

---

# High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking

G. Ullman, M. Wallden, E. G. Marklund, A. Mahmutovic, Ivan Razinkov and J. Elf

*Phil. Trans. R. Soc. B* 2013 **368**, 20120025, published online 24 December 2012

---

## Supplementary data

["Data Supplement"](#)

<http://rstb.royalsocietypublishing.org/content/suppl/2012/12/14/rstb.2012.0025.DC1.html>

## References

[This article cites 15 articles, 7 of which can be accessed free](#)

<http://rstb.royalsocietypublishing.org/content/368/1611/20120025.full.html#ref-list-1>

[Article cited in:](#)

<http://rstb.royalsocietypublishing.org/content/368/1611/20120025.full.html#related-urls>

## Subject collections

Articles on similar topics can be found in the following collections

[biophysics](#) (57 articles)  
[cellular biology](#) (112 articles)  
[microbiology](#) (25 articles)  
[molecular biology](#) (132 articles)  
[systems biology](#) (61 articles)

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

[rstb.royalsocietypublishing.org](http://rstb.royalsocietypublishing.org)



## Research

**Cite this article:** Ullman G, Wallden M, Marklund EG, Mahmutovic A, Razinkov I, Elf J. 2013 High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Phil Trans R Soc B* 368: 20120025. <http://dx.doi.org/10.1098/rstb.2012.0025>

One contribution of 12 to a Theme Issue 'Single molecule cellular biophysics: combining physics, biochemistry and cell biology to study the individual molecules of life'.

### Subject Areas:

biophysics, microbiology, molecular biology, systems biology, cellular biology

### Keywords:

microfluidics, single molecule, image analysis, cell tracking, gene expression, growth

### Author for correspondence:

J. Elf  
e-mail: [johan.elf@icm.uu.se](mailto:johan.elf@icm.uu.se)

†These authors contributed equally to this study.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2012.0025> or via <http://rstb.royalsocietypublishing.org>.

# High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking

G. Ullman<sup>1,2,†</sup>, M. Wallden<sup>1,†</sup>, E. G. Marklund<sup>1</sup>, A. Mahmutovic<sup>1</sup>, Ivan Razinkov<sup>3</sup> and J. Elf<sup>1</sup>

<sup>1</sup>Department of Cell and Molecular Biology, Science for Life Laboratory, and <sup>2</sup>Division of Scientific Computing, Department of Information Technology, Uppsala University, 75105 Uppsala, Sweden

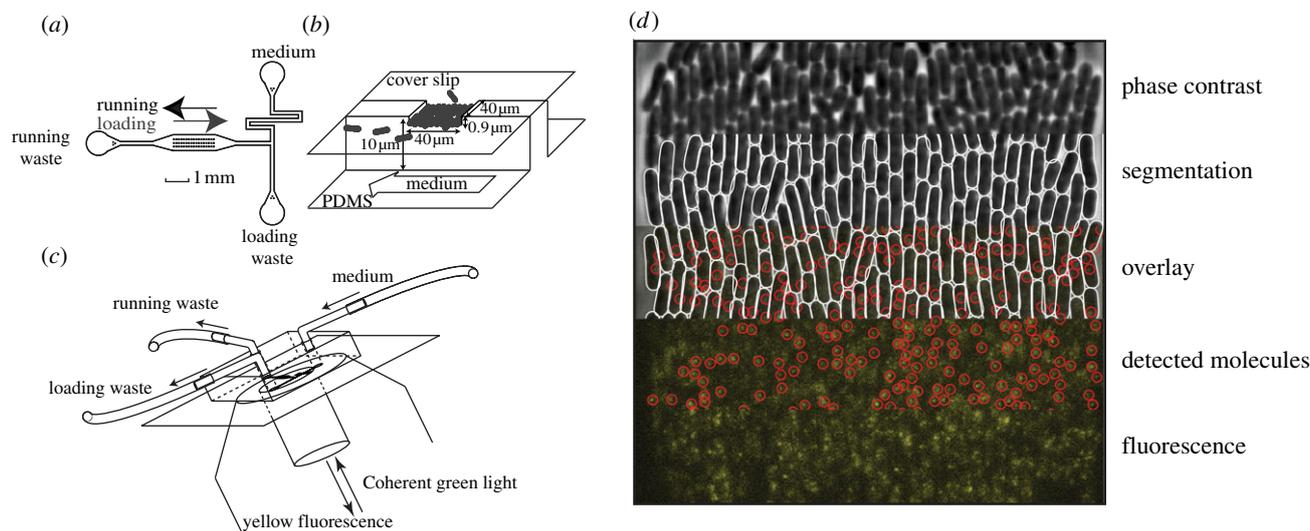
<sup>3</sup>Department of Bioengineering, University of California, La Jolla, San Diego, , CA 92093, USA

We have developed a method combining microfluidics, time-lapsed single-molecule microscopy and automated image analysis allowing for the observation of an excess of 3000 complete cell cycles of exponentially growing *Escherichia coli* cells per experiment. The method makes it possible to analyse the rate of gene expression at the level of single proteins over the bacterial cell cycle. We also demonstrate that it is possible to count the number of non-specifically DNA binding LacI–Venus molecules using short excitation light pulses. The transcription factors are localized on the nucleoids in the cell and appear to be uniformly distributed on chromosomal DNA. An increase in the expression of LacI is observed at the beginning of the cell cycle, possibly because some gene copies are de-repressed as a result of partitioning inequalities at cell division. Finally, a size–growth rate uncertainty relation is observed where cells living in rich media vary more in the length at birth than in generation time, and the opposite is true for cells living in poorer media.

## 1. Introduction

Using time-lapsed phase-contrast and fluorescence microscopy, it is possible to monitor live bacterial cells and simultaneously quantify the expression of their highly expressed genes as the activity of introduced fluorescence reporters [1]. However, for many of its native protein species, a bacterial cell expresses only a few copies per generation [2]. In order to study processes involving these proteins, fluorescence microscopy methods sufficiently sensitive to resolve individual molecules have been developed. For instance, Yu *et al.* [3] reported on the use of a fast maturing yellow fluorescent protein (YFP) variant, Venus [4], fused to a membrane tag, Tsr, to profile the absolute expression of the *lacZ* gene, in live *Escherichia coli* cells, in its repressed state. The Tsr domain immobilizes the fluorophore at the membrane so that it appears stationary for periods of 50–100 ms and can be detected as a diffraction-limited spot. However, tethering to the membrane will disable molecules that rely on intracellular mobility for their function. For this reason, methods for counting expression events for cytoplasmic proteins are limited. A possible solution is suggested by the single-molecule tracking experiments where stroboscopic illumination pulses were used to image the transcription factor LacI–Venus non-specifically bound to DNA in live *E. coli* cells [5]. This suggests that short excitation pulses could be used also to profile the synthesis of cytoplasmic low copy number transcription factors or other proteins binding to relatively immobile intracellular targets.

Single-protein counting experiments *in vivo* reveal that isogenic cells under seemingly identical experimental conditions display considerable diversity in expression [6]. In order to confidently draw conclusions on the nature of this



**Figure 1.** The experimental set-up and data processing. (a) The microfluidic device has three ports designated for medium, running waste and loading waste. The chamber houses three rows, each containing 17 traps. The direction of the flow through the chamber is alternated between the loading and running phase of the experiment. The cells are introduced from the running waste and are caught in the traps. (b) Each trap is a  $40 \times 40 \times 0.9 \mu\text{m}$  compartment which is bounded by two rigid walls and two openings. Cells that reach the openings are released from the traps into the  $10 \mu\text{m}$  deep surrounding. (c) The device is connected to reservoirs at the ports and imaged using an inverted microscope. The various parts of the microfluidic chip are not drawn in scale. (d) Data processing: cells are detected and segmented from the phase-contrast image (top). Molecules are detected within the fluorescence images (bottom). The coordinates from the detected molecules and cells are used to map molecules to cells (middle).

diversity, it is necessary to sample a sufficient number of cells. Several microfluidic devices have been reported to substantially increase experimental throughput by harnessing the reproduction of bacterial cells to continuously regenerate the sample and also allowing imaging of many replicate colonies in parallel [7,8]. However, the sheer size of image datasets that can be generated in this fashion overwhelms manual analysis efforts and consequently several initiatives of automation have been undertaken [9,10]. In this study, we report on a method combining microfluidics, single-molecule fluorescence microscopy and automated image analysis, enabling the study of the expression and super-resolution localization of low copy number transcription factors throughout thousands of bacterial lifespans per experiment. To illustrate the performance of the method, we quantify the dynamics of synthesis and intracellular localization of the lactose repressor by monitoring LacI–Venus expressed from its native promoter in live *E. coli* cells. We compare these observations with those obtained under identical conditions for cells expressing the reporter construct Tsr–Venus from the lactose permease gene, *lacY*, of the lactose operon.

## 2. Material and methods

### (a) Design, fabrication and use of the microfluidic device

The chip design was inspired by Mather *et al.* [11]. The features of the microfluidic chip used in this study were designed in three layers using AUTOCAD. The layers correspond to structures of different step heights of the mould and ultimately to the different depths of the structures of the finished microfluidic device (described under ‘mould fabrication’ and ‘chip fabrication’). The device contains four structural motifs: ports, channels, a chamber and traps (figure 1a). The chamber houses three

evenly spaced rows, each containing 17 traps (figure 1a). Each trap is  $40 \times 40 \times 0.9 \mu\text{m}$  (figure 1b), and is bounded by two opposite walls and two open sides connecting the trap to the  $10 \mu\text{m}$  deep surrounding. This geometry restricts the cells to form a monolayer colony in the focal plane while imaging. Cells close to the openings are released as the colony expands (figure 1b). The microfluidic device is connected to media reservoirs and imaged using an inverted microscope (figure 1c).

The master mould was fabricated using standard UV-soft lithography techniques. Three masks for microfabrication were printed in chrome. Custom formulations of SU8 Photoresist (MicroChem) were deposited on clean polished silicon wafers (University Wafer) using a spin coater. The wafers were then aligned to the mask and exposed using a mask aligner (Süss MA6). This process was repeated to deposit layers of step heights 0.9, 2.7 and  $10 \mu\text{m}$  per wafer. The first layer corresponds to the trap depth of the microfluidic device; the intermediate layer enables the alignment of the first and third layer, corresponding to the channels and ports. Each layer of the moulds was measured using a stylus profilometer and inspected under a microscope before applying the next.

A master cast of the mould was made from polydimethylsiloxane (Sylgaard 184, Dow Corning), using the master mould. Bubbles were removed by vacuum desiccation. The cast was cured at  $80^\circ\text{C}$  for 30 min. One master cast contained 12 identical chip structures, which could be excised and used individually.

When fabricating each device, port holes ( $0.5 \text{ mm}$  diameter) were punched out of the device cast. Debris was removed from the cast by vortexing in ethanol. The chip cast was bonded to a coverslip ( $40 \text{ mm}$  diameter,  $200 \mu\text{m}$  thick, Thermo-Scientific) after oxygen/UV plasma treatment (UVO-cleaner 42–220, Jelight Co.) for 5 min at 0.5 bar oxygen pressure. The bond was stabilized by incubating at  $80^\circ\text{C}$  for 10 min. Just prior to loading and running the device, the ports were treated with a high-frequency generator (model BA 20 D, Electro-Technic Products Inc.), and the device was flooded with de-ionized water.

Gravity flow was used to control the direction and the magnitude of the flow inside the microfluidic device. The pressure gradients between the different ports of the device were established by differences in elevation relative to the sample of the

connected reservoirs. During loading, the seeding culture was introduced into the device through the running waste port. The cells were caught in the traps by introducing pressure waves into the tubing. Once all traps were sufficiently occupied (10–100 cells per trap), the direction of the flow in the chamber was reversed, exchanging the seeding culture with fresh medium (figure 1a). The cells were allowed to acclimatize and grow until the traps were fully occupied (approx. 4 h) before imaging. The temperature of the sample was maintained at 37°C using a custom-fitted incubator hood (OKO LAB).

### (b) Strains and medium

Two bacterial strains, SX701 and JE116, based on *E. coli* strain BW25993 [12], were used in this study. In strain SX701, the lactose permease gene, *lacY*, was replaced with the *tsr-venus* construct [13]. Strain JE116 is based on strain JE12 [5], in which the *lacI* gene was modified to encode a C-terminal fusion of LacI and Venus. The auxiliary lactose operator site, *O3*, was replaced with the main operator sequence, *O1*, to increase auto-repression by LacI threefold. Further, in strain JE116 the downstream sequence including the native *O1*, *O2* binding sites as well as parts of the *lacZ* gene was removed, leaving only one specific binding site sequence for LacI–Venus molecules per chromosome copy [14].

Cells were grown in M9 minimal medium, with 0.4 per cent glucose, either with or without supplemented amino acids (RPMI1640 (R7131), Sigma–Aldrich). An overnight culture was diluted 200 times in 40 ml fresh medium and incubated for 3–5 h (6–8 h for cells grown without amino acids) at 37°C and shaking at 225 rpm. During this incubation, the microfluidic device was prepared. Cells were harvested into a seeding culture by centrifugation at  $5000 \times \text{rcf}$  for 2.5 min and the pellet resuspended in 50–100  $\mu\text{l}$  fresh medium. In order to prevent the cells from sticking to the surfaces of the microfluidic device a surfactant, Pluronic F108 (prod. Number 542342, Sigma–Aldrich) was added to all media to a final concentration of  $0.85 \text{ g l}^{-1}$ .

### (c) Microscopy and imaging

Imaging was performed using an inverted microscope (Ti Eclipse, Nikon) fitted with a high numerical aperture oil objective (APO TIRF 100 $\times$ /N.A 1.49, Nikon) and external phase contrast to minimize loss of fluorescence signal. The phase-contrast channel and the fluorescence channels were imaged using separate cameras, a model CFW-1312M (Scion Corporation) and an Ixon EM plus (Andor Technologies), respectively. Focus was maintained by the *perfect-focusing-system* of the microscope. The light source for fluorescence excitation was an Argon ion laser (Innova 300, Coherent Inc.) dialled to 514 nm for excitation of YFP-reporters in the sample. For fluorescence imaging, a slower shutter (LS6Z2, Uniblitz) was used for strain SX701 (Tsr–Venus) and a fast shutter (LS2Z2, Uniblitz) was used for strain JE116 (LacI–Venus). The fast shutter was controlled using a signal generator (AFG3021B, Tektronix), which was triggered by the Ixon camera, exposing the sample for 1 ms. A 2 $\times$  magnification lens was used in the fluorescence emission path to distribute the point spread function ideally on the 16  $\mu\text{m}$  pixels of the EMCCD. Image acquisition was performed using RITACQUIRE, an in house GUI-based plugin for MICRO-MANAGER (v. 1.3.4.7, [www.micro-manager.org](http://www.micro-manager.org)). In each experiment, three positions (traps) were subjected to the following acquisition program in parallel: every 30 s (every frame), a phase-contrast image (125 ms exposure) was taken for all positions. Every 3 min (1/6 frames) for all positions, in addition to the phase-contrast image, two fluorescence images (50 ms exposure for SX701 and 1 ms exposure for JE116) were taken in rapid succession, followed by a bright field image (100 ms exposure) of the fluorescence channel, i.e. using the white-light lamp of the

microscope as illumination source. This programming cycle was repeated for 1001 frames (8.3 h). Fluorescence images were acquired in tandem to account for the effects of bleaching on molecular counting (see §2f). The bright field images were acquired to allow alignment of phase-contrast and fluorescence images for each frame. Our automatic method for cropping the phase images and aligning them to the fluorescent images is described in the electronic supplementary material, methods.

### (d) Cell segmentation and tracking

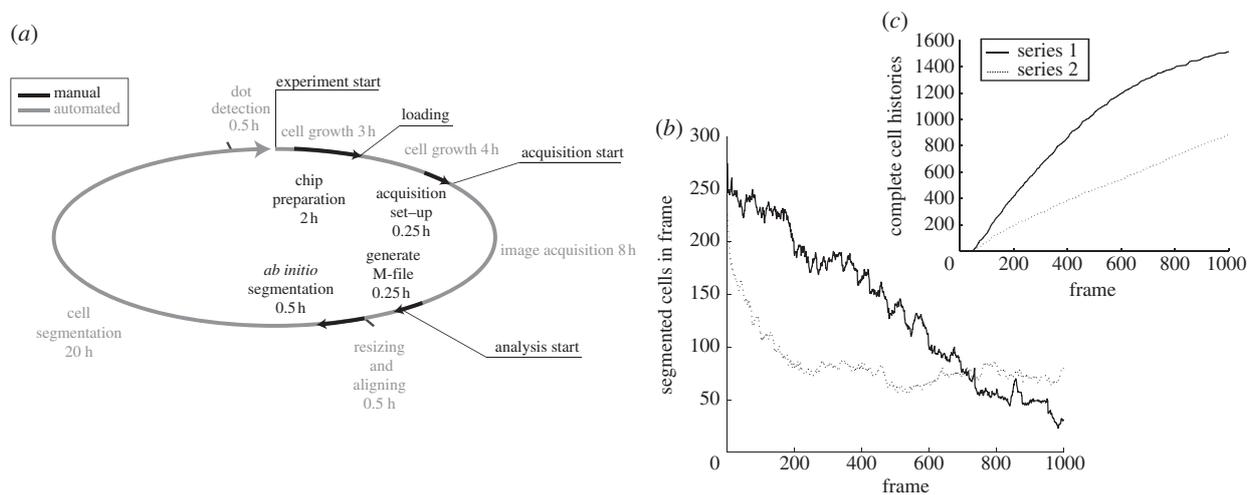
For segmenting and tracking individual cells in the microfluidic device, we have modified and further developed existing MATLAB software, MICROBETRACKER [10]. MICROBETRACKER uses the position of cells in the previous frame as an initial guess and applies an active contour model [15] to fit each cell with a sub-pixel resolution boundary. In order to accurately track mobile cells over several generations, three additional supervised algorithms [16] were implemented in MATLAB to complement MICROBETRACKER (see the electronic supplementary material, Methods): a *cell pole tracker* and two separate *error detectors*. The cell pole tracker is used to help the active contour model find the cell poles correctly for moving cells; otherwise this will lead to error propagation in the subsequent frames. The first error detector identifies errors made by the cell pole tracker. This is usually the result of an occasional large displacement of the cell between frames. This activates the cell tracker, which attempts to correct the segmentation of the erroneous cell. The accuracy of the cell tracker is in turn monitored by a second error detector. Any cell histories triggering this detector are terminated. In addition, a novel *division function* was added to MICROBETRACKER in order to more accurately detect cell divisions for densely growing *E. coli*. Each supervised algorithm was constructed by first identifying features that efficiently discriminate between two classes, for instance, true or false cell division. In the second step, training data were extracted manually from the image sets for creating training examples for the algorithm in order to achieve accurate classification. A linear classifier [16] was used in all supervised algorithms. The algorithms, cell tracker and the classification method are described in detail in the supplementary methods. To increase the computational speed, parts of MICROBETRACKER were rewritten to allow parallel computing, using MATLAB's parallel computing toolbox.

### (e) Single molecule detection, localization

Fluorescent particles in the sample were detected as diffraction limited spots in the fluorescence micrographs according to the method described in Ronneberger *et al.* [17], in which the normalized cross-correlation between the fluorescence image and an idealized optical point spread function (a symmetric bi-variate Gaussian function) is calculated. The standard deviation (s.d.) for this function is obtained experimentally by imaging and the signatures of immobilized highly fluorescent beads (data not shown). The image resulting from the correlation is transformed using the Fisher transform. A Fisher transformed Gaussian function with s.d. corresponding to the point spread function is fitted to the Fisher transformed correlation image using the Levenberg–Marquardt method [18] implemented in MATLAB, and the obtained parameters are used to localize each molecule with super-resolution accuracy and estimate the localization error.

### (f) Maximum-likelihood estimate of synthesis

For gene expression studies, we want to estimate how many molecules have been newly synthesized between two fluorescence images given that there is a chance that some of the fluorophores present in the previous frame have not been bleached. We formulate this as a maximum-likelihood problem



**Figure 2.** Throughput of the method. (a) Time distribution for an experiment. The minimum time expenditure for an experiment is around 36 h. Currently, the method generates around 3000 complete cell generations per experiment. The protocol contains mostly automated steps, i.e. they require no attention from the operator. The manual time expenditure accounts for less than 10% of the total time and less than 2% of the analysis time. (b) The number of segmented cells remaining at different frames during the analysis of two different time-series. Owing to large movements in the cell colony, the loss of correctly segmented cells varies between series. The nature of the decay is observed to depend on the pattern in which the cells grow in the trap, for which no sufficiently accurate prediction model has yet been found. (c) The integrated numbers of cells acquired from cell division to cell division, i.e. the number of complete cell histories, for the two series in (b).

where there are  $M$  molecules observed in frame  $i - 1$  and  $N$  molecules observed in frame  $i$ . The number of molecules surviving bleaching,  $m$ , can be calculated by maximizing the probability

$$p(m|M, N, p, \lambda) = \text{Bin}(m, M, 1 - p) \cdot \text{Po}(N - m, \lambda),$$

where Bin is the binomial distribution and Po is the Poisson distribution. The maximum-likelihood estimate of the number of new synthesized molecules is  $n_{\max} = N - m_{\max}$ , where  $m_{\max}$  maximizes  $p(m|N, M, p, \lambda)$ . The parameter  $p$  is the bleaching probability per fluorophore per frame and is assumed to be constant.  $\lambda$  is the number of molecules synthesized between two frames.

In the special case of cell division between frame  $i - 1$  and  $i$ , where  $N1$  molecules are found in one daughter cell and  $N2$  in the other, the most likely number of newly synthesized molecules  $n_{\max}$  are calculated for both cells based on  $N = N1 + N2$ . Given  $n_{\max}$  the most likely number of newly synthesized molecules in daughter cell 1 is the  $n1$  that maximizes  $\binom{N1}{n1} \binom{N2}{n_{\max} - n1}$  because this gives the number of possible combinations of picking  $n1$  molecules from  $N1$  and  $n - n1$  from  $N2$ .

### (g) Availability

All programs and scripts developed for this study will be made available on request.

## 3. Results

### (a) Throughput

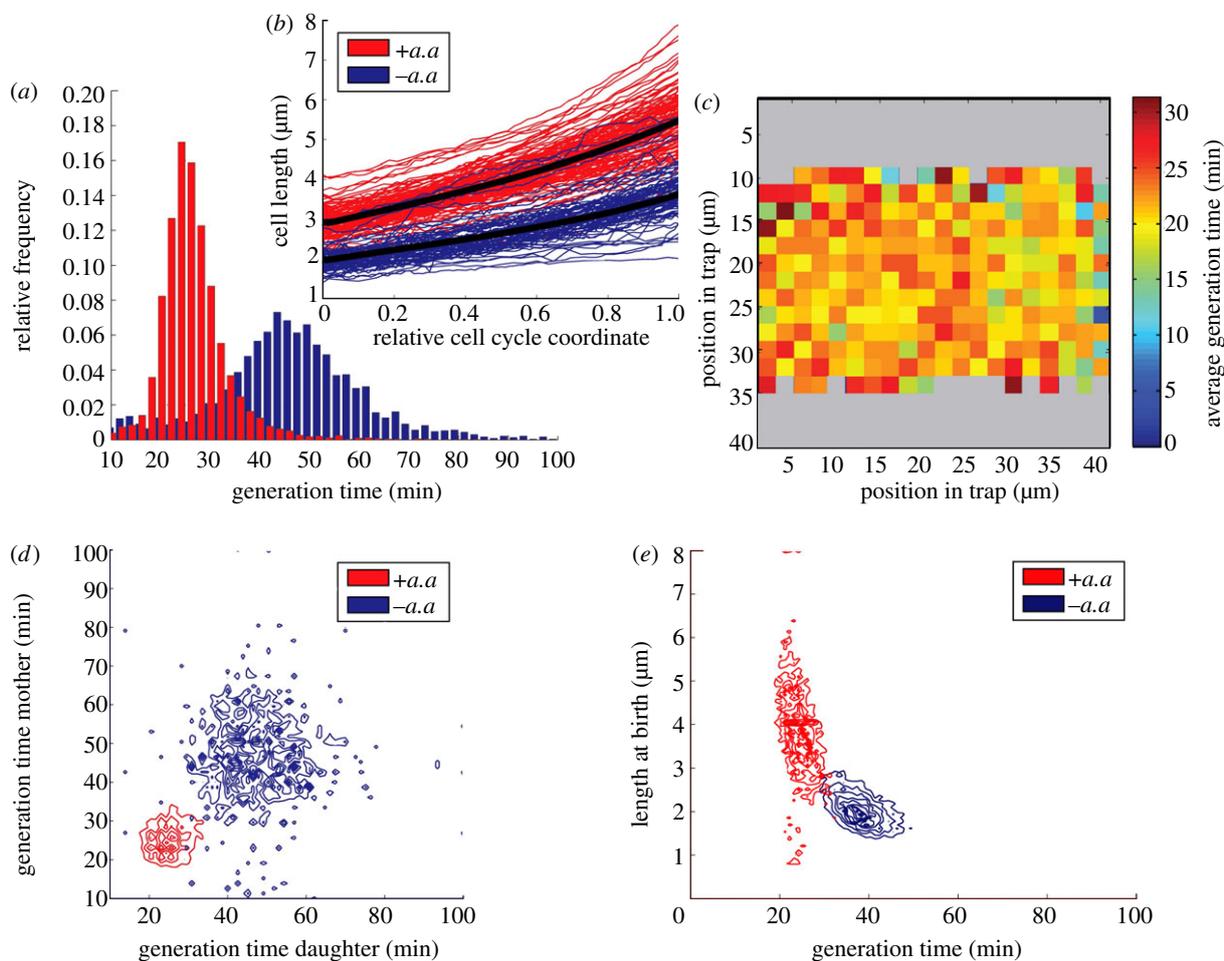
Currently, one experiment returns approximately 3000 complete cell histories from three traps imaged in parallel. The total time of expenditure is 36 h. The manual effort of a single operator amounts to 3 h, of which roughly 80 per cent is spent prior to image acquisition. The manual work effort to acquire and analyse the images constitutes less than 2 per cent of the total time required to complete these processes (figure 2a). Several overlapping experiments can

be performed to use the alternating availability of the microscope and the computational framework to further improve throughput. The number of cell histories acquired from an image series is determined in the segmentation process.

The cells sometimes make large displacements between two frames. When the cell tracker fails to track the cell, the cell history is terminated. Therefore, the number of cells that the program keeps track of decreases over time. The rate of decay varies considerably between image series, even when acquired under seemingly identical conditions (figure 2b). Only the set of cell histories that completely cover the time from division-to-division enter the analysis (figure 2c).

### (b) Morphology and growth in the microfluidic device

The generation time defines the growth rate of exponentially growing cells and is often used as an indicator of the health or fitness. We compare cells grown with and without amino acids in the medium (figure 3a, red and blue, respectively) and observe average generation times of  $26.4 \pm 7.2$  and  $46.8 \pm 17.0$  min, respectively. Further, we observe an exponential growth of the cell length over the cell cycle (figure 3b). In contrast to previous reports [11], we observe no obvious dependencies of the growth rate on the position the cell occupied in the trap (figure 3c). This uniformity also holds for morphology and bacterial age, i.e. the number of divisions during which the oldest pole of a cell has been observed. We find that the generation times of mother and daughter cells are weakly correlated ( $r = 0.27 \pm 0.02$  with amino acids,  $r = 0.07 \pm 0.05$  without amino acids; figure 3d). The relation between the length at birth and the generation time of a cell history displays a correlation (figure 3e), indicating that comparatively longer newborns complete their cell division faster. Although this holds qualitatively for cells grown both with and without supplemented amino acids (red and blue), it is less pronounced for cells grown without amino acids. Also, cells grown with amino acids vary more in length at birth than in generation time and the opposite is observed



**Figure 3.** Morphology and growth of cells in the microfluidic device. (a) The distribution of generation times of cells grown with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (b) Cell length as a function of relative cell cycle coordinates, i.e. time from birth normalized by the generation time, for randomly selected cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (c) Growth rate as indicated by average generation time over the geometry of the trap for cells growing with supplemented amino acids ( $n = 6755$ ). The figure is oriented so that the outlets of the trap are on top and bottom (figure 1b). (d) Joint distribution of generation times for daughter and mother cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids. (e) Joint distribution of generation time and cell length at birth for cells with (red,  $n = 6755$ ) and without (blue,  $n = 2298$ ) supplemented amino acids.

for cells grown without supplemented amino acids. The correlation for cells with amino acids is  $r = -0.43 \pm 0.02$  and without amino acids  $r = -0.28 \pm 0.04$ . No significant differences in growth or morphology between strains SX701 and JE116 are observed.

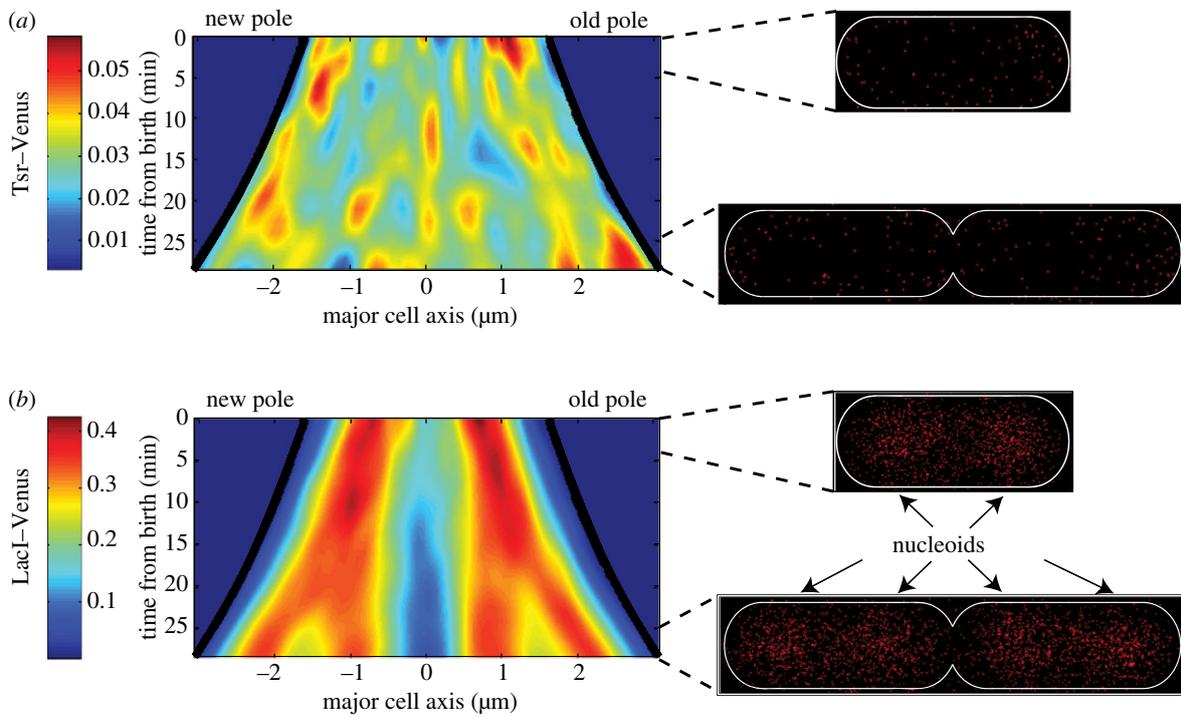
### (c) Localization of transcription factors during the cell cycle

In figure 4, we compare the intracellular localization of the reporter constructs, Tsr-Venus (figure 4a) and LacI-Venus (figure 4b), over the cell cycle. A localization distribution function (figure 4, left) is constructed by mapping the detected molecules to their position along the major axis of the cell ( $x$ -axis) at the time in the cell cycle they were detected ( $y$ -axis) and smoothed using a Gaussian filter. To increase synchronicity, only observations occurring in cells with generation times between 25 and 32 min and terminal lengths of 4–7  $\mu\text{m}$  are included (780 for SX701 and 1176 for JE116). In the right of figure 4, we visualize the detected molecules of each construct as bi-variate symmetric Gaussian functions to create a PALM style super-resolution plot of the intracellular distribution. We do not observe the typical polar

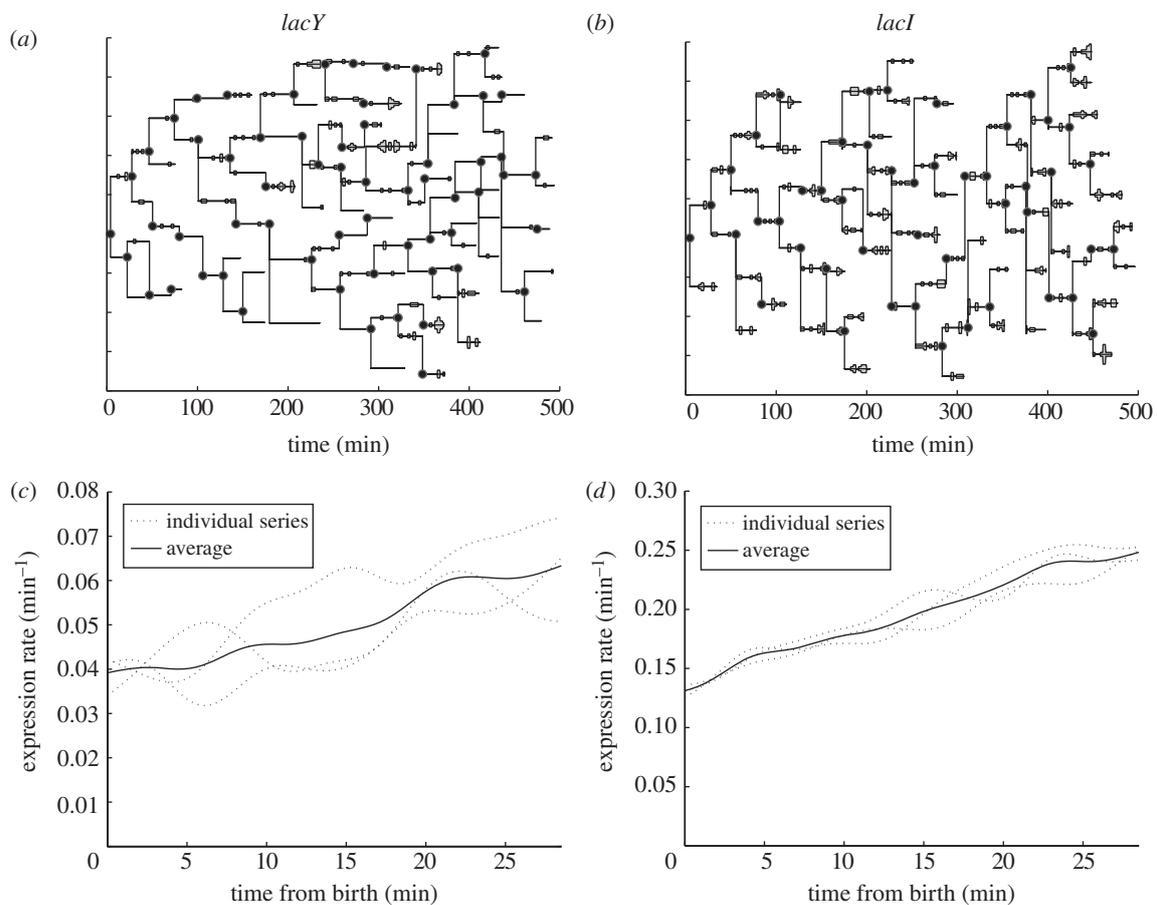
localization that may be expected for Tsr (figure 4a). This is most likely because the protein is inserted at random positions in the membrane and bleaches before reaching the Tsr clusters in the polar regions [4]. For LacI-Venus molecules (figure 4b), we observe a tendency to cluster at positions corresponding to the nucleoids of chromosomal DNA. The number of nucleoids doubles from two, early in the cell history, to four in the later stages, which is consistent with expectations for our growth conditions.

### (d) Synthesis dynamics of an auto-repressed transcription factor throughout the cell cycle

Figure 5 shows lineage trees of cell histories stemming from a single ancestral root of strain SX701 (figure 5a) and JE116 (figure 5b) with bars corresponding to the number of Tsr-Venus and LacI-Venus molecules at the times they were synthesized. The trees are pruned as cells are lost from the segmentation and/or from the trap. For Tsr-Venus expressed from the *lacY* gene, we observe  $1.5 \pm 0.1$  molecules per expression event and  $1.7 \pm 0.1$  events per cell cycle. For LacI-Venus,  $2.2 \pm 0.05$  molecules per expression event and  $2.5 \pm 0.04$  events per cell cycle are observed. The average expression



**Figure 4.** Intracellular localization of molecules over the cell cycle. On the left are the observed molecule densities projected onto the major axis of the cell as a function of time from birth to division for (a) Tsr–Venus and (b) LacI–Venus. The units are the average number of molecules per minute and  $\mu\text{m}$ . The black line at the edge indicates boundary of cell at the apex of the cell poles and expands as the cell grows. On the right is shown the localization of individual (a) Tsr–Venus and (b) LacI–Venus molecules with super-resolution accuracy for early and late stages of the cell cycle.



**Figure 5.** The rate of gene expression over the cell cycle. (a,b) Representative lineage trees for strains (a) SX701 and (b) JE116 stemming from one ancestral root. The absolute number of newly synthesized molecules expressed from the *lacY* and *lacI* gene are indicated as grey bars at the time they are detected. (c,d) The average expression rates from the (c) *lacY* and (d) *lacI* genes over the cell cycle. Solid lines show the average of all three series for each construct. The average of the individual series is shown as dotted lines as an indication of the uncertainty in determining the mean.

rates of Tsr–Venus and of LacI–Venus molecules over the cell cycle are shown in Figure 5*c,d*. Both show relatively large statistical errors, especially Tsr–Venus. The cell histories with generation time 25–32 min and terminal length 4–7  $\mu\text{m}$  are used. For strain JE116, 1418 complete cell histories and 7910 LacI–Venus molecules are observed. For strain SX701, 780 cell histories are retained from the experiment and 1176 Tsr–Venus molecules. The combination of fewer cell histories and lower expression levels leads to larger statistical uncertainty in determining the expression rate of Tsr–Venus from the *lacY* gene. However, our results indicate a greater expression rate of LacI–Venus at the beginning of the cell cycle.

## 4. Discussion

In this study we report on a method combining microfluidics, time-lapsed single-molecule microscopy and automated image analysis capable of monitoring the growth and absolute number of gene expression events throughout approximately 3000 complete individual *E. coli* life-spans per experiment. Further, we demonstrate that it is possible to use a functional transcription factor, LacI–Venus, non-specifically interacting with DNA, to retrieve information on both expression dynamics and super-resolution localization dynamics throughout the cell cycle. We show that the microfluidic chip provides a beneficial and stable environment for exponentially growing *E. coli* cells and a high degree of control and reproducibility. We observe a significant variability in generation times of individual cells. However, we find that generation time is relatively memory-less from generation to generation. More interestingly, cells living in richer media vary more in length at birth than in generation time and that the opposite is true for cells living in poorer media. The underlying causes for this *size–generation time uncertainty relation* and for which range of conditions it holds are presently unclear. LacI–Venus molecules localize onto the nucleoids in the cell. It appears that non-specifically interacting transcription factors are uniformly distributed over the DNA. As expected, we find that LacI–Venus is more highly expressed than Tsr–Venus from the *lacY* gene position. Our result for the latter is consistent with the findings of Yu *et al.* [3] in the number of gene expression events from the *lacZ* gene position during the cell cycle. However, we observe fewer Tsr–Venus

molecules per expression event ( $1.7 \pm 0.1$  instead of  $4.2 \pm 0.5$ ). Given that *lacZ* and *lacY* are transcribed to a polycistronic mRNA, we conclude that the translation rate at the *lacY* position is two to three fold lower than that of the *lacZ* position. The average rate of LacI–Venus expression is slightly higher in the beginning of the cell cycle. We propose that this may be due to partition inequalities at cell division, in which disfavoured cells replenish their transcription factor pools. The experiments confirm the highly variable nature of *in vivo* single-molecule observations (figure 5). We estimate that to obtain a 5 per cent accuracy of the mean expression rate per minute for all points in the cell cycle would require to a total of 4000 and 16 000 complete cell histories of JE116 (LacI–Venus) and SX701 ( $\Delta\text{lacY}::\text{Tsr–Venus}$ ), respectively. Sufficient observations could therefore be obtained with three additional experiments for JE116 and fifteen additional experiments for SX701. The Mather design can potentially sustain a population of bacterial cells in a state of exponential growth indefinitely. Many biological phenomena, such as the development of antibiotic resistance, occur in a small subpopulation of all cells and on longer time-scales than the current longevity of an experiment using our method. Further increasing the throughput and the longevity of the method to enable the study of such phenomena represents a formidable image analysis challenge. However, to our advantage is that the performance of supervised algorithms improves and can be made more advanced as more training data accumulate. We are confident that more advanced algorithms can be implemented to increase both accuracy and speed, which would make it possible to acquire an arbitrary number of cell histories from a single experiment.

The authors thank Jeff Hasty and the members of his laboratory for guidance in establishing the microfluidic protocols. We specifically thank William Mather and Octavio Mondragon-Palomino and colleagues for designing and characterizing the functions of the trap motif used in this study. We also want to thank Christine Jacobs-Wagner and Oleksii Sliusarenko for providing the original version of MICROBETRACKER. The SX701 strain was kindly donated from the laboratory of Sunny Xie. Further, we thank Fredrik Persson for helpful discussions on photonics and dot-detection. We are also grateful for the efforts of Intakhar Ahmad for validation work on the analysis protocols. This study was funded by grants from the European Research Council, the Swedish Research Council, Knut and Alice Wallenbergs stiftelse and Görans Gustafssons stiftelse. A.M. was funded by the Center for Interdisciplinary Mathematics (CIM), Uppsala University. The paper was accepted for publication July 16 2012.

## References

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002 Stochastic gene expression in a single cell. *Science* **297**, 1183–1186. (doi:10.1126/science.1070919)
2. Xie XS, Choi PJ, Li GW, Lee NK, Lia G. 2008 Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* **37**, 417–444. (doi:10.1146/annurev.biophys.37.092607.174640)
3. Yu J, Xiao J, Ren X, Lao K, Xie XS. 2006 Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603. (doi:10.1126/science.1119623)
4. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. 2002 A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90. (doi:10.1038/nbt0102-87)
5. Elf J, Li GW, Xie XS. 2007 Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194. (doi:10.1126/science.1141967)
6. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538. (doi:10.1126/science.1188308)
7. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, Jun S. 2010 Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103. (doi:10.1016/j.cub.2010.04.045)
8. Danino T, Mondragon-Palomino O, Tsimring L, Hasty J. 2010 A synchronized quorum of genetic docks. *Nature* **463**, 326–330. (doi:10.1038/nature08753)
9. Young JW, Locke JCW, Altinok A, Rosenfeld N, Bacarian T, Swain PS, Mjolsness E, Elowitz MB. 2012 Measuring single-cell gene expression

- dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88. (doi:10.1038/nprot.2011.432)
10. Sliusarenko O, Heinritz J, Emonet T, Jacobs-Wagner C. 2011 High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80**, 612–627. (doi:10.1111/j.1365-2958.2011.07579.x)
  11. Mather W, Mondragon-Palomino O, Danino T, Hasty J, Tsimring LS. 2010 Streaming instability in growing cell populations. *Phys. Rev. Lett.* **104**, 208101. (doi:10.1103/PhysRevLett.104.208101)
  12. Datsenko KA, Wanner BL. 2000 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645. (doi:10.1073/pnas.120163297)
  13. Choi PJ, Cai L, Frieda K, Xie XS. 2008 A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446. (doi:10.1126/science.1161427).
  14. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. 2012 The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598. (doi:10.1126/science.1221648)
  15. Kass M, Witkin A, Terzopoulos D. 1988 Snakes: active contour models. *Int. J. Comput. Vis.* **1**, 321–331. (doi:10.1007/bf00133570)
  16. Bishop CM. 2006 *Pattern recognition and machine learning*. Berlin, Germany: Springer.
  17. Ronneberger O, Raffel M, Kompenhans J. 1998 Advanced evaluation algorithms for standard and dual plane particle image velocimetry. *Proc. 9th Int. Symp. Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal, 13–16 July 1998*, paper 10.1.
  18. Moré J. 1978 The Levenberg–Marquardt algorithm: implementation and theory, chapter 10. In *Numerical Analysis* (ed. GA Watson), Lecture Notes in Mathematics, pp. 105–116. Berlin, Germany: Springer.