overlapping molecules. Where rare alleles are to be quantitated, it will therefore be obvious that more molecules will have to be counted to give the same level of confidence in the frequency.

While these results are a proof-of-principle demonstration, the method described here could be improved in a

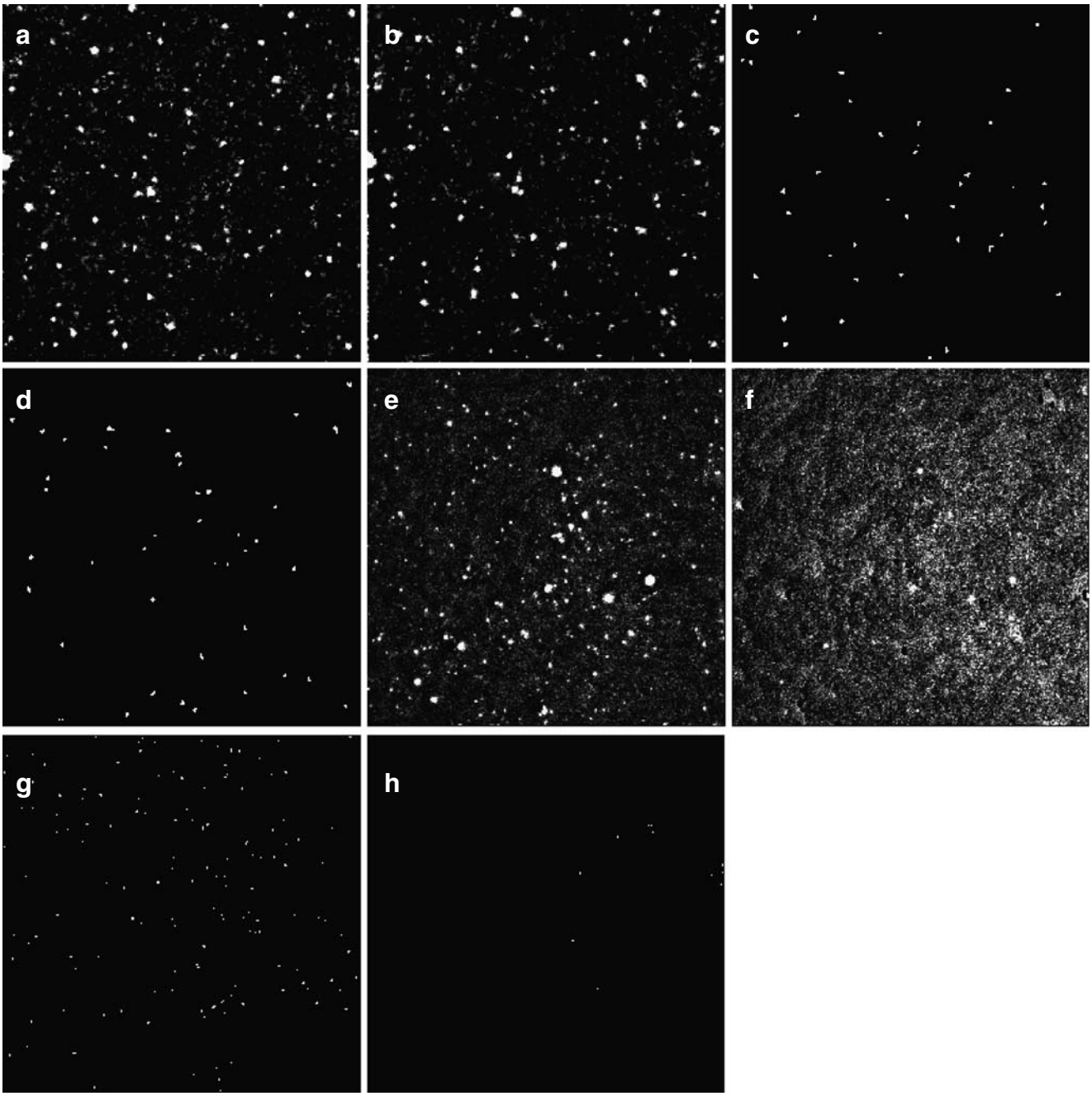

Fig. 2 Single molecule fluorescence images of labelled DNA molecules. Two experiments are shown: **a**–**d** an equimolar (1:1) mix of the SK(+) and KS(+) forms of the plasmid; **e**–**h** a control with KS (+) only. **a** TAMRA fluorescence image and **b** R110 fluorescence image. Note the roughly equal numbers of spots in each image; typically, each spot corresponds to a single molecule. Some of the larger spots may be aggregates or contaminating particles and are excluded from the analysis, as are spots that occur in both channels. The contrast in each image has been enhanced for clarity. **c** and **d** Results of feature analysis of the images in **a** and **b**. The images were thresholded and objects consisting of groups of connected pixels were analysed. Objects which contained too few, or too many, pixels to be consistent with the point spread function were excluded, as were any objects which shared pixels with an object on the other channel. The remaining objects were then counted. In this case, 41 R110-labelled and 43 TAMRA-labelled objects were identified. **e** TAMRA and **f** R110 images from control experiment. Note that **e** resembles **a** and **b** in that many spots of the expected size are present. In **f** the contrast is enhanced so that the background signal is visible. It is clear that very few spots of the expected size are present; the large spots that are seen are likely to represent aggregates and are visible in both channels. In the corresponding post-analysis images **g** and **h**, 156 and 12 objects are identified, respectively. The images are 41 μm (256 pixels) square and represent an average of 100 frames with each filter. The exposure time per frame was 1 s

number of ways. Firstly, long acquisition times were required due to a high background signal. This may have arisen from the polylysine surfaces used to immobilise the DNA molecules, from non-specific binding of fluorophore or from fluorophore free in solution. If this background signal could be reduced, for example by the use of amine-modified surfaces, or the inclusion of a wash step, acquisition times could be reduced by at least an order of magnitude. This should also decrease the level of background counts observed, improving the accuracy of the method.

Secondly, the image processing and analysis algorithms could be improved to give greater automation and enhanced counting accuracy.

Thirdly, the method could be adapted into a scaleable high-throughput format, where many SNPs could be quantitated simultaneously for each sample. This would be achieved by adapting the method to a microarray format.

This method would operate as follows:

1. The primers corresponding to each SNP of interest would be printed as a DNA microarray.
2. The template DNA sample would be incubated with the array so that it hybridised to the complementary primers.
3. The labelling reaction would then be performed *in situ* on the array, by adding labelled and unlabelled terminators and polymerase.
4. After washing as necessary, the array would then be visualised as described previously.
5. Image processing would then be used to count the numbers of molecules of each allele in each printed spot of the array.

The last step would require that each spot on the array be visualised individually, so acquisition times are critical to allow an array of many spots to be scanned in a reasonable time. This would also require the spot positions to be known with considerable accuracy, although a typical spot is likely to have a diameter several times greater than the field of view of the instrument. Alternatively, the array could be visualised using the synchronous scanning approach which has recently been demonstrated for expression arrays [7].

In summary, we believe that the principle we have described could be developed into a high-throughput method for obtaining quantitative allele frequencies for single nucleotide polymorphisms in un-amplified samples, although this would require increased accuracy of counting and an improved understanding of the sources of uncertainty in the measurement. This would be of benefit to studies that attempt to identify SNPs that correlate with phenotypes of interest, and demonstrates once again the power of single molecule approaches to biological measurement.

Acknowledgements This work was funded by the Strategic Research Programme of the National Physical Laboratory. SR was supported by a secondment under the Measurement for Innovators programme of the National Measurement System, funded by the Department for Innovation, Universities and Skills. We would like to thank Prof. Maurice Cox for the helpful discussions regarding data analysis, and Drs. Rainer Macdonald and Marc Bailey for facilitating the secondment.

References

1. Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2(12):930–942
2. Syvänen AC (1999) From gels to chips: “minisequencing” primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 13(1):1–10
3. Nowotny P, Kwon JM, Goate AM (2001) SNP analysis to dissect human traits. Curr Opin Neurobiol 11(5):637–641
4. Knight A, Mashanov G, Molloy J (2005) Single molecule measurements and biological motors. Eur Biophys J 35(1):89
5. English BP, Min W, van Oijen AM, Lee KT, Luo G, Sun H, Cherayil BJ, Kou SC, Xie XS (2006) Ever-fluctuating single enzyme molecules: Michaelis–Menten equation revisited. Nat Chem Biol 2(2):87
6. Xie XS (2001) Single-molecule approach to enzymology. Single Molecules 2(4):229–236
7. Hesse J, Jacak J, Kasper M, Regl G, Eichberger T, Winklmayr M, Aberger F, Sonnleitner M, Schlapak R, Howorka S, Muresan L, Frischauf AM, Schütz GJ (2006) RNA expression profiling at the single molecule level. Genome Res 16(8):1041–1045