

A Microfluidic Setup for Quantifying Single-Cell Transcription Regulatory Dynamics

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Matthias Kaiser
aus Biberist, SO

Luzern, 2018

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Erik van Nimwegen

Prof. Dr. Dirk Bumann

Basel, den 22.3.2016

Prof. Dr. Jörg Schibler
Dekan

Contents

Abstract	2
1 Introduction	3
1.1 Gene Regulation in Bacteria	3
1.2 Measurements of Gene Regulation on the Single-Cell Level in Bacteria	5
1.3 Regulation of the <i>E. coli lac</i> Operon	7
2 The Dual Input Mother Machine: From the Design to the Experiment	13
2.1 Photolithography	13
2.2 The Dual Input Mother Machine: Implementing Environmental Control	17
2.3 Mask Design	19
2.4 Challenges in the UV-Lithography	22
2.5 The Procedure of an Experiment	24
2.5.1 Preparation of the Device	24
2.5.2 Loading of the Passivation Buffer	28
2.5.3 Cell Loading	29
2.5.4 Imaging	32
2.5.5 Stopping the Experiment	32
3 Main Publication	33

4 Summary and Future Prospects	51
References	57
Acknowledgements	63
Curriculum Vitae	65

Abstract

Bacteria are exposed to fluctuations in their environment and can respond to such changes by regulating gene expression, often at the level of transcription. Since gene expression is an inherently stochastic process, identical cells within a single environment display heterogeneous expression levels. To understand how the stochastic processes in gene expression affect the dynamics of single-cell gene regulation it is necessary to observe gene expression in single cells in changing environments. Recently developed microfluidic devices combined with quantitative fluorescence time-lapse microscopy allow lineages of single cells to be followed over long time-scales and to measure their growth and gene expression phenotypes simultaneously. However these devices are missing the environmental control needed to study gene regulation. Therefore we set out to find a way to combine the long-term observation of single cells with precise environmental control in a single microfluidic chip. As a basis we chose a device called the Mother Machine in which single files of cells are growing in small dead end growth channels. These growth channels are connected to a main channel with a constant flow of medium for nutrient diffusion into the growth channels. The cells at the dead end of the growth channels are trapped and when dividing push their progeny into the main channel where they are removed by the flow. Therefore

the trapped cell can be monitored essentially for its whole lifetime, while its progeny can only be observed for a short timeframe before they leave the growth channel. By combining the Mother Machine design with a specialized dual input junction and mixing serpentine for environmental control we developed a device that offers new prospects in studying gene regulation. Together with the device we developed an easy to use software solution to analyze data from Mother Machine like devices together with our collaboration partners. This integrated experimental and computational setup will be an important tool to understand the genetic basis for differences in single-cell expression distributions, and to understand how natural selection has shaped single-cell gene regulation. As a first example we show how single cells differ in the regulation of the expression of the *lac* operon when exposed to alternating changes in the available carbon source switching between glucose and lactose every 4h.

Chapter 1

Introduction

1.1 Gene Regulation in Bacteria

Bacteria live in diverse environments across the whole planet. In many of those habitats bacteria are confronted with fast and drastic changes in conditions to which they are directly exposed. Such changes might include fluctuations in available nutrients, in temperature or in various other stresses. To survive these fluctuations and increase fitness in the new condition it is crucial to sense changes and to adapt. One way to do so is to change the transcriptional program in response to external stimuli. The most well known example of such an adaptation is the change in expression of the genes involved in lactose metabolism in *Escherichia coli* depending on the availability of glucose and lactose [1].

Regulation of gene expression often happens at its first step where the RNA polymerase (RNAP) binds to the promoter sequence to initiate transcription [2–4]. The bacterial RNAP “core enzyme” is a protein complex consisting of five subunits $\beta\beta'\alpha_2\omega$. The “core enzyme” alone can only bind DNA and synthesise RNA nonspecifically [4, 5]. To bind to promoter DNA the RNAP “core enzyme” has to form the “holoenzyme” together with a sigma factor that specifically binds

promoter DNA to initiate transcription [6–9]. Prokaryotic sigma factors can be classified into two families, the σ^{70} and the σ^{54} family [6, 10, 11]. In *E. coli* there are seven known sigma factors that recognize different promoters and globally regulate gene expression in response to environmental conditions (for a review check [12]). Further regulation of transcription initiation is mediated via transcription factors that act as regulatory switches. Such transcription factors do sense specific signals and in response affect transcription initiation by directly interacting with the target DNA. The regulation can be both positive [13] and negative [1]. Negative regulators normally bind directly to the promoter region to interfere with RNAP recruiting while positive regulators usually bind upstream of the promoter region to facilitate RNAP recruitment (for a review see [14]).

Even though bacteria tightly regulate their gene expression in response to external stimuli as mentioned above, genetically identical cells in a homogeneous environment still show some heterogeneity in gene expression because it is an inherently stochastic process [15–18]. Both fluctuations in the number and activity of cellular components involved in gene expression and noise coming from the stochastic nature of biochemical reactions in gene expression contribute to cell-to-cell variation [16]. The phenotypical heterogeneity resulting from noise in gene expression might be beneficial in changing environments on the population level because some fraction of the population might by chance be in a state that is beneficial in the new environment and thus will survive even fast and drastic changes [19]. For example it was observed that fluctuations in the expression of a toxin called HipA can be linked to the formation of persister cells that are tolerant

to antibiotics [20]. Therefore fluctuations in gene expression seem to have functional consequences (for a review see [21]).

1.2 Measurements of Gene Regulation on the Single-Cell Level in Bacteria

To get a better understanding of how stochastic processes affect gene regulation and what consequences differential regulation has for the fitness of a cell in changing environments, it is crucial to measure the growth phenotype and the dynamics of gene expression in single cells exposed to environmental fluctuations.

There are several experimental methods that allow measurements of growth and gene expression of single cells using quantitative fluorescence time-lapse microscopy. One widespread method is the use of agarose patches on which the cells can grow and form microcolonies [22]. Combined with microfluidics this method and similar approaches even allow for changing the environment [23–25], but have one big drawback in common: The size of a growing microcolony quickly exceeds the size of the field of view in the microscope, additionally cells start to form multilayered structures in big microcolonies which leads to the loss of the single cell resolution.

Here, novel microfluidic approaches open new perspectives. By flushing away the progeny of observed cells the possible observation times are prolonged dramatically [26, 27]. While these approaches make longer measurements on the single-cell level possible, they are unfortunately not equipped for precise environmental control yet. Therefore to measure transcriptional responses to changing environments in sin-

gle cells the field is missing a device that brings together the advantages of a system for long-term experiments with a system to carefully control the growth environment. One especially interesting microfluidic device that got a lot of attention in recent years is the so-called Mother Machine. The Mother Machine is a device published in 2010 designed to study long-term growth in *Escherichia coli* [26]. The main features of the Mother Machine are small dead end growth channels in which the bacteria are trapped. Nutrients and waste products are diffusing in or out, respectively, through the end of the growth channel that is connected to a main channel with a constant flow of medium. These growth channels are $\sim 20\mu\text{m}$ long and have a cross section of $\sim 1\mu\text{m} \times \sim 1\mu\text{m}$. Therefore they accommodate exactly one column of bacteria with a cell trapped at the closed bottom end. This mother cell will divide and push its progeny up until they leave the growth channel and are removed by the flow in the main channel. While the progeny of the mother cell can only be observed for a certain time before they leave the growth channel, the mother cell can be monitored for its whole life time. The big advantage of this system is that it allows following single cells over a long time-scale in separated growth channels, therefore no neighbouring effects can affect the results. For these reasons the Mother Machine is a good basis to design a device that allows for both single-cell observations over a long time-scale and environmental control. First applications of medium switches in a Mother Machine have been shown [28]. In the respective study short switches to a medium containing IPTG were used to temporally induce a construct in the cells growing in the Mother Machine. To do so a Y-junction connector feeding into the device was used

to bring the tubings of two different syringes together. Manual switches were generated by activating one or the other syringe, while the inactive line was shut with clamps. To achieve more precise switches and expand the possibilities of controlling the environment, a Dual Input Mother Machine device was developed in this thesis to switch between or mix two different inputs in the device under automatic control of programmable syringe pumps.

1.3 Regulation of the *E. coli lac* Operon

This section will introduce the regulatory system controlling the expression of the *lac* operon in *E. coli*, which is probably the best known gene regulatory system in biology and was studied to demonstrate the capabilities of the newly designed microfluidic device presented here.

In 1960 Jacob *et al* defined the term operon for genes organized in transcriptional units that are controlled by an operator [29]. In the year after Jacob and Monod presented their model for gene regulation in bacteria [1, 30] with the regulation of the expression of the *lac* operon as a prominent example. The *lac* operon consists of the three genes *lacZ*, *lacY* and *lacA* that are needed to metabolize lactose. Adjacent to the *lac* operon is the gene for the *lac* repressor *lacI* (Figure 1.1).



Figure 1.1: Organization of the *lac* operon in *E. coli*

The *lacZ* gene codes for the β -galactosidase. The LacZ β -galactosidase is a glycoside hydrolase enzyme that fulfills three different functions in an *E. coli* cell. First LacZ cleaves lactose into glucose and galactose [31], second it produces allolactose by catalyzing the transgalactosidation of lactose [32, 33] and third it can also cleave allolactose into glucose and galactose [34]. The second gene *lacY* codes for the lactose permease which is a lactose/proton symporter responsible for lactose uptake into the cell [35–37]. The last gene in the operon *lacA* codes for the galactoside O-acetyltransferase. The function of this enzyme is not fully understood yet but there is a hypothesized role of LacA as a detoxifying enzyme [38]. The transcription of these genes is negatively regulated by the LacI repressor. There are three different operator sites to which the repressor LacI can bind when no lactose is present (Figure 1.1). The main operator O1 is located 11 base pairs (bp) downstream of the transcription start site. When LacI binds to O1 it interferes with RNAP binding and blocks transcription initiation [39]. The two auxiliary operators O2 (412 bp downstream of the transcription start site) and O3 (82 bp upstream of the transcription start site) cooperate with O1 in repression of the expression of the *lac* operon. This cooperative interaction between the operators is due to DNA loop formation with tetrameric LacI [40, 41]. When lactose is present it is partially converted into allolactose by the LacZ enzyme. Allolactose is the inducer of the *lac* operon and is bound by the tetrameric repressor LacI [32, 33]. Upon binding the inducer the repressor undergoes an allosteric transition with the resulting state having a lower affinity for the operators [42, 43]. Therefore the repressor acts as a switch that only allows transcription

from the *lac* operon when lactose is available as a carbon source. Once the *lac* genes are expressed the number of permease molecules in the membrane rises, lactose is imported more easily and is converted to allolactose by the increasing amounts of LacZ, which leads to even stronger induction of the expression of the *lac* operon. But the expression of the *lac* operon is not only regulated by the availability of lactose. Besides the negative regulation via LacI there is also positive regulation of the expression of the *lac* operon exerted by the catabolite activator protein (CAP). Long before Jacob and Monod published their model for gene regulation in bacteria, Monod observed that *E. coli* does not metabolize glucose and lactose simultaneously when both carbon sources are present but shows two phases of growth, first metabolizing glucose, which is more easily metabolized, and only when the glucose is gone changes the metabolism to lactose usage. This phenomenon was called diauxic shift and it was shown that glucose effectively prevents expression from the *lac* operon [44–46]. When Makman and Sutherland 1965 showed that that glucose-starved *E. coli* cells accumulate high amounts of cAMP under certain conditions, the question arose as to whether cAMP has something to do with the glucose effect or catabolite repression that prevents expression of the *lac* genes when glucose is present [47]. In several subsequent studies CAP was discovered and was shown to activate transcription from the *lac* operon when bound to cAMP. In addition it was shown that strains lacking the adenylate cyclase and CAP fail to induce expression of the *lac* genes normally [48–56]. Taking these findings together, it was thought that with CAP and cAMP the main players in the glucose effect were found. Later it was shown that the cAMP levels are essen-

tially the same in cells growing in glucose and cells growing in lactose and it was suggested that the main reason for the inhibitory effect of glucose on the expression of the *lac* genes is due to inducer exclusion [57].

Glucose is imported into the cell through the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS). The PTS is a phosphorelay system found in many bacterial species and usually consists of 5 proteins (some of these proteins can be fused together in some cases): enzyme I (EI), histidine protein (HPr) and enzyme IIA, IIB and IIC. When glucose is available *E. coli* imports it via the PTS, the sugar is phosphorylated during import and the glucose specific EIIA^{Glu} component is mainly present in its non-phosphorylated state (for a review about the PTS see [58]). In its non-phosphorylated state EIIA^{Glu} inhibits the lactose permease LacY and therefore expression of the *lac* genes is prevented in the presence of glucose through inducer exclusion (Figure 1.2). Only when glucose is gone lactose can be imported and converted to the inducer allolactose. Interestingly non-phosphorylated EIIA can only interact with LacY when the permease is in the conformation it takes when lactose or another substrate is present [59, 60]. Since the same inactivation mechanism is also used for other sugar permeases this ensures that EIIA is only directed to the permeases that really need inactivation.

Despite all these regulatory mechanisms stochasticity is also observed in the expression of the *lac* operon [16, 18]. Therefore the regulation of the expression of the *lac* operon is an ideal system to study how stochastic processes in gene expression can affect the dynamics of gene regulation in single cells. There are some recent studies that monitored the

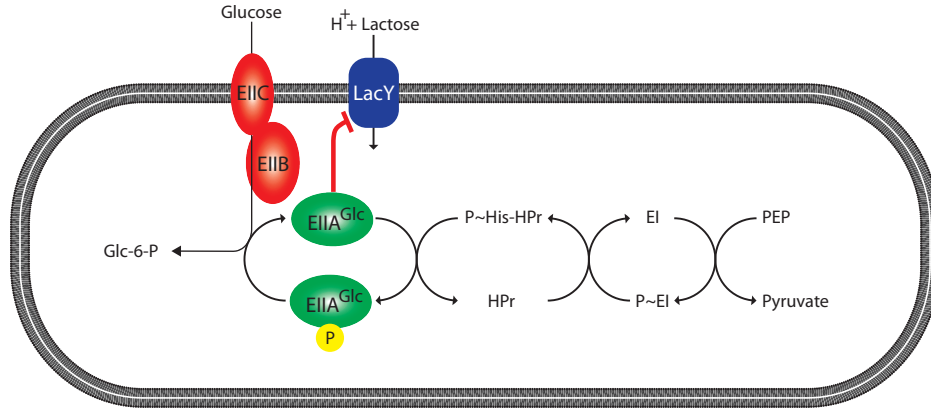


Figure 1.2: Glucose import via PTS and the inhibitory effect of non-phosphorylated EIIA^{Glu} on lactose import (adapted from [58])

regulation of the expression of the *lac* operon in changing environments on the single cell level [24, 25]. But the experimental setups used do not allow performing experiments over long time-scales and therefore limit the duration and frequency of the medium switches. In contrast, Lambert and Kussel 2014 [61] monitored growth rates and expression of the *lac* operon during several alternating switches of the carbon source between glucose and lactose but without single-cell resolution. By providing both, single-cell data and the possibility to perform experiments over long time-scales, the Dual Input Mother Machine can bring the advantages of the experimental systems used in these studies together. The kind of data that can be collected using the Dual Input Mother Machine will be presented in Chapter 3.

Chapter 2

The Dual Input Mother Machine: From the Design to the Experiment

2.1 Photolithography

The Dual Input Mother Machine device presented in this work is a microfluidic device made out of polydimethylsiloxane (PDMS). PDMS is a silicon-based organic polymer and is widely used in microfluidics because it is chemically inert, non-toxic and optically transparent. To get a microfluidic chip, liquid PDMS is poured onto a master that serves as a template. After curing the rubber-like PDMS block can be removed from the master and can be bonded to glass. To do so, both materials are chemically activated by plasma treatment to get an irreversible covalent bond between the PDMS and the glass slide [62] (Figure 2.1).

To go through the different steps in the development of such a microfluidic device, first the method for producing the master shall be introduced here. The master is produced using a technique called photolithography. With this technique, structures made out of light sensitive materials called

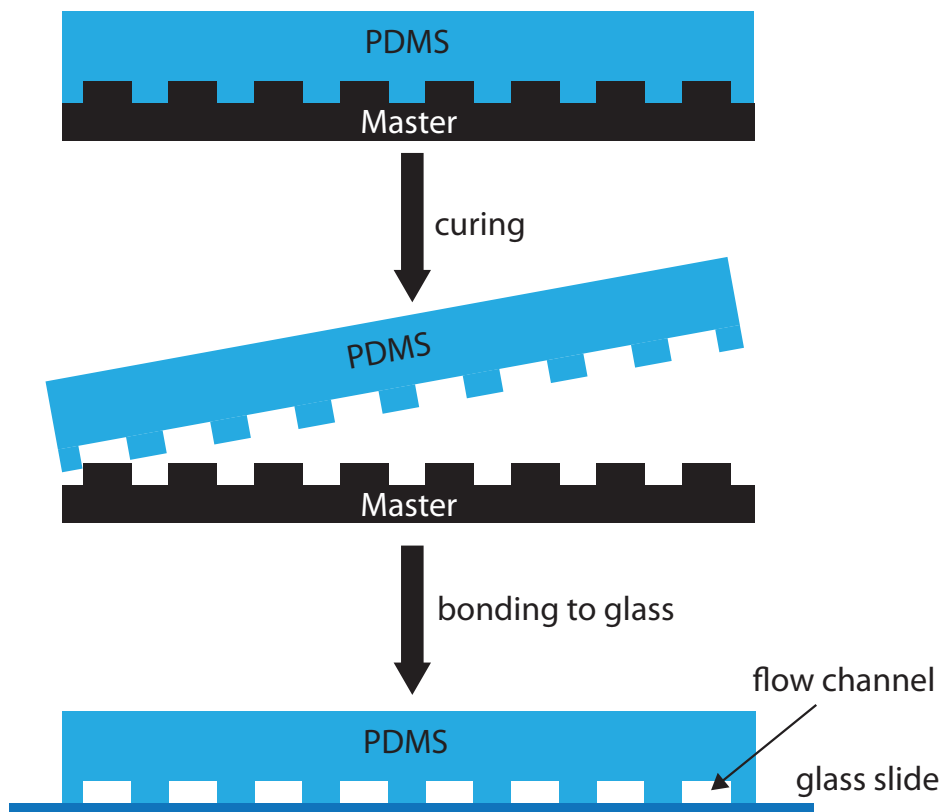


Figure 2.1: The process of building a PDMS microfluidic device from an existing master: 1) Liquid PDMS is poured onto the master, 2) After curing the PDMS block with the imprinted structures can be removed from the master, 3) The PDMS block is bonded to glass after chemical surface activation, channels are formed between the PDMS and the glass slide

photoresists are built up layer by layer on a flat substrate (often silicon wafers). Photoresists can be classified into two groups: 1) Positive photoresists that become soluble to the developer when exposed to light (unexposed areas remain insoluble) and 2) negative photoresists that only become insoluble when they are exposed to light [63]. The masters for the Dual Input Mother Machine were built using different SU-8 negative photoresists (micro resist technology GmbH) that are exposed with a wavelength of 365nm. Therefore the process of photolithography will be explained using negative photoresist here. If a microfluidic device consists of structures with different heights, there is one photoresist layer for each height and each layer is exposed separately (Figure 2.2). The photoresist layers are spread by a process called spin coating. A certain amount of liquid photoresist is placed on the wafer which is spun afterwards to get an even layer of photoresist. The height of the layer is controlled by the spin speed, polymer concentration and the intrinsic viscosity of the resist. Next, the photoresist layer can be exposed with UV-light, a photomask is used to only expose the desired areas of the resist layer. There are two different exposure methods: 1) contact printing where the mask and the substrate with photoresist layer are in contact and 2) proximity printing where there is a small gap between the mask and the substrate with the photoresist layer. Contact printing allows high resolution structures to be built, but because of the contact there is the risk of small debris damaging the mask and causing defects in the structures. This risk is minimized with proximity printing at the cost of a lower resolution because of diffraction. There is a third more complex exposure method called projected printing where the image of the patterns on

the mask is projected to the photoresist layer using optics [63]. This method will not be discussed here in detail since it was not available in the facility used for the structuring attempts. In a developing and cleaning step unexposed resist is removed and more layers can be added to the structure (Figure 2.2)[63, 64]. For the Dual Input Mother machine a first layer with a height of $0.9\mu\text{m}$ was used for the fine growth channels and a second layer with a height of $5\mu\text{m}$ for the flow channels.

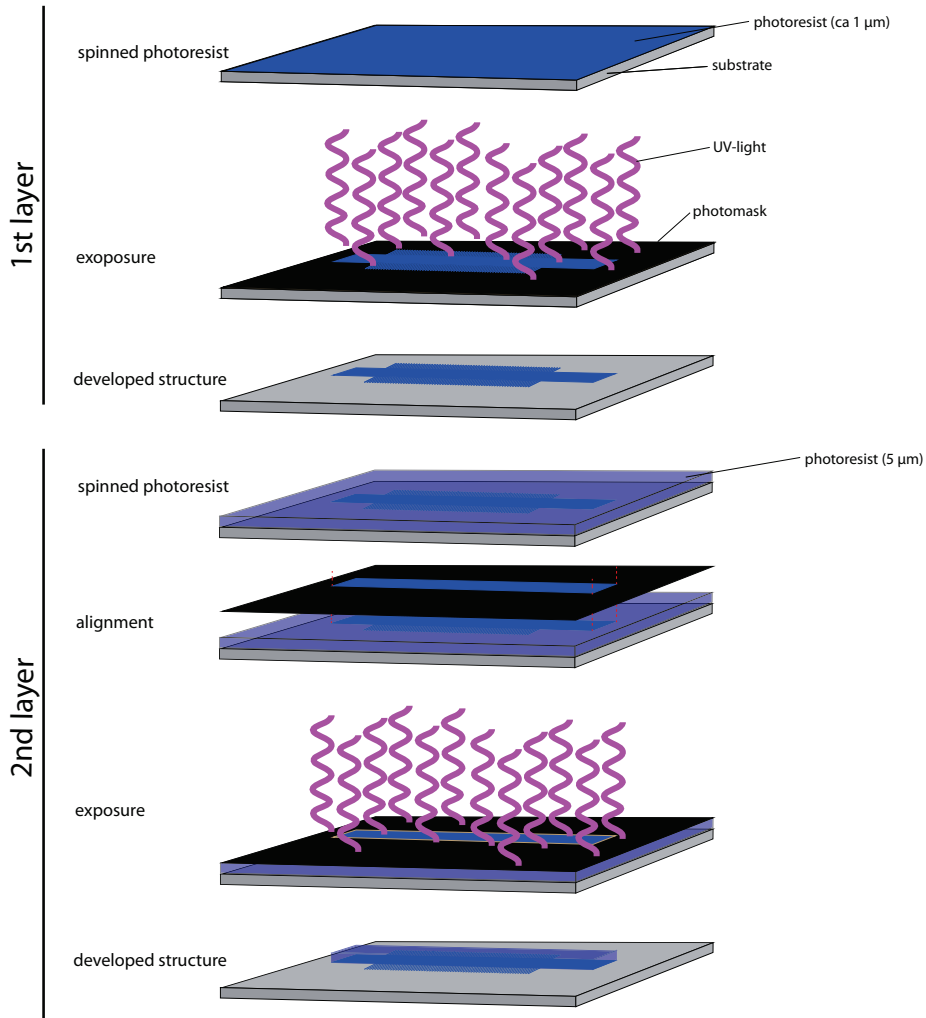


Figure 2.2: Photolithography: Illustration of the structuring of a two layer Mother Machine master

2.2 The Dual Input Mother Machine: Implementing Environmental Control

The Mother Machine [26] is a microfluidic device to follow the growth and gene expression of single bacteria over long periods of time. However to use the advantages of this design to study gene regulation on the single cell level it needs to be coupled to a precise environmental control.

The device presented here does not only allow to switch between two different inputs but also to create different mixtures of two inputs. This requires a special junction to bring the two inputs together upstream of the growth channels and to allow for both switching and mixing. When inputs are switched the junction needs to be designed to prevent backflow from the active into the inactive input. For mixing on the other hand the design needs to allow for precise changes of the mixing ratio and further a solution has to be found to guarantee for sufficient mixing of the two inputs before the mix reaches the growth channels. We decided to use a so called Dial-A-Wave junction (DAW junction) [64] combined with mixing serpentine. This setup fulfills all our criteria and still results in a relatively simple device. The DAW junction brings the flow in the two inlets together and divides it into three outlets. While the middle outlet is connected to the main channel with the growth channels, the two outer outlets are connected to waste channels (Figure 2.3A, left). This design allows to have flow from one input only in the main channel, while the other input can still be preassured sufficiently to prevent backflow. This is possible since the excess flow from the inactive input can be redirected to the waste channels (Figure 2.3B, middle and right). By

changing the flow rates of the two inputs the junction can also be used to create different mixing ratios between the two inputs (Figure 2.3B, left).

For the subsequent mixing of the inputs it is important whether a flow is laminar or turbulent and this depends on the ratio between the inertial and the viscous forces of a fluid's flow. This ratio can be expressed in the Reynolds number Re :

$$Re = \frac{\rho \nu D_h}{\mu} \quad (2.1)$$

Here ρ is the density of the fluid, ν is the mean fluid velocity, D_h is the hydraulic diameter of the channel and μ is the viscosity of the fluid. Flows in microfluidics have typically very low Reynolds numbers and are thus strictly laminar [65]. Therefore there is only diffusive mixing happening in these devices. To make sure the inputs joining at the DAW junction flow together long enough to ensure full mixing before the mixture reaches the cells we introduced a series of mixing serpentes between the DAW junction and the growth channels (Figure 2.3A, right). The length required for effective diffusive mixing depends on the flow speed (fluid velocity) ν , the width of the micro-channels L and the diffusion coefficient of the molecule of interest D [66]:

$$\Delta y_m \approx \frac{\nu L^2}{D} \quad (2.2)$$

Such simple mixing serpentes work well for the device presented here but for some applications simple 2D mixing serpentes would need to be impractically long to ensure effective diffusive mixing. In this case more complicated mixing devices have to be used to save space on the microfluidic chip.

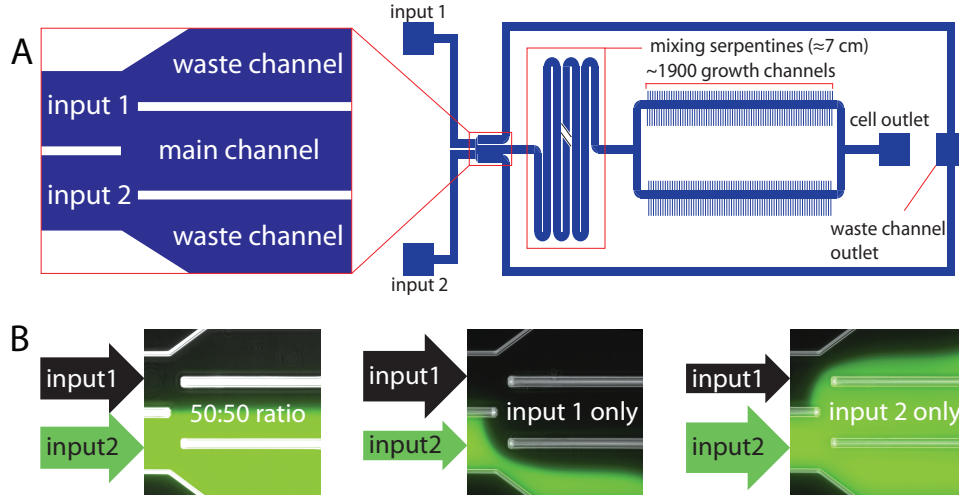


Figure 2.3: The Dual Input Mother Machine: A) Scheme of the Dual Input Mother Machine with a blow-up of the DAW junction, B) DAW junction in different states, left: 50:50 mixture, middle: input 1 only (input 2 is directed to the lower waste outlet), right: input 2 only, input 1 is directed to the upper outlet

2.3 Mask Design

In this section the design of the photomask for the Dual Input Mother Machine using computer-aided design (CAD) software will be explained and some changes compared to the original Mother Machine other than the introduction of environmental control shall be highlighted. Once the design is complete a specialized company can use the plans to fabricate a photomask for the photolithography step. The design of the mask was done using the widespread AUTOCAD[®] software from AUTODESK[®] (free academic licences are available). To discriminate areas that shall remain clear on the photomask from the ones that shall be opaque, the features are drawn in two different layers. In the first layer every object enclosed by lines will end up transparent on the mask, while objects enclosed by lines in the second layer will be opaque (Figure 2.4A). As mentioned above the Dual Input Mother Machine

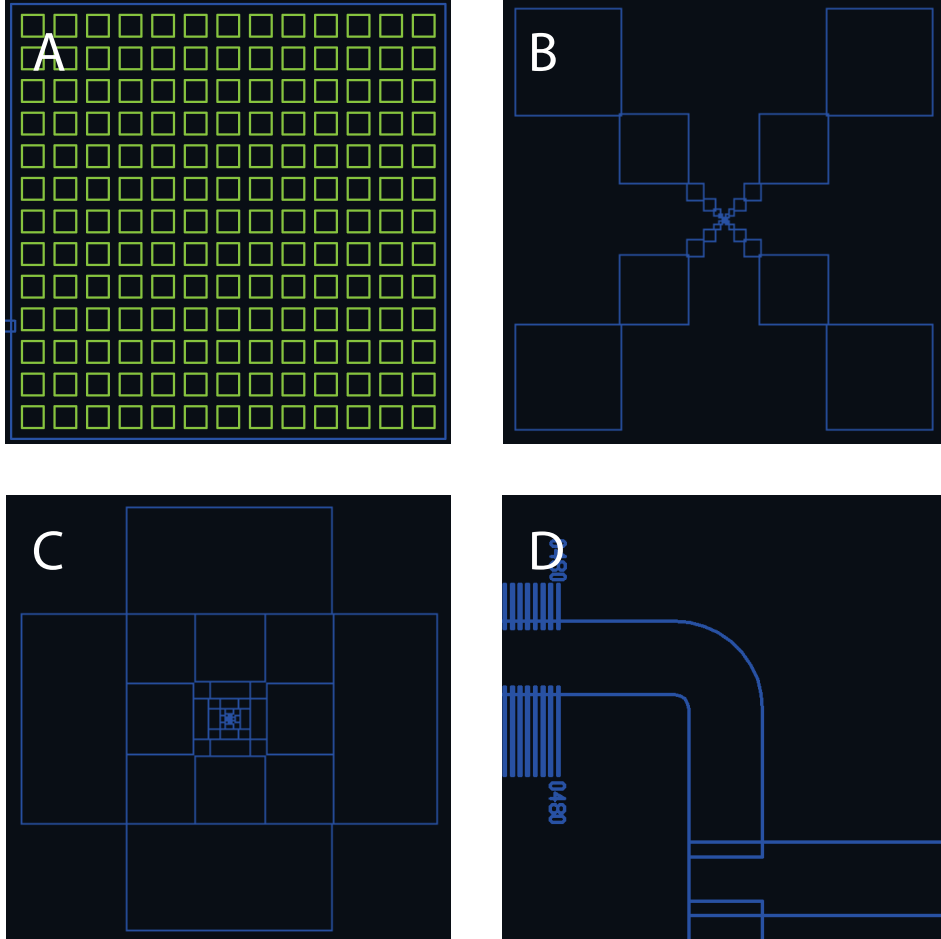


Figure 2.4: Design of the Dual Input Mother Machine: A) The design features quadratic in- and outlets with a length of 2mm. In the CAD design the structures that will remain clear on the photomask are outlined in blue (layer 1) while structures that will be opaque on the mask are outlined in green (layer 2), B) and C) To align the structures for the exposure of the second layer of photoresist the structures are combined with alignment marks on the mask, D) Rounded edges assure good flow in the bends without any regions with no flow in the edges

master is built up by two photoresist layers with different heights. Therefore two different designs have to be printed on the mask to expose the first and second layer. The first layer consists of the growth channels and the lowest part of the flow channels (Figure 2.6A top) while the second layer consists of the rest of the flow channel (Figure 2.6A bottom). It is important to perfectly align the structures of the first and the second layer in the process of exposure. To make this task easier alignment marks were put next to the structures for the two layers on the photomask (Figure 2.4B and C).

Besides the introduction of environmental control we also made some more changes to the original Mother Machine design in order to optimize the system for our needs. The high medium flow used with the original design is impractical for long experiments and we therefore decided to reduce the dimensions of the flow channels. In the Dual Input Mother Machine all flow channels were designed to be $50\text{ }\mu\text{m}$ wide only (compared to $100\text{ }\mu\text{m}$ in the original design) and in the photolithography step the height of flow channels was reduced from $25\text{ }\mu\text{m}$ to $6\text{ }\mu\text{m}$. Also the design of the in- and outlets was changed. The Dual Input Mother Machine features two inlets and two outlets. Since handling of a microfluidic device gets more complicated with a higher number of in- and outlets, we decided to design the in- and outlets rather large to make working with them easier. We designed square in- and outlets with a length of 2mm. To avoid collapsing of these structures, small pillars are placed inside them (Figure 2.4A). Last, to avoid flow free areas in edges all bends downstream of the DAW junction were designed to have rounded edges as shown in figure 2.4D. From

these plans a 5" photomask (quarz with a chrome layer) was ordered from Compugraphics Jena GmbH.

2.4 Challenges in the UV-Lithography

The masters for the Dual Input Mother Machine device presented here were fabricated using SU-8 negative photoresists exposed with UV-light (365nm) (micro resist technology GmbH). While the second layer with the flow channels turned out to be unproblematic to produce, the exposure of the first layer with its fine growth channel structures and the perfect alignment of the two layers turned out to be very challenging. Spin coating the resist for the first and second layer was unproblematic. Because of the small dimensions of the growth channels ($0.9\mu\text{m}$ high, $0.8\mu\text{m}$ wide and $25\mu\text{m}$ long) on the first layer we worked in contact mode for the exposure. Nevertheless with the conditions in the local cleanroom we failed to get reproducible results for the small structures. It turned out that it is very hard to avoid any dust or debris ending up on the photoresist layer in this facility. These objects than lead to bad contact during the exposure which in turn leads to a loss of resolution and uneven results across the master (Figure 2.5A). Because of these problems the structuring attempts resulted in either too large or poorly developed structures (Figure 2.5B). Since the first layer of the Dual Input Mother Machine includes both the growth channels and part of the flow channel (Figure 2.6A top) it is crucial to perfectly align this layer with the flow channel parts in the second layer. If the alignment fails, the growth channels on one side of the flow channel can end in a $0.9\mu\text{m}$ high flow channel because the higher part of the flow

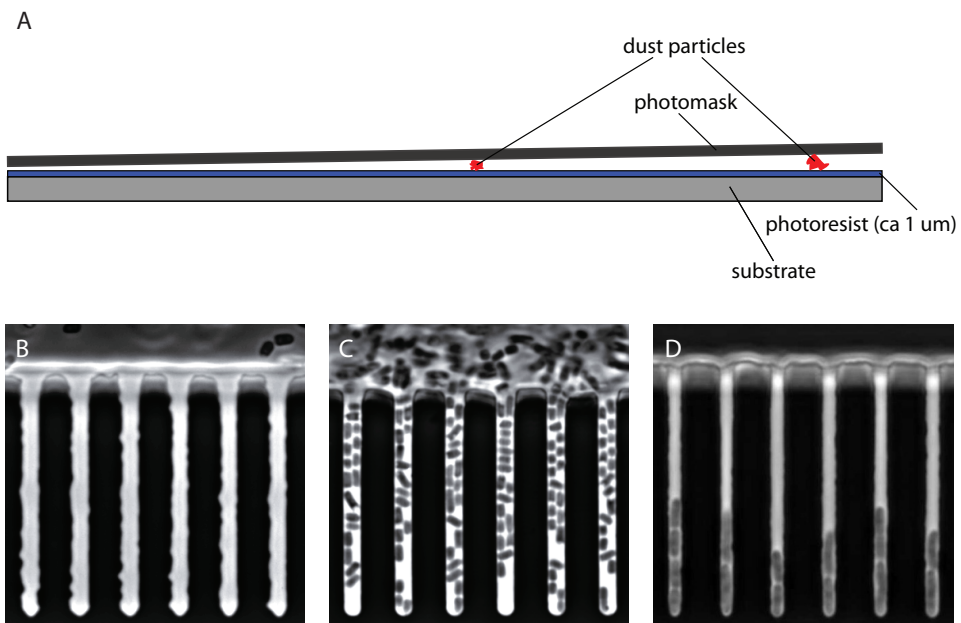


Figure 2.5: Problems with structuring growth channels probably because of bad contact between mask and photoresist: A) Illustration of how dust particles prevent proper contact between the photoresist layer and the photomask B) Structures that are not developed properly and show uneven contours, C) Structures that are too big (several times wider than the structures on the mask), D) improvement over the first two cases but still not narrow and smooth enough

channel is shifted (Figure 2.6B). In an experiment such a device causes problems because cells don't load very well on the side where the second layer of the flow channel is shifted away from the growth channels. In addition, cells that would grow in these growth channels would not be flushed away properly when leaving the growth channel. While the problem with the alignment would have been possible to solve within the environment we worked in (probably by changing the mask design) we came to the conclusion that the setup we were using in the local clean room was insufficient to get well developed structures with the right size. Therefore we asked a company (micro resist technology GmbH) with an advanced clean room environment to do the structuring of the masters for us. Even in this, more professional, setup reproducibility seems to be limited and we got slightly varying structure sizes on different masters.

2.5 The Procedure of an Experiment

2.5.1 Preparation of the Device

Before the experiment the device has to be prepared. It is recommended to prepare several devices in advance to have backups in case some step is going wrong at the day of the experiment. All steps should be carried out wearing gloves to protect the device.

- 1) First the master for the device has to be cleaned with pressurized air. Other cleaning methods (solvents, tape or tissue) should be avoided because they might destroy the delicate features.

- 2) In the next step the PDMS needs to be prepared. In this work the Sylgard[®] 184 Silicone Elastomer Kit was used

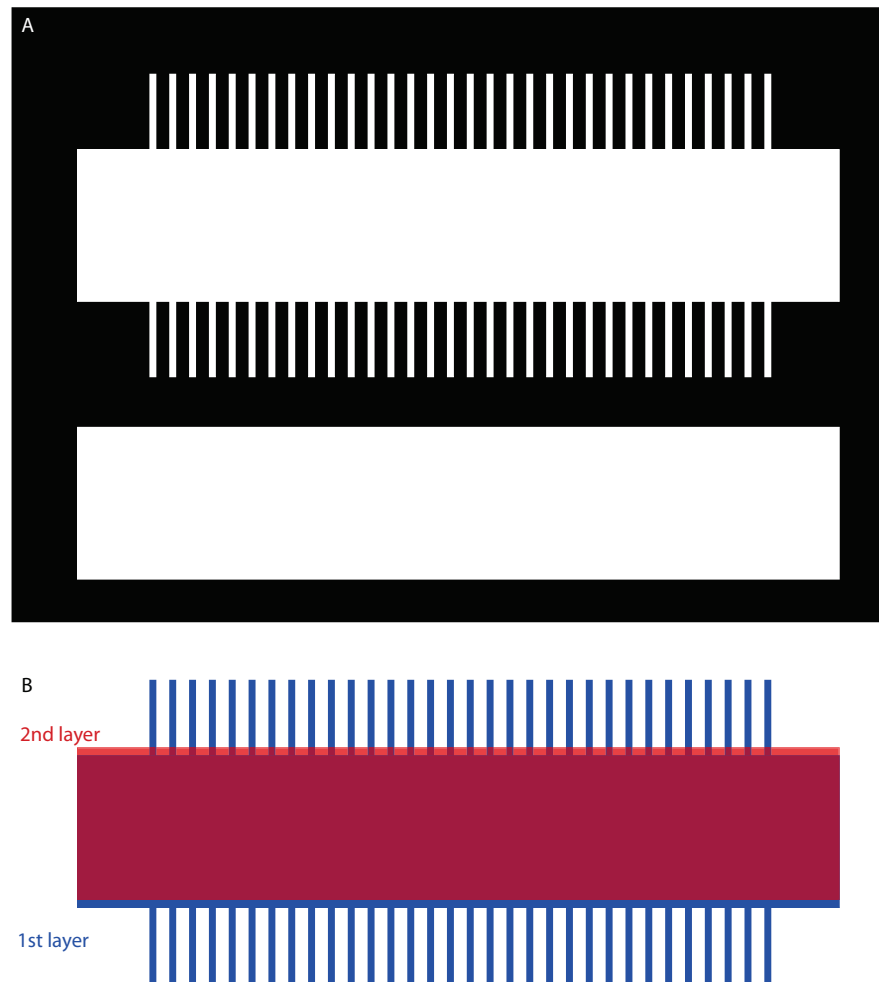


Figure 2.6: Problems with the Alignment of the two Photoresist Layers: A) Schematic view of the photomask for the first layer (lower part of the flow channel plus growth channels) in the top and for the second layer (rest of the flow channel) in the bottom, B) Topview of a misaligned structure

with a 1:10 curing agent to base ratio. The correct amounts of the two components can be measured using a balance. It is crucial to thoroughly mix the two components to assure complete curing afterwards.

3) Before pouring the PDMS it should be checked again that there is no dust left on the surface of the master. It is recommended to pour to a thickness of about 7mm. Since it is impossible to mix and pour the PDMS without creating bubbles, trapped air has to be removed before curing. One can either use a vacuum chamber to accelerate the process or leave the master on room temperature for around 2h to let the bubbles come to the surface. Remaining bubbles can be removed with a pipette tip.

4) To accelerate the curing, the master is incubated at 80°C for at least 4h or overnight (the curing time can depend on the thickness of the device but for the thickness mentioned above these times should be sufficient).

5) Once the device is fully cured it can be cut out. One needs to make sure not to use too much pressure to avoid breaking the silicone wafer and it is recommended to cut far away from the structures. If multiple structures are on one master it is easier to separate them after removing the mold from the master.

6) To later connect the tubing to the device, in- and outlets need to be punched. In the device presented here the in- and outlets are fairly large and should be easy to see. If not it might help to mark their location with a pen on the PDMS. To punch the hole a Harris Uni-Core™ 0.75mm biopsy punch was used. It is easiest to punch the holes from the side with the structures on your mold. Before proceeding with the

bonding step it should be double checked that all in- and outlets are punched properly.

7) The device is bonded to a coverslip with a thickness of 0.17mm (#1.5). To get good bonding both the PDMS block with the device and the coverslip need to be cleaned thoroughly. First the PDMS block is cleaned with 100% isopropanol. Sometimes rinsing is not enough and dirt needs to be cleaned away manually. In a last step the PDMS block is rinsed again with isopropanol to flush away remaining dirt and can then be dried using pressurized air. It is important to dry the block as quickly as possible to prevent any smears from slowly drying isopropanol. The coverslip needs to be cleaned in the same way and once both the PDMS block and the coverslip are free of dust and smears they can be placed in the plasma cleaner. In this work we used a plasma cleaner from Harrick Plasma (PDC-32G). The chamber of the plasma cleaner is evacuated and the pressure is set to be around 2mbar. Once this pressure is reached, the switch for the RF level is switched to "high" for around 45s. The purple glow of the plasma can be observed through the ventilation slit to check for proper plasma formation. After switching off the RF level switch, the chamber can be slowly filled with air again until the pressure has equilibrated with the pressure in the room. Now the coverslip can be put on a black background (helps to check the bonding) and the PDMS block can be placed on it (make sure the side with the structures faces down). When good bonding is achieved one should see how the contact between glass and PDMS spreads over the whole surface without applying any pressure. After bonding the device can be heated to 65°C for an hour to enforce the bonding.

2.5.2 Loading of the Passivation Buffer

As described for the original Mother Machine device [26] the device needs to be passivated before using it. The passivation buffer consists of salmon sperm DNA (10mg/ml) and BSA (10mg/ml) (mixing ratio 1:3). To prepare the mix, the salmon sperm DNA is denatured at 95°C for 10min and can be mixed with the BSA after cooling down.

In the Dual Input Mother Machine the two waste channels cannot be pressured separately and therefore dried up passivation buffer might block one waste channel without a chance to get rid of the clog. To prevent this from happening it is recommended to load water into the waste outlet while loading the passivation buffer into the cell outlet. To do so, two 1ml syringes are prepared with 23G needles and short pieces of tubing (in our experiments we use PTFE tubing with an inner diameter of 0,56mm and an outer diameter of 1,07mm) that are connected to the syringes by directly putting the tubing over the needles. Now one syringe is loaded with water and the other one with passivation buffer and the end of the tubings are connected to the waste outlet or the cell outlet respectively. To prevent the inlets from drying out later two pipette tips are put into both inlets. Now first the syringe with the passivation buffer is pressured by hand until the main channel is filled up to the DAW junction (can be seen by eye). Afterwards both syringes are pressured equally until the liquid has filled both inlets and some part of the pipette tips. Like this, the device can now be incubated at 37°C for 1h without any blockage of the waste channels or drying out of the inlets. The incubation step should not be much longer than 1h otherwise the passivation buffer can block the main channel.

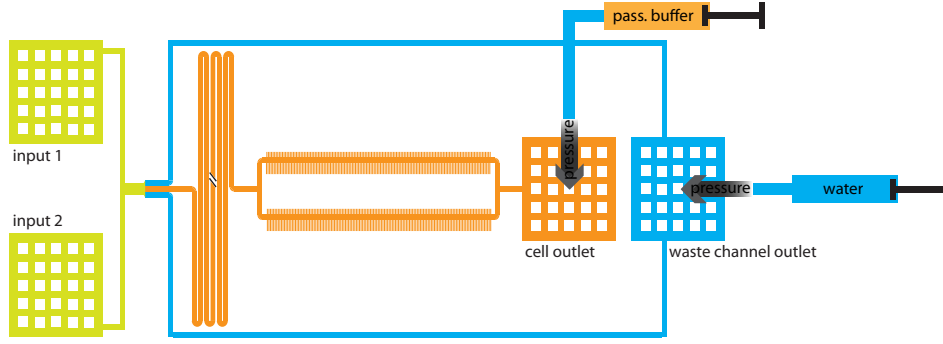


Figure 2.7: Loading of the passivation buffer: While the main channel is filled with passivation buffer loaded through the cell outlet (orange), the waste channels are flushed with water loaded through the waste outlet (blue) at the same time. Therefore the waste channels stay clear of passivation buffer that could dry up and block the channels.

2.5.3 Cell Loading

The night before the experiment an overnight culture with the medium that shall be used first in the experiment is inoculated from a single colony. On the day of the experiment the cells are diluted 1/100 into the same medium and are grown to mid-to-late-exponential phase before the loading. The cell loading step is performed in the microscope and can take some time. Therefore it is recommended to heat up the microscope incubator some hours before the loading step so the system can equilibrate and the cells always stay at 37°C during loading. While the device is incubated with the passivation buffer one can already prepare the two media that shall be used in the experiment. To monitor the flow at the DAW junction one of the media has to be labeled (here non-fluorescent microspheres (Polybead® Polystyrene 1 μ m Beads) were used in a dilution of 1/100 of the original concentration). Syringe size is chosen according to the duration of the experiment and both syringes are equipped with 23G needles. Now the syringes can be loaded with the different

media and the tubing can be connected by putting the end of the tubing directly over the needle.

Once the passivation of the device is done, the short tubing from the loading of the passivation buffer can be removed, the tubing from the medium syringes is connected to the inlets (try to avoid any bubbles) and both pumps are started with a flow rate of $0.025\mu\text{l/s}$ to get rid of the water and the passivation buffer. A new tubing can already be connected to the waste outlet with the other end placed in a waste container. Once the system has equilibrated and all channels are filled with medium the device is mounted on the microscope to check the mixing ratio at the DAW junction. If it fails to reach a 50:50 mixing ratio with a flow rate of $0.025\mu\text{l/s}$ for both syringes the resistance of one of the waste channels is probably altered by a blockage or there is a leak somewhere. If this problem cannot be fixed the device has to be discarded. If one observes a 50:50 ratio one can proceed and change the syringe input to reach the mixing ratio in which the experiment shall be started. To start with a 100% input from only one inlet we use flow rates of $0.04\mu\text{l/s}$ for the active inlet and $0.01\mu\text{l/s}$ for the inactive one. Loading of the cells should only start after all the medium in the outlets has been exchanged with the correct mixture to ensure the cells do not see any other condition before. This will take some minutes depending on the flow rates. It is important that the pumps are always running during the loading step to prevent cells that make it up to the DAW junction from entering the inlets where they could stick and grow.

Now the cells can be taken out of the incubator and have to be concentrated by centrifugation (100-200x). The cells will be loaded through the cell outlet. To do so the tubing that

will serve as a waste tube from the cell outlet after the loading is connected to a 1ml syringe and cells are loaded into the end of the tubing. Now this end is connected to the cell outlet. To fill the main channel the 1ml syringe is carefully pressured by hand and the cells are observed under the microscope while they are entering the main channel. After some minutes the main channel should be filled with a high density of cells and now a custom made screwing clamp can be used to hold the pressure on the syringe to make sure the flow in the main channel stops. It is important to monitor the loading step because the pressure has to be adjusted with the clamp from time to time to ensure zero flow in the main channel and let the cells enter the growth channels. Once a sufficient number of growth channels is loaded one can remove the syringe from the cell outlet tubing, place the end of the tubing in the waste container and let the cells recover for at least 2h before the actual experiment starts.

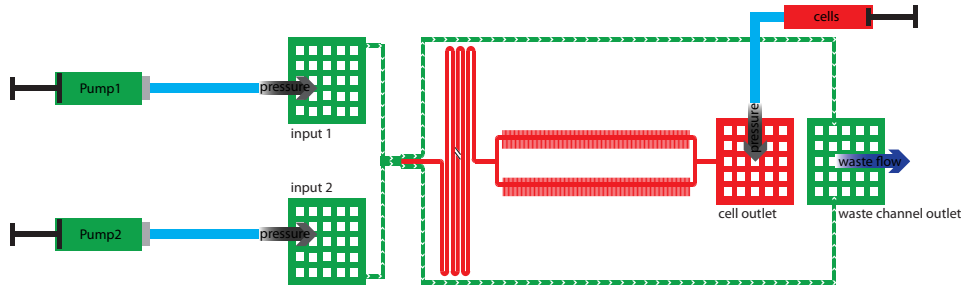


Figure 2.8: Cell loading procedure: There is a constant from both inlets (input 1 and 2) during the whole loading procedure to make sure no cells can enter the inlets during loading. The concentrated cells are loaded through the cell outlet. First the cells are flushed into the main channel. Once the main channel is filled with a high density of cells, the pressure at the cell outlet is controlled to reach zero flow in the main channel (red). The medium flow from the inlets is directed to the waste channels (green) to remove cells that make it up to the DAW junction.

2.5.4 Imaging

For imaging we use an inverted Nikon TI-E microscope with a motorized xy-stage with linear encoders. The complete microscopic setup is placed in an incubator and experiments are typically done at 37°C. For the Mother Machine experiments a CFI Plan Apochromat Lambda DM 100x objective (NA 1.45, WD 0.13mm) and a CMOS camera (Hamamatsu Orca-Flash 4.0) are used to record images. Typically 5 to 6 positions with 26 growth channels per field of view are imaged. For each position an image in phase contrast plus one image for each fluorescent channel is taken. For the phase contrast images a LED transmitted light source at 100% intensity with an exposure time of 100ms is used. The settings for the fluorescence channels depend on the constructs used. For a chromosomal *lacZ-gfp* fusion used in this work, an exposure time of 2s with the Lumencor[®] SpectraX light engine[®] set to 17% output and a ND4 filter in the light path was used. Also the DAW junction is imaged but instead of fluorescence images an additional phase contrast image with a short exposure time (10ms) is imaged to visualize the beads in the flow. Good results were obtained with taking around 30 frames per cell cycle, to have enough data for analysis and still not too much photodamage.

2.5.5 Stopping the Experiment

At the end of the experiment the tubing can be removed from the medium syringes and the device can be discarded. It is recommended to remove all pressure from the syringes before removing the tubing to avoid splashes of medium.

Chapter 3




Main Publication

ARTICLE

DOI: 10.1038/s41467-017-02505-0

OPEN

Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software

Matthias Kaiser¹, Florian Jug², Thomas Julou ¹, Siddharth Deshpande ^{3,4}, Thomas Pfohl ³, Olin K. Silander^{1,5}, Gene Myers² & Erik van Nimwegen¹

Much is still not understood about how gene regulatory interactions control cell fate decisions in single cells, in part due to the difficulty of directly observing gene regulatory processes in vivo. We introduce here a novel integrated setup consisting of a microfluidic chip and accompanying analysis software that enable long-term quantitative tracking of growth and gene expression in single cells. The dual-input Mother Machine (DIMM) chip enables controlled and continuous variation of external conditions, allowing direct observation of gene regulatory responses to changing conditions in single cells. The Mother Machine Analyzer (MoMA) software achieves unprecedented accuracy in segmenting and tracking cells, and streamlines high-throughput curation with a novel leveraged editing procedure. We demonstrate the power of the method by uncovering several novel features of an iconic gene regulatory program: the induction of *Escherichia coli*'s *lac* operon in response to a switch from glucose to lactose.

¹ Biozentrum University of Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50/70, 4056 Basel, Switzerland. ² Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany. ³ Department of Chemistry, University of Basel, Spitalstrasse 51, 4056 Basel, Switzerland. ⁴ Present address: Department of Bionanoscience, TU Delft, Van der Maasweg 9, 2629 HZ Delft, The Netherlands. ⁵ Present address: Institute of Natural and Mathematical Sciences, Massey University Auckland, Private Bag 102904, North Shore 0745, New Zealand. Matthias Kaiser, Florian Jug and Thomas Julou contributed equally to this work. Correspondence and requests for materials should be addressed to O.S. (email: olinsilander@gmail.com) or to G.M. (email: myers@mpi-cbg.de) or to E.v.N. (email: erik.vannimwegen@unibas.ch)

Gene regulation is one of the key processes that underlie the complex behavior of biological systems, allowing cells to adapt to varying environments, and allowing multi-cellular organisms to express a large number of phenotypically distinct cell types from a single genotype. In spite of more than half a century of intense study since the discovery of the basic mechanism of gene regulation¹, much remains to be understood about the ways in which gene regulatory interactions control cell fate decisions. Because of a number of challenges, it is still difficult to directly observe and measure gene regulation in vivo. First, gene regulation is inherently stochastic, and genetically identical cells in homogeneous environments often exhibit heterogeneous behaviors^{2,3}. This implies that bulk expression measurements are often misleading, thus necessitating methods for studying gene regulation in single cells. Second, while methods such as flow cytometry, smFISH, and single-cell RNA-seq provide snapshots of gene expression distributions across single cells (see e.g. refs. ^{3–5}), understanding the processes that shape these distributions often requires that single-cell gene expression be followed in time (e.g. refs. ^{6,7}). The most common approach in such studies is to grow cells on a surface while tracking gene expression and growth using quantitative fluorescence time-lapse microscopy (QFTM).

Three key issues currently limit the power of such studies. First, to capture crucial regulatory events, long-term observations stretching over many cell cycles are often required. Second, measuring gene regulatory responses requires the ability to accurately control and vary environmental conditions. And third, to accurately characterize the statistics of single-cell responses, powerful image-analysis tools are needed to automatically extract large numbers of quantitative phenotypes from the time-lapse measurements. Considering bacteria, while it is possible to expose cells growing on surfaces to changing conditions^{8–10}, gathering long time courses is not possible because the microcolonies grow out of the field of view or start to form multiple layers.

Recently developed microfluidic devices solve this problem by growing cells in micro-fabricated geometries that confine their location and movement^{11–13}. An especially attractive design is the so-called Mother Machine¹¹, in which cells grow single-file within narrow growth-channels that are perpendicularly connected to a main flow-channel that supplies nutrients and washes away cells extruding from the growth channels. However, all current designs expect a single media to be used as input, necessitating manual switching of the input to alter conditions, e.g. refs. ^{14,15}, which precludes the accurate temporal control of the growth environment that is desired to study gene regulation in vivo.

In addition, beyond specific technical problems, many researchers are likely discouraged from studying gene regulation using a combination of microfluidics and time-lapse microscopy, because of the high costs associated with establishing the necessary methods. One not only needs to obtain designs for microfluidic devices, learn how to manufacture these, and work out experimental protocols for performing time-lapse experiments, one also needs sophisticated image-analysis and post-processing methods to obtain accurate quantitative information from the data. While there are a number of software tools for analyzing QFTM data of microcolonies on agar^{16–18}, they perform poorly on data from microfluidic devices such as the Mother Machine, because cells undergo larger movements between consecutive frames. In addition, phase contrast images in microfluidic devices often suffer from non-uniformity due to varying background and opacity. For this reason, most require a dedicated fluorescent reporter to assist segmentation. Although a number of labs are analyzing data from microfluidic devices using various inhouse image-analysis solutions^{11,14,19–21}, there is currently no publically available tool that allows automated analysis of such data with the throughput and accuracy required for quantifying growth and gene expression in large data sets.

To address these problems, we here present an integrated experimental and computational setup for studying gene regulation in single cells using microfluidics in combination with time-lapse microscopy. Our approach consists of the combination of, first, a new microfluidic device, called the dual-input Mother Machine (DIMM), that allows arbitrary time-varying mixtures of two input media, such that cells can be exposed to a precisely controlled set of varying external conditions. Second, to enable high-throughput and high accuracy analysis of phenotypic measurements from the DIMM, we accompany it with a software suite, called MoMA (Mother Machine Analyzer). The Mother Machine Analyzer takes specific advantage of the geometry of the device to accurately segment and track cells using only phase-contrast images, and further provides a curation user interface with leveraged-editing, meaning that a set of related errors are often fixed with a single click. The combination of MoMA's accuracy and curation efficiency allows analyses of data sets involving millions of single-cell observations. Third, we provide several methods for precise quantification and characterization of the accuracy of growth and gene expression measurements. By making the entire framework including the microfluidic device's design, protocols for manufacture and time-lapse experiments, the open source MoMA software, and post-processing methods, all jointly available, we aim to dramatically lower the entrance costs for researchers to adopt this methodology. To demonstrate the power of the method, we apply it to the iconic *lac* operon regulatory system that underlies the discovery of gene regulation, and uncovers several novel unexpected features of its stochastic induction dynamics.

Results

The dual-input Mother Machine. The design of our DIMM device closely follows that of the original Mother Machine¹¹, consisting of a main channel and small dead-end growth channels that open into the main channel (Fig. 1a, c). Nutrients diffuse from the main channel into the growth-channels in which cells are trapped (Fig. 1c), and as the cells in the growth-channels grow and divide, cells closest to the channel's exit are pushed out and are transported away by the flow in the main channel. In contrast to previous designs, our device has dual-input ports and mixing serpentine which, in combination with programmable pumps, allow for arbitrary time-dependent mixing of two input media. The two inputs meet in a dial-a-wave junction²² consisting of two inlets and three outlets (Fig. 1b). While the middle outlet feeds into the main channel of the device, the outer outlets function as waste channels and allow the flow in the middle outlet to vary from carrying only one of the two inputs (black in Fig. 1b), to carrying only the other input (green in Fig. 1b), without getting backflow into the inactive inlet. Note that arbitrary mixtures of the two input media are possible (see Methods, Performance of the environmental control) so that, for example, dynamically changing concentrations of particular nutrients or stressors can be realized. Details on the loading of the DIMM are provided in the Methods (Priming and loading of the microfluidic devices).

To demonstrate the power of our approach, we applied it to the archetypical example of a gene regulatory system: the induction of *Escherichia coli*'s *lac* operon when switching between glucose and lactose as a carbon source. We used a modified *E. coli* MG1655 strain that carries a translational *lacZ-gfp* (green fluorescent protein) fusion at its native locus⁷. Time lapse movies of 22–24 h were obtained in duplicate for three different setups (1. a constant supply of M9 minimal media+0.2% glucose, 2. a constant supply of M9+0.2% lactose, and 3. switching between these two media every 4 h), taking a frame every 3 min (see Supplementary Movie 1

(<https://www.youtube.com/watch?v=2Tznm868fmc> (2015))). Together with additional control conditions (strain without GFP, and switching media where lactose is supplemented with 500 μM IPTG (isopropyl β -D-1-thiogalactopyranoside), we thus

analyzed eight different time-lapse experiments all together, amounting to data from 180 growth-channels, more than 10,000 full cell cycles, and more than 500,000 single-cell observations (Supplementary Table 1).

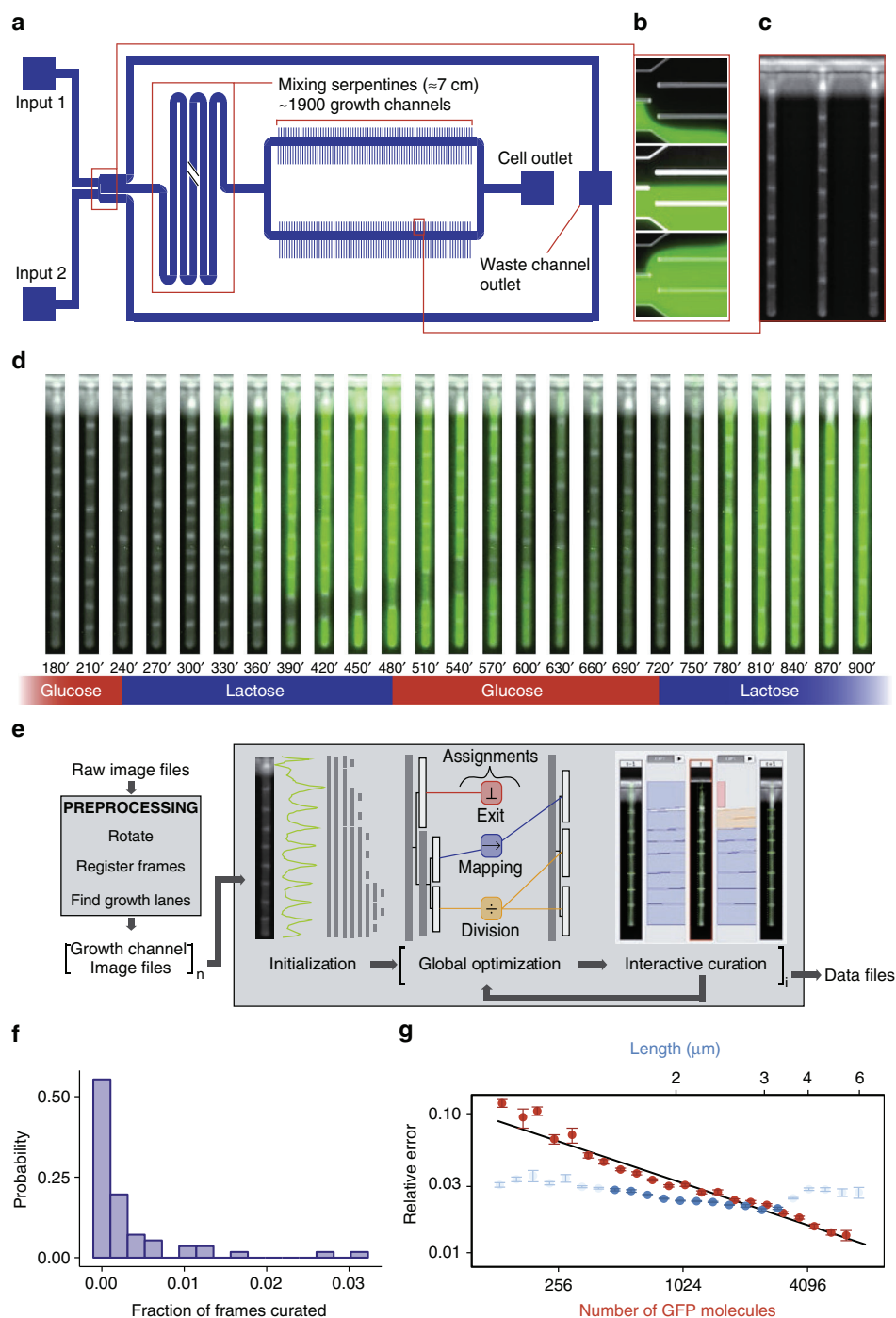


Fig. 1 The dual-input Mother Machine. **a** Overview of the dual-input Mother Machine (DIMM) design. **b** Dial-a-wave junction in three different switching states, top: 100% from input 1 (unlabeled) and 0% from input 2 (green), middle: 50% from both inputs, bottom: 0% from input 1 and 100% from input 2. **c** Phase contrast image of growing *Escherichia coli* cells in three growth-channels of the DIMM. **d** A time series of a single growth-channel containing *E. coli* cells expressing LacZ-GFP from the *lac* promoter while being exposed to media which alternate between containing glucose and lactose as a carbon source. **e** Overview of the automated and curation phases of the MoMA analysis pipeline. **f** Histogram of the fraction of curated frames per single growth-channel time course. **g** Estimated measurement errors on cell size (blue) and number of GFP molecules (red). Dark blue points indicate the typical range of cell sizes. Error bars show standard errors. The black line shows the fitted function $1.01/\sqrt{x}$.

Image analysis and data processing. The analysis of the image sequences acquired by a DIMM is performed in three phases by the MoMA software suite (see Methods, The Mother Machine Analyzer, and following sections). Although MoMA, by default, uses phase contrast images to segment and track the cells, leaving all fluorescent channels for measurement of gene expression and allowing tracking on non-fluorescent (e.g. wild-type) cells, the user can opt to let MoMA use fluorescence images for tracking. The first automated phase begins by registering the frames of a movie to sub-pixel accuracy to correct for jitter and stage drift. Then the growth-channels in each time frame are cropped out and reorganized into a time-series for each channel. Each growth-channel movie is then segmented and tracked. Unlike most image analysis tools that first segment each of the images and then link these segmentations into a tracking, MoMA uses an algorithm that first over-predicts a hierarchy of feasible cell objects (segmentations) for each time point and then simultaneously selects what it thinks are the true cells and the tracking links between them²³. This is accomplished by formulating prior information as a collection of integer linear constraints that guarantee only valid cell trackings satisfy the constraints, and finding among this space of valid trackings, the one of minimum cost. Since cost reflects the likelihood of the solution considering both the observations and prior constraints, this is equivalent to finding the maximum a posteriori solution in Bayesian statistics. We use Gurobi, a potent off-the-shelf integer linear program solver to do so (see Methods).

In the second curation phase, an interactive graphical user interface is opened that allows users to browse the results, identify errors, and correct them. In contrast to existing methods, where data curation is performed by directly editing the segmentations or linking graphs, MoMA offers various possibilities to browse through alternate segmentation hypotheses and tracking paths. Once a user makes an adjustment, e.g. by selecting an alternative segment or link, MoMA formulates the user's choice as an additional constraint and restarts the optimization in order to find the new optimum solution incorporating this constraint. In this way corrections automatically percolate to nearby time points, typically fixing multiple mistakes at once. For the individual growth-channels of the 22–24 h time courses analyzed here, an average 0.3% of frames required a curation directive, and roughly half of the growth-channels required no curation at all (Fig. 1f). In our hands, it typically takes 1–2 min to curate 100 frames (see Methods, Curations statistics).

In the final quantification phase, we developed methods to quantitate the sizes of cells and the amount of fluorescent reporter, as well as to quantify the size of the errors on these measurements. When growing in a constant environment, cell sizes across the cell cycle closely follow an exponential growth curve in both conditions (median squared-correlation $R^2 \approx 0.99$, see Methods, Cell size and growth rate estimation) and this allows us to estimate an upper bound on the errors of individual size measurements, which we find to be approximately 3% (Fig. 1g, and see Methods, Cell size and growth rate estimation). Growth rates of individual cell cycles can be estimated within an error of 1–3% and we find average growth rates of 0.75 (glucose) and 0.69 (lactose) doublings per hour, which vary by 17% across cells (see Methods, Cell size and growth rate estimation). Growth rates during the lactose and glucose phases of the switching conditions have virtually the same distribution as in the corresponding constant conditions (see Methods, Cell size and growth rate estimation).

We observed that cell fluorescence spreads significantly beyond the cell, approximately as a Cauchy distribution as a function of distance from the cell, and we use a Bayesian mixture model to accurately estimate the fluorescence of a given cell (see Methods, Cell fluorescence estimation). This procedure removes auto-

fluorescence due to the PDMS (polydimethylsiloxane) but not the auto-fluorescence of the cell and media. Using measurements on wild-type cells, we observed that auto-fluorescence is proportional to cell size and used this to subtract the contribution of auto-fluorescence to GFP fluorescence measurements (see Methods, Cell auto-fluorescence estimation). Finally, to estimate the conversion factor between fluorescence level and the number of GFP molecules we adapted the method of Rosenfeld et al.²⁴ which is based on the assumption that fluctuations in the fluorescence levels of two daughter cells immediately after division derive from random binomial partitioning of the mother's GFP molecules to the two daughters. We substantially improve on this method by (a) taking advantage of the DIMM design to use data only from the glucose phases in which no GFP synthesis occurs, (b) incorporating the slow fluorescence decay due to bleaching and protein decay (see Methods, Estimating GFP's bleaching and degradation), and (c) taking into account that fluctuations in the sizes of the daughters contribute significantly to the fluorescence fluctuations. We integrated all this into a Bayesian procedure and determined the posterior distribution of the conversion factor between fluorescence and number of LacZ-GFP tetramers (see Methods, Estimating the conversion factor between fluorescence and number of GFP molecules). Using this we find that, when growing in lactose, cells contain 3000–6000 GFP molecules at birth and 6000–12,000 GFP molecules just before division. Finally, we estimated the measurement errors of individual GFP measurements by quantifying the deviations of measured GFP levels from a simple exponential decay during the glucose phases of the switching experiment. In contrast to the relative error on size estimates, which are approximately independent of absolute size, we find that the error on GFP molecule number G scales as $1/\sqrt{G}$ (Fig. 1g), which suggests that this measurement error is likely dominated by shot noise.

One problem encountered with sophisticated image analysis for cell tracking is that methods often poorly generalize to data from setups other than the specific one used in developing the methods. However, MoMA's novel approach in which segmentation and tracking are treated as a joint optimization problem under a system of constraints ensures a high level of robustness to changes in the setup. To directly demonstrate MoMA's general applicability, we reached out to MoMA's emerging user community and obtained time-lapse data sets that were produced in other labs, using different microfluidic devices, different strains and species of bacteria, and different growth conditions (Supplementary Table 2). We confirmed that MoMA shows excellent performance on these data sets, both in terms of the needed curation interactions (Supplementary Fig. 1), and the quality of the resulting growth curves (Supplementary Fig. 2). We find that, depending on strains and conditions, growth rate fluctuations range between 10 and 20% of the average growth rate (Supplementary Fig. 3), and that the accuracy of estimated growth rates is determined to a large extent by the number of measurements per cell cycle (Supplementary Fig. 3).

Single-cell dynamics of the *lac* operon. Figure 2a illustrates how our methodology allows accurate tracking of growth and gene expression across lineages of single cells as the environment is varied. As an example application, we focused our analysis on the single-cell dynamics of *lac* operon induction. Even before the discovery of gene regulation, it was already known that the induction of the *lac* operon is stochastic, with different single cells inducing at different times²⁵. Further support for the stochasticity of this system has come from studies showing that, when exponentially growing populations are treated with artificial inducers

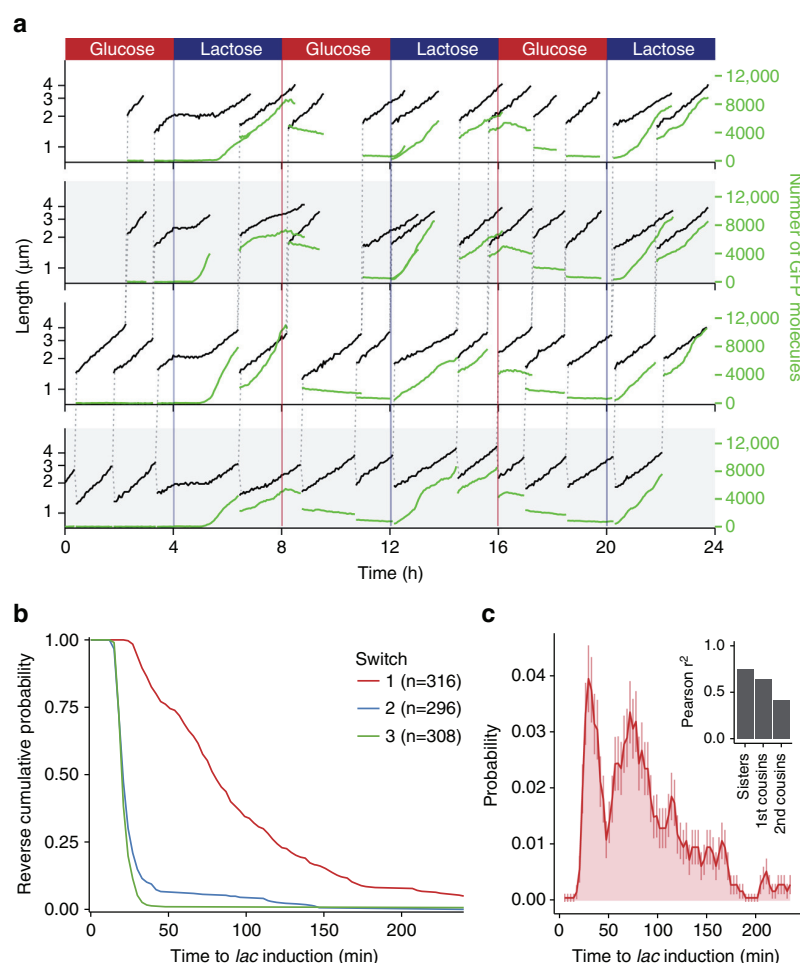


Fig. 2 Tracking single-cell dynamics of *lac* operon induction. **a** Dynamics of growth and gene expression in lineages of single cells in an environment that switches between M9+0.2% glucose and M9+0.2% lactose every 4 h. Cell size (black, logarithmic scale) and expression of LacZ-GFP (green, linear scale) are shown as a function of time for a lineage of cells at the bottom of the growth-channel (bottom row) together with first-generation offspring and second-generation offspring cells (second row from the bottom, and top two rows, respectively). The dashed vertical lines show the lineage of cell divisions by connecting each mother cell to its two daughter cells. **b** Reverse cumulative distributions of lag times to LacZ-GFP induction in single cells at the first (red), second (blue), and third (green) switch from glucose to lactose. **c** Estimated probability distribution (mean and standard deviation) of single-cell lag times for the first switch in 3-min intervals. The inset shows the correlation in lag times for pairs of cells that had the same mother 1, 2, or 3 generations in the past

such as IPTG or TMG (methyl- β -D-thiogalactoside), population snapshots often show a bimodal distribution of *lac* expression in single cells. The current consensus is that, in order for a cell to switch from a low expression to a high expression state, a sufficiently large stochastic burst of *lac* operon expression is needed^{26–28}. A first attempt to measure the distribution of *lac* induction lag times in single cells was made by Boulineau et al.¹⁰, and a wide distribution of lag times was observed. However, the lack of a precise control of the growth media in that work not only precluded accurate time resolution of the lag times, but also caused the switch from glucose to lactose to be so gradual that only some cells experienced a transient reduction in growth rate, while others continued without any change in growth rate.

In contrast, we find that upon a controlled sudden switch from glucose to lactose, the effect on growth is not stochastic at all: all cells completely arrest growth within 3 min of the switch (Fig. 2a). Other aspects that are extremely reproducible are the fact that all cells exit growth arrest as soon as LacZ-GFP is at detectable levels (i.e. 100–200 molecules), and that LacZ-GFP production is halted almost immediately after switching back to glucose (Supplementary Fig. 4). Thus, while induction of the *lac* operon is highly

stochastic, its shutdown and the coupling of growth to *lac* expression appears essentially deterministic.

Interestingly, while it might have been expected that, after exiting growth arrest, initial growth rates would be low when LacZ levels are still far below steady-state levels, we find that cells immediately grow at rates that are close to those observed in constant lactose, and reach steady-state growth rates within an hour of induction (Supplementary Fig. 5). We estimated instantaneous growth rate as a function of LacZ-GFP concentration and found only a substantial decrease when concentration is more than 10-fold below the steady-state levels of 2000–3000 molecules per μ m of cell length (Supplementary Fig. 5). That is, cells can sustain high growth rates in lactose with *lac* operon expression that is fivefold or more below steady-state levels, in line with previous observations¹⁰. This raises the question as to why LacZ steady-state levels are so much higher than required for growth. One intriguing possibility is that such high expression levels allow for a memory of growth in lactose that lasts over several generations, something that has been observed previously at the population level¹³. Indeed, during the glucose phase the total fluorescence in each cell shows a slow exponential decay,

mostly due to bleaching, and approximately halves at each cell division (Fig. 2a). By the time of the second switch to lactose, LacZ-GFP levels have diluted back to low levels, but the remaining *lac* expression is enough to ensure that all progeny of cells that induced in the first switch continue growth without an obvious decrease in growth rate, and quickly recommence LacZ-GFP production (Supplementary Fig. 6).

Our methodology allows, for the first time, the accurate measurement of the distribution of lag times for single cells to exit their growth arrest after the first switch from glucose to lactose. We not only observe a wide distribution of lag times, but find that this distribution is multi-modal: 27% of cells induce within 25–45 min, 68% induce within 50–240 min, and 5% of cells do not induce at all (Fig. 2b, c). Importantly, this observation is directly at odds with the current view in the literature that all lags are determined by the waiting time to a single stochastic event. Instead, the multi-modal distribution suggests that naive cells can exist in different states that determine their ability to respond to lactose.

We investigated whether lag times correlate with simple physiological quantities such as growth rate, cell cycle stage, or LacZ-GFP levels at the time of the switch, but found that none of these variables correlate with lag times (Supplementary Fig. 7). However, we find strong correlations of the lag times of cells that had the same mother, grand-mother, or even great-grandmother cell (Fig. 2c and Supplementary Fig. 8). It is especially striking that these genealogical correlations are larger for lag time than for any other physiological quantity that we measured, including quantities such as cell size and LacZ-GFP concentration, that are known to be directly inherited from the mother (Supplementary Fig. 9). In particular, only lag time shows significant correlations in cousins and second cousins. These results strongly suggest that lag time is controlled by an inheritable epigenetic factor that, in contrast to other physiological quantities such as LacZ-GFP expression, growth rate, and cell size, shows significant correlations over 2–3 generations.

Although a full investigation of the mechanistic interpretation of the multi-modal lag time distribution is beyond the scope of this work, we can propose an interpretation that we consider most plausible. We propose that the first and second modes of the lag distribution correspond to cells that, at the time of the switch, have either nonzero expression of both LacZ-GFP and LacY permease, or zero expression of either of these molecules. When both LacY and LacZ-GFP molecules are present at the switch, lactose can presumably immediately enter the cell, where it is metabolized into allolactose, causing *lac* operon derepression and LacZ-GFP production. In contrast, when no LacY/LacZ-GFP is present, lactose can either not enter the cell, or it cannot be metabolized, and cells first have to wait for a stochastic burst of *lac* operon expression, causing a longer lag time. If this interpretation is correct, then no long lags should be observed when an artificial inducer is added that can diffuse into the cell without LacY permease and binds directly to the LacI repressor. Indeed, when we add IPTG to the medium containing lactose we only observe short lags (Supplementary Fig. 10). Our interpretation also requires that, when growing in glucose, the majority of cells should contain either no LacY or no LacZ-GFP at all. This prediction is consistent with our LacZ-GFP measurements in glucose that predict the distribution of lacZ-GFP per cell significantly overlaps zero molecules (Supplementary Fig. 11). It is also broadly consistent with previous observations that, in similar growth conditions, roughly 50% of cells contain no LacY²⁷, and 65% of the cells contain no LacZ²⁹. Finally, we note that Choi et al.²⁷ estimated that, when growing in the absence of lactose, small bursts in which around 40 LacY molecules are produced occur every 2–3 cell cycles, which is consistent with the

waiting times of up to 240 min that we observe for cells of the second mode of the distribution.

Discussion

We have here presented an integrated experimental and computational setup for quantifying gene expression dynamics at the single-cell level over long periods of time in dynamically changing environments that are precisely controlled. This methodology opens up a wide array of possibilities for studying gene regulatory mechanisms at the single-cell level. A single experiment with our setup was able to uncover several novel and unexpected features of one of the most intensely studied regulatory systems: *lac* operon expression under growth conditions that change between glucose and lactose as a carbon source. However, the technology enables many other types of investigations, e.g. it can be used to quantify how expression fluctuations affect growth rates at the single-cell level, to investigate how regulatory responses depend on the concentration and length of exposure to an inducing nutrient or stress, and how memories of regulatory responses are maintained across lineages of cells. More generally, its power extends beyond the scope of gene regulation studies. For example, it is becoming increasingly appreciated that single-cell heterogeneity plays an important role in persistence and evolution of resistance to antibiotics, and our methodology could be used to accurately quantify how single-cell growth and survival varies as a function of both the concentration and time of exposure to particular antibiotics, and as a function of the physiological states of the cells. In summary, we believe that the integrated experimental and computational methodology that we present here will be an important tool for studying gene regulatory mechanisms at the single-cell level. As detailed in the Data Availability section below, to facilitate access of other labs to our integrated methodology, we have collected all relevant information in a web repository, including files with the CAD designs of the device, information on fabrication of the device, detailed protocols for running the experiments, and links to the open source MoMA software. MoMA and its documentation, including tutorial videos are available online³⁰ and, to make MoMA easily available to any user of ImageJ, we have also made MoMA available as a Fiji plugin.

Methods

Design and fabrication of the microfluidics devices. *Escherichia coli* cells take on different sizes depending on the media they are grown in, e.g. LB versus M9 minimal medium. Since the growth-channels aim to trap the cells growing in single file, the width of the channels needs to match the width of the cells as closely as possible. To account for this, our DIMM device contains channels with a range of widths, ranging from 0.8 to 1.6 μm , and lengths of 25 μm on one side of the device, and 55 μm on the other. For the results presented here, the growth-channel sections were $\sim 0.9 \mu\text{m}$ (height) $\times \sim 0.8 \mu\text{m}$ (width), and 25 μm (length). These dimensions worked nicely with cells growing in M9+0.2% glucose or 0.2% lactose respectively. Experiments with other media and strains might require slightly different dimensions. In order to reduce the flow rates compared to the original mother machine device, the dimensions of the main channels were reduced to a diameter of 6 μm (height) by 50 μm (width) in the device presented here. The resulting flow rates are discussed in more detail in the section discussing loading and flow control. We note that reflections from the PDMS in the main channel can affect the phase contrast images near the top of the growth-channels, such that a segment of the growth-channels nearest to the exit needs to be discarded. To minimize this effect it is advisable to keep the main channel relatively shallow.

The device was designed using AutoCAD® (AUTODESK®) and is freely available at Metafluidics, an open repository for fluidic systems³¹. We outsourced both the production of the photomask and the production of the masters to pour the PDMS devices from. A 5" quartz mask with chrome layer was ordered from the Compugraphics Jena GmbH. Using this mask, Microresist (Berlin) produced the master using UV-lithography with SU-8 photoresists (for more details see ref. ¹¹). To make the chips, we use the Sylgard Elastomer Kit 184 with a 1:10 curing agent to base ratio. Curing was performed at 65 °C overnight or longer. Harris Uni-Core 0.75 mm biopsy punches were used to create in- and outlets. Before bonding, both the PDMS mold and a cover slip were washed with isopropanol and dried with air. Surface activation was done in a plasma cleaner (PDC-32G, Harrick Plasma)

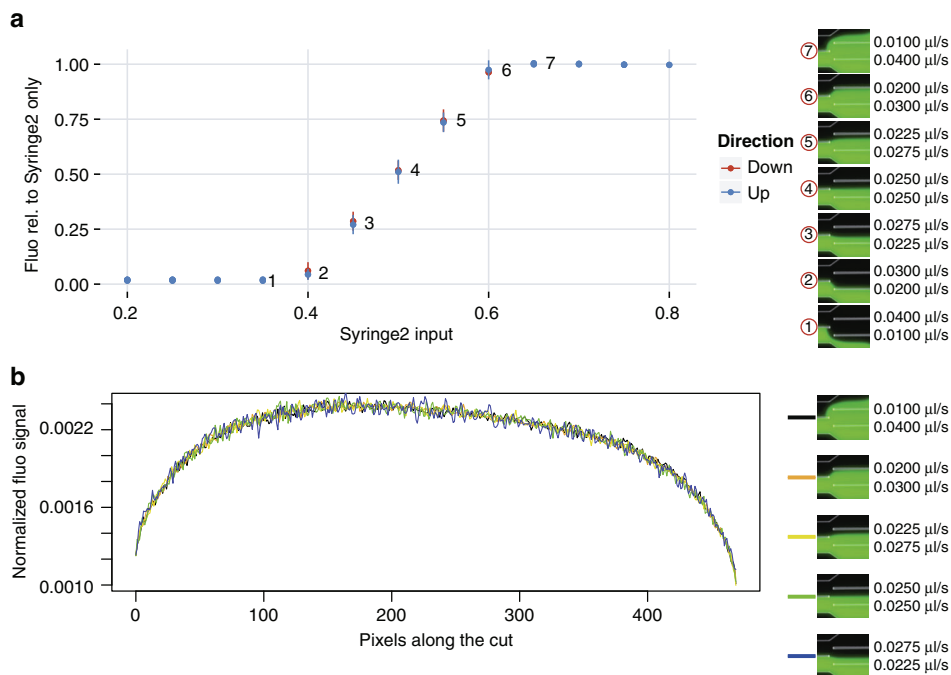


Fig. 3 Mixing of fluorescein-labeled medium with non-labeled medium at different input flow rate ratios. **a** Total fluorescence was measured in a square region in the main channel downstream of the mixing serpentine as the input ratio was changed in a stepwise manner from 0% fluorescein input to 100% fluorescein input (up: blue) and back to 0% fluorescein input again (down: red). The fluorescence measured for the mixture relative to the fluorescence measured for pure fluorescein-labeled medium (Syringe 2 only) is plot against the input of Syringe 2 as a fraction of the total input from both Syringes 1 and 2. **b** Normalized fluorescence along a section through the main channel downstream of the mixing serpentine at different mixing ratios

Table 1 Experimental conditions used in this study		
Condition	Sequence of growth media	Strain
No GFP	12 h: M9+0.2% glucose 12 h: M9+0.2% lactose	MG1655
Glucose	22 h: M9+0.2% glucose	ASC662
Lactose	22 h: M9+0.2% lactose	ASC662
Switch	4 h: M9+0.2% glucose 4 h: M9+0.2% lactose 4 h: M9+0.2% glucose 4 h: M9+0.2% lactose 4 h: M9+0.2% glucose 4 h: M9+0.2% lactose	ASC662
Switch IPTG	4 h: M9+0.2% glucose 4 h: M9+0.2% lactose +500 μM IPTG 4 h: M9+0.2% glucose 4 h: M9+0.2% lactose +500 μM IPTG	ASC662

Strain MG1655 is a reference K12 strain⁴⁶, and ASC662 was derived from it by integrating a translational fusion *lacZ-gfp* in the native *lac* operon⁷. Note that for each condition, the first step of its sequence of growth media was preceded by 2 h in the same media (in order to reach growth steady-state under fluorescence illumination conditions) that were discarded from the data analysis

operated at high intensity with vacuum at 1500 mTorr for 40 s. After bonding the devices were incubated at 65 °C for at least 1 h.

Performance of the environmental control. The design presented here not only allows switching between different media but also allows for continuous control over the ratios of two different input media. Because flows in micro-channels are strictly laminar, only diffusive mixing occurs at these scales²². To keep the design

simple we introduced 2D mixing serpentine to the device. These serpentine guarantee that the media coming together in the junction are flowing together long enough to allow for diffusive mixing before the mix reaches the cells. The required length of these mixing serpentine depends on the flow speed (fluid velocity), the width of the micro-channels, and the diffusion coefficient of the molecule of interest in the medium used³².

To demonstrate mixing in the device we used M9 minimal medium labeled with fluorescein (1 g/ml) (Syringe 2) mixed with unlabeled M9 minimal medium (Syringe 1). We first obtained a reference fluorescence level for the medium containing fluorescein by measuring fluorescence every 15 s for 70 min, and taking the average of these 280 measurements. For 13 different relative flow rates of the two syringes, ranging from 20% of the total flow from Syringe 2 to 80% of the total flow from Syringe 2, we then measured fluorescence every 15 s for 10 min (40 min) and divided this by the reference fluorescence level to obtain a relative fluorescence. We then calculated the mean and standard deviation of 40 relative fluorescence levels for each relative flow rate. The results are shown in Fig. 3a, demonstrating how the system presented here can generate different mixing ratios and thus can be used to precisely control the growth environment. Figure 3b shows the normalized fluorescent intensity along a section through the main channel downstream of the mixing serpentine at different flow regimes. Because of small imperfections in the mold the intensity profile is not perfectly symmetrical even in the unmixed state (black line). However in the different mixed states, the shape of the profile stays the same indicating complete mixing is guaranteed in the flow regimes tested here.

Strains and growth conditions. Strains were streaked from freezer stocks onto LB plates before experiments. Overnight precultures were grown from single colonies in M9 minimal medium supplemented with the same carbon source that the cells were to experience in the initial phase of the experiment (0.2% glucose or 0.2% lactose). The next day, cells were diluted 100-fold into fresh medium with the same carbon source. Cells were harvested after 4–6 h to be concentrated and loaded into the microfluidic device (typically, a culture at OD \approx 0.2 was concentrated 100-fold). Growth occurred at 37 °C for both the precultures and the microscopy experiments. The growth conditions used during the microscopy experiments are described in Table 1.

Priming and loading of the microfluidic devices. The DIMM design presented here has two inlets and two outlets. This leads to some complications in the cell loading process compared to the original Mother Machine design. Here we describe the adjusted loading procedure we developed. As described in ref. ¹¹, a mixture of salmon sperm DNA (10 mg/ml) and BSA (bovine serum albumin, 10

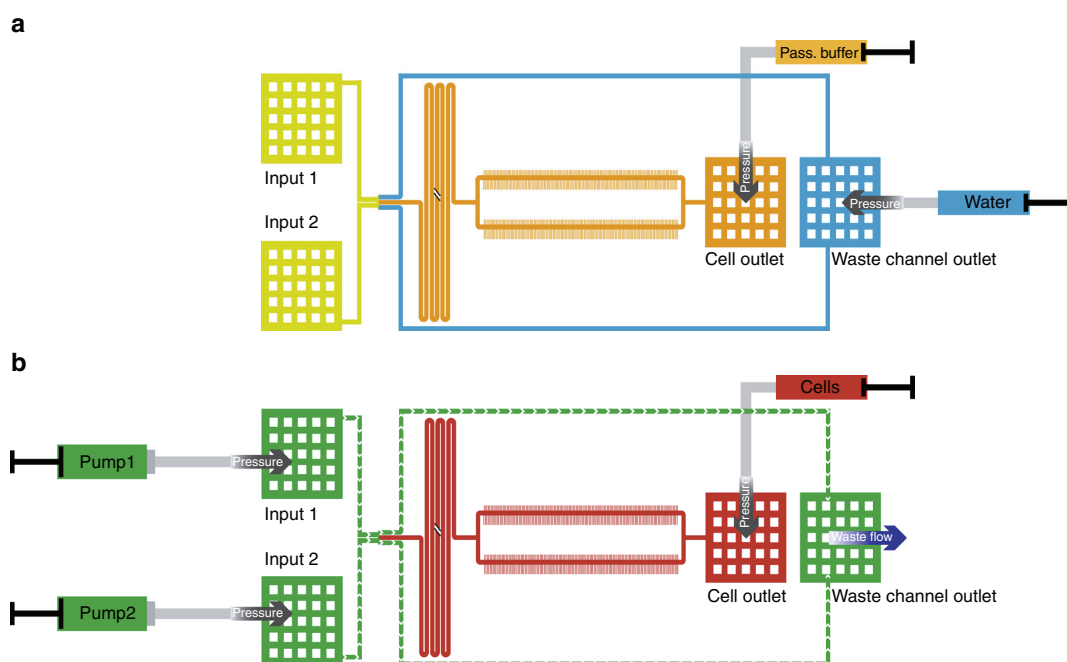


Fig. 4 Priming and loading of the device. **a** Passivation buffer loading. To prevent blocking of the waste channels by passivation buffer, the waste channels are loaded with water through the waste channel outlet (blue) while loading of passivation buffer is done through the cell outlet (orange). Putting both outlets under pressure assures complete loading of the main channel with passivation buffer while the waste channels stay clear of passivation buffer. **b** A constant flow in both inlets (input 1 and input 2) prevents cells entering the inlets during the loading. The concentrated cell solution can be loaded through the cell outlet. First some pressure is applied to fill the whole main channel with cells. Afterwards the pressure is controlled to maintain zero flow in the main channel (red) while there is a constant flow through the inlets and in the waste channels (green) to remove cells that make it up to the dial-a-wave junction

mg/ml) (at a ratio 1:3) is used to passivate the device before loading the cells. The salmon sperm DNA is denatured at 95 °C for 10 min and is mixed with the BSA after cooling down. This passivation buffer is also added to the growth medium in the experiment in a concentration of 1/100. In addition, one medium was always labeled with non-fluorescent microspheres (Polybead® polystyrene 1 µm beads) to monitor medium flow at the dial-a-wave junction. As shown in Fig. 4a, the two dial-a-wave waste channels cannot be pressured separately because they both end in the same outlet. Therefore to prevent blockage of one of the waste channels by passivation buffer it is recommended to flow water into the waste channel outlet while the passivation buffer is loaded into the cell outlet. Once the main channel (with the growth-channels) is filled with passivation buffer and the inlets (input 1 and input 2) are full of liquid (mixture of water and passivation buffer), both the flow of water and of passivation buffer can be stopped. The device is now incubated for ca. 1 h at 37 °C before the loading of the cells is started.

After the passivation step, cell loading can begin. To get rid of the passivation buffer, the two inlets are connected to the pumps with the two different media that will be used in the experiment. At this point the tubing for the waste outlet can also be installed and connected to a waste container. Both pumps are now set to a flow rate of 1.5 µl/min. When all channels are clear, this flow regime will lead to a 50:50 ratio between the two inputs at the dial-a-wave junction. If the device leaks at this point or fails to establish a 50:50 ratio at the dial-a-wave junction (one medium is labeled with beads to check the flow under the microscope), most likely the resistance of some channel is altered by a blockage and the device cannot be used. If the device works properly, the dial-a-wave junction can be switched to the medium that will be used first. This step is necessary to ensure that the cells that are loaded afterwards only encounter the media condition in which they will begin growth. For a complete switch we use flow rates of 0.6 µl/min for the inactive inlet and 2.4 µl/min for the active one (Fig. 3a). After a few minutes (depending on the flow rate) the main channel and cell outlet should be free of the medium from the initial input and the cell loading process can begin. The cells are harvested in exponential phase and are concentrated by centrifugation (~100–200×). Once the device is fully switched to the desired input, one can load the cells using a 1 ml syringe into the tubing that will later serve as the waste tube. This tubing is inserted into the cell outlet and can be pressured by hand to flow the cells into the main channel (the loading process is observed under the microscope). It is important not to stop the flow at the inlets during the whole loading process. This allows cell loading without getting cells into the inlets where they might become stuck and grow. Once the cells reach the growth-channels we used a custom-made clamp to hold a precise level of pressure on the 1 ml syringe for cell loading. The pressure here has to be

continuously adjusted to make sure the cells stop flowing in the main channel and can enter the growth-channels. As shown in Fig. 4b there is a constant flow through the inlets and the waste channels (green) while the main channel is pressured to achieve zero flow where the growth-channels are (red). If cells move up to the dial-a-wave junction they are removed through the waste channels and the inlets stay clear. Loading continues until a satisfactory number of channels contain cells (typically 20–60 min). When complete, the 1 ml syringe used for loading is removed, and the end of its tubing is put into the waste container together with the tubing from the waste channel outlet. After loading the cells are allowed to recover for at least 2 h before the experiment starts.

Growth media are delivered from air-tight glass syringes (Hamilton) that are connected to the device using PTFE tubing with an inner diameter of 0.56 mm and an outer diameter of 1.07 mm. The syringes are controlled by two low pressure pumps (Cetoni GmbH) so that the total flow during the experiment is 3 µl/min.

Microscopy and data preprocessing. An inverted Nikon TI-E microscope equipped with a motorized xy-stage with linear encoders was used to perform the experiments. All experiments were performed in an incubator maintained at 37 °C. The sample was fixed on the stage using metal clamps and focus was maintained using the Perfect Focus System from Nikon. Images were recorded using a CFI Plan Apochromat Lambda DM ×100 objective (NA 1.45, WD 0.13 mm) and a CMOS camera (Hamamatsu Orca-Flash 4.0). The setup was controlled using Micro-Manager³³ and timelapse movies were recorded with its Multi-Dimensional Acquisition engine (customized using runnables). Every 3 min one phase contrast image and one GFP fluorescence image were acquired, typically for six different positions. Phase contrast images were acquired using 100 ms exposure with the transmitted light source at full power (CoolLED pE-100). Images of GFP fluorescence were acquired using 2 s exposure, illuminating the sample with a Lumencor SpectraX (Cyan LED) set to 17% and dimmed using a ND4 filter in the light path; the excitation (475/35 nm) and emission filters (525/50 nm) were used with a dichroic beam-splitter at 495 nm. For the switching experiments images of the dial-a-wave junction were also acquired. Here the GFP channel was replaced with an additional phase contrast image with a short exposure time (10 ms) to visualize the beads in the flow.

The MoMA tracking software expects to be given image data sets in which a single growth-channel is present, with the growth-channel opening at the top, and with phase contrast being the first channel. With our microfluidic design, the camera field of view covers ca. 30 growth-channels so the images must be split into

individual growth-channels and preprocessed in order to match MoMA's requirements. The preprocessing consists of the following tasks:

1. Load the microscopy data set, one position at a time, in a format-independent manner using the Bio-Formats library (in order to open a specific position, one must use the Java API rather than functions available in ImageJ).
2. Register all frames to the first frame of the first channel in order to correct the sample drift over time, as well as the jitter introduced by acquiring multiple positions in parallel. To do this, we develop HyperStackReg, a custom extension of the StackReg ImageJ plugin that is able to handle hyperstacks, i.e. data sets with several channels.
3. Crop the image to keep only the area of the growth-channels and rotate the images (so that the growth-channel opening is at the top).
4. Save images as a tiff data set with one file per frame.
5. Straighten the image so that growth-channels are oriented vertically (using bicubic interpolation).
6. Identify the growth-channels in the first-phase contrast frame and save one data set per cropped growth-channel.

All steps are run in Fiji with the help of two utility plugin released together with MoMA: HyperStackReg and MMPPreprocess. This preprocessing step is documented extensively on MoMA's Wiki³⁰, including how to run it from the command line. Note that in order to preprocess data sets from the command line, Fiji must be run using a virtual window environment (using Xvfb), since the headless mode is not compatible with some important ImageJ features.

The Mother Machine Analyzer. Today's predominant tracking methods originated in the 1960s^{34,35} and were developed to track single or a hand-full of objects with complex distinguishing features such as ships or airplanes. However, here we require the tracking of objects that are visually almost identical. In some cases this can be resolved by maintaining multiple association hypotheses over multiple time points³⁶. However, although particle trackers and state space models can produce high-quality results, proofreading (data curation) is always required in order to guarantee error-free tracks. Notably, computer-assisted approaches for proofreading are usually not related to the method that produced the automated results in the first place.

Interactive error correction is rarely part of available tracking systems and usually turns out to be difficult to implement and integrate, leaving the user with an inflexible patchwork of multiple tools. Part of the reason for this is the way classical tracking models work. Their local and iterative solvers are highly specialized, not offering intrinsic possibilities to constrain the space of possible solutions in a user-driven way. In other words, they intrinsically do not offer any interaction capabilities that can be employed by users to prevent the tracking system from making certain mistakes.

Assignment Models promise to make a difference in all these respects. The novelty of this type of tracking system is the way in which solutions are found. A tracking problem is formulated as a global optimization problem under constraints that can be solved using discrete optimization methods. MoMA is based on such an optimization-based assignment model that allows the user to furnish constraints in an interactive manner. Hence, we can offer unprecedented user interactions for data curation—a process we call leveraged editing.

In particular, MoMA offers the following leveraged editing primitives: (i) Forcing solutions to contain a selected cell (segment), (ii) forcing solutions not to include specific segments, (iii) forcing a cell to a given movement or division (assignment), or to (iv) avoid such an assignment, and (v) specifying the number of cells visible at a given time. We will show that the very nature of the underlying optimization problem allows us to seamlessly incorporate these leveraged editing primitives.

Automated tracking with MoMA. Here we briefly review the class of tracking methods called assignment models^{23,37–40}. We provide sufficient technical detail to prepare the reader for later sections, introducing leveraged editing primitives used in MoMA.

Tracking consists of two equally important tasks: Cells need to be segmented in each frame, and segments of the same cell in consecutive frames need to be linked. Tracking by assignment approaches these tasks by formulating and solving a joint global optimization problem. In this context, the segmentation task consists of selecting a subset of segments in each image, i.e. corresponding to the cells in the image. To do this, the algorithm first generates a large collection of possible segment hypotheses that are contained in a (possibly heavy) oversegmentation of the images. Joint segmentation and tracking then boils down to enumerating many potential subsets of segments together with potential ways of linking (assigning) these between consecutive frames. To identify, among all these possible joint segment/assignment subsets, an optimal solution, each of the potential segments and assignments is given a cost. The cost of a joint segmentation/assignment hypothesis aims to reflect how unlikely it is that the corresponding dynamics occurs in the real system, i.e. the corresponding movement, growth, and division of the cells in our system. That is, the total cost can be thought of as a negative log-likelihood of the segmentation/assignment hypothesis^{38,41} and an optimal solution minimizes this cost.

The cost function is designed to reflect the knowledge of domain experts. To give an example, in our application, the cost function for a cell division assignment that links one segment to two segments in the next frame contains a term that evaluates the size of the three segments to be linked which implements the physical constraint that the sum of the sizes of the two daughter cells should be similar to that of the mother cell. Structural knowledge about which assignments can be chosen simultaneously is encoded in terms of constraints that ensure that only physically meaningful solutions can be chosen. That is, solutions that do not describe impossible events like cells popping into existence out of nowhere, cells moving to two places at once, etc. In our implementation, these constraints force or prohibit certain segments and assignments to be jointly contained in a segmentation/assignment solution. Notably, in formulating these constraints we of course take advantage of the fact that the microfluidic device organizes cells into one-dimensional arrays.

Once the segmentation and tracking problem has been formalized in this manner in terms of costs and constraints, well-established discrete optimization methods can be used to obtain a solution that is (i) feasible, i.e., free of conflicts, and (ii) cost-minimal. In the following we will put these notions on formal grounds. A more in-depth description can be found in ref. ²³, where we described in detail the assignment model upon which MoMA operates. In the next section we will briefly summarize this model in order to lay the foundation to understand the leveraged editing primitives introduced thereafter.

The assignment tracking system in MoMA. First, an excess of segment hypotheses $H^{(t)}$ is generated for each frame t , with many hypotheses partially overlapping and thereby providing alternative and mutually exclusive interpretations of where the cells are appearing in the image²³. To represent possible solutions, a binary segmentation variable $h^{(t)}$ is associated to each possible segment hypothesis in $H^{(t)}$. Whenever $h^{(t)} = 1$, it indicates that this segment hypothesis is part of the proposed solution. Similarly, a set of assignment hypotheses $A^{(t)}$ and associated binary assignment variables $a^{(t)}$ are generated, that link segment hypotheses at time point t to segment hypotheses at $t+1$. For example, a mapping assignment $a_{i \rightarrow j}^{(t)}$ connects two segment hypotheses $h_i^{(t)}$ and $h_j^{(t+1)}$.

Thus, a proposed segment/assignment solution consists of a selection of binary segmentation and assignment variables v that are set to 1. As mentioned above, a cost function is defined that associates to every such variable v , a cost $c_v \in \mathbb{R}$ of including it in a solution. For details on the cost function used for mother machine devices, we refer to ref. ²³. In a nutshell, the cost measures how much a segment/assignment deviates from the expected appearance/dynamic behavior of bacterial cells. The total cost C of a particular solution is then simply the summed cost over all active variables

$$C = \sum_i v_i \cdot c_{v_i}. \quad (1)$$

Linear constraints are used to restrict the solution space to only include conflict-free and structurally sound solutions. As a simple example, two segment hypotheses that offer conflicting explanations of a particular pixel cannot simultaneously be active in any feasible solution. To introduce some notation that will be required below, we look in detail at one particular constraint. Continuity constraints ensure that each active segment at frame t (i.e. each cell) must be involved in exactly one assignment entering from frame $t-1$ and must also be involved in exactly one assignment towards $t+1$. In other words, each cell must have a past and a future. Formally, this statement can be written as

$$\forall t \in \{2, \dots, T-1\}, \forall h^{(t)} \in H^{(t)} : \sum_{a^{(t-1)} \in \Gamma_L(h^{(t)})} a^{(t-1)} = h^{(t)} = \sum_{a^{(t)} \in \Gamma_R(h^{(t)})} a^{(t)}. \quad (2)$$

Here we image time frames ordered from left to right and use the notation $\Gamma_L(h)$ to denote the set of assignments directly to the left of segmentation variable h (i.e. its left neighborhood) and $\Gamma_R(h)$ to denote the set of assignments directly to the right of h (its right neighborhood). That is, the left neighborhood $\Gamma_L(h)$ is the set of all assignments entering h from the previous frame and the right neighborhood $\Gamma_R(h)$ is the set of all assignments leaving h towards the next frame. The equation above then says that, for each cell at time t , there should be one assignment in the previous time frame, and one in the following time frame.

A globally optimal solution, i.e. picking a set of conflict-free assignments of minimal cost can be achieved by solving an integer linear program (ILP)^{23,38–40}.

An ILP is an optimization problem that is fully specified by (i) an objective function that is a linear function of a set of variables \mathcal{V} , and (ii) a set of constraints that are formalized as (in-)equalities on these variables. The space of feasible solutions is defined by all variable assignments that obey all constraints. An optimal solution is a feasible solution that minimizes the objective function.

The joint segmentation and tracking formulation introduced above is already in ILP form: The set of variables \mathcal{V} comprises binary segmentation and assignment variables. The objective to minimize is the cost C defined in Eq. (1). Note that this is a linear function of the variables $v \in \mathcal{V}$. In Eq. (2) we also gave an example of how constraints can be formalized as linear equalities.

Integer linear programming is a well-understood problem⁴², and given the above formulation we can turn to off-the-shelf solvers like Gurobi to find an

optimal solution. Although finding an optimal ILP solution is NP-hard, recent success solving relatively large tracking problems^{23,38–40} suggests that assignment models pose well-natured instances to be solved as ILPs.

In the following we will make use of a particular feature of many ILP solvers, namely the ability to perform “warm-starts”. One speaks about a warm start if a solver can benefit from residual intermediate results created during a preceding optimization. This can speed-up optimization significantly as shown in ref. ²³.

Additional performance for solving the ILP underlying a tracking instance can be gained by reducing variable redundancy via substitution. The set of variables \mathcal{V} contains variables h for available segments, and a , for available assignments. However, note that whenever the segmentation variable for a segment i is active, i.e. $h_i = 1$, then at least one assignment a that involves a segment i must be active as well. Using these constraints, the segmentation variables can be removed from the model entirely²³. That is, after adequate constraints are added to the ILP, each occurrence of $h_i^{(t)}$ can be substituted by a sum over all assignment variables in $\Gamma_L(h_i^{(t)})$ (or $\Gamma_R(h_i^{(t)})$).

Leveraged editing of tracking solutions. In this section we discuss how MoMA modifies the underlying optimization problem in response to user feedback. MoMA first of all provides the user with a graphical interface that allows the user to browse through the tracking solution that the optimization has provided for a given movie. The basic idea of leveraged editing is simple: When a user identifies a segmentation or tracking error, (s)he suggests the correct alternative or simply points at the error in the graphical interface, leaving the algorithm to search for a corrected solution to the model. In MoMA, the given feedback is incorporated into the ILP via additional constraints. Using warm-starts allows optimizing the modified problem fast enough for interactive use. Fixing a single error will usually resolve a bulk of transitive errors. These interaction-based modifications and re-optimizations are iterated until the found solution is satisfactory to the user, i.e., appears to be free of errors.

Here we introduce five specific interaction primitives implemented in MoMA. We will see that they do not introduce significant changes to the existing assignment tracking formulation and can be implemented efficiently. To illustrate how leveraged editing works in practice, a tutorial movie is available on MoMA's Wiki page³⁰, showing several of these primitives in action.

One possible error is that the tracking may have failed to include a particular cell, possibly even across multiple frames. In this case, the user wants to choose an adequate segment and force it to be included in the tracking solution. In MoMA this can be achieved by hovering the mouse over the part of the image where a cell was not picked up by the original optimization. Segment hypotheses located at the mouse position will be highlighted interactively, and simply clicking on any highlighted segment will cause (i) adequate modifications of the ILP (as described below), and (ii) a re-run of the solver to obtain an optimal solution for the given data, now constrained to include the forced segment.

Technically this can be achieved by adding a single constraint to the ILP, namely $h_i = 1$ where h_i is the chosen segment. Applying the redundancy reduction discussed in the previous section, the constraint to be added can be expressed in terms of assignment variables as

$$\sum_{a \in \Gamma_R(h)} a = 1, \quad (3)$$

where $\Gamma_R(h)$ is the right neighborhood of h , i.e. the set of all assignments leaving h towards the next frame.

In addition to allowing users to force missing segments to be included, the user can also tell MoMA to exclude certain segments from solutions. The re-solved ILP will correspond to the minimal cost solution for the data, constraint to exclude the chosen segment. Analogously to forcing segments, the constraint to be added to the ILP is

$$\sum_{a \in \Gamma_R(h)} a = 0. \quad (4)$$

Instead of interacting with segments, a user might want to directly work with individual assignments. To do so, users can browse through a library of available assignments. Assignments can be included or excluded from tracking solutions.

Browsing the library of available assignments can be done in only a few mouse-clicks. Since there is precisely one binary variable a corresponding to the chosen assignment, the constraint to be added to the ILP to force or exclude this assignment is simply $a = 1$ and $a = 0$, respectively.

The last interaction primitive of MoMA is particularly powerful, often capable of fixing multiple tracking errors at once. The idea is simply to let MoMA know how many cells are contained in a given time point. We constrain the solution space to only allow solutions that contain k segmented cells at time point t . Formally this is accomplished by adding the constraint

$$\sum_{h \in H^{(t)}} \sum_{v \in \Gamma_R(h)} v = k, \quad (5)$$

where $H^{(t)}$ is the set of all segments existing at time t .

Installation of MoMA. The installation of MoMA can be performed via Fiji^{43,44}. In Fiji, simply activate the MoMA update site. Once installed, the Fiji updater will automatically install future versions of MoMA containing new features and bug-fixes. The MoMA Wiki pages contain further information about how to install and use MoMA³⁰.

Implementation of MoMA. MoMA is implemented in Java, using the imaging library ImgLib2⁴⁵ and other components from the open source universe around ImageJ and Fiji^{43,44}. For solving ILPs we use Gurobi. The source code of MoMA is a Maven project, hosted on GitHub³⁰.

Additional features of MoMA. Additional useful features of MoMA include (i) the ability to optimize (solve) only parts of a loaded data set, (ii) save a fully or partially curated data set, and (iii) the possibility to export a found tracking solution for downstream processing.

If a loaded data set contains 1000 or even more time points, the optimization of MoMA's assignment model can take tens of seconds. While this is still fast, e.g. when compared to the data acquisition time for such a data set, leveraged editing can become cumbersome when the user is forced to wait tens of seconds between interactions for the optimization to finish. In order to guarantee fast interactive response times, MoMA allows users to define a subrange of time points $[t_a, t_b]$ across which to perform the optimization.

All assignments that are not in $[t_a, t_b]$ are either set to the value computed at a previous (partial) optimization run, or simply clamped to be 0. Formally this can be expressed by

$$\forall t \notin [t_a, \dots, t_b], \forall a^{(t)} \in A^{(t)} : a^{(t)} = \begin{cases} 1 & \text{if } a^{(t)} \text{ was set to 1 previously, or} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Once correct solutions are found, it is important that users can save and load the curations that they performed. Leveraged editing primitive introduces additional constraint to the underlying optimization problem, and MoMA is capable of serializing all edits to file.

But not only leveraged edits can be saved, also MoMA's segmentation and tracking results can be exported for subsequent downstream processing. An exhaustive list of exportable data is given below. MoMA's Wiki page contains a formal specification of the used data format³⁰.

- Data source.
- Total number of cells observed in the data set.
- Number of channels in the raw data, i.e. phase contrast and fluorescent channels.
- Growth channel (growth-channel) height and image height in pixels.
- (Vertical) position of the growth-channel in the image.
- For each cell, its cell id, and lineage information (the ids of its ancestors).
- For each time point in the life of each cell: position in the growth-channel [pixels and cell number]; bounding box area; intensity histogram, intensity percentiles, and pixel intensities for all channels.

Curation statistics. MoMA was tested on Mother Machine data with ~30 frames per cell cycle, stable focus over the experiment and both phase contrast and fluorescence imaged. To estimate the time the user needs to spend to curate data sets, we analyzed an unbiased selection of growth-channels and measured the time spend curating. For the selection of the growth-channels there was no visual inspection of the growth-channels other than checking that they harbor cells on the first frame. Therefore, this sample also harbored growth-channels in which the cells are lost during the experiment, and some that show structural defects. We only used the times for the growth-channels in which we had cells until the end of the experiment. Defect growth-channels were excluded as well. There are also rare growth-channels in which a cell is lysing or shows other abnormalities. In such cases, even with the eyes of an experienced observer, it is difficult to decide on the border of such cells, and such growth-channels were excluded as well.

Figure 1f shows a histogram of the fraction of frames needing curation across the growth-channels. Roughly half of the growth-channels required no curation at all, and most growth-channels require less than 1% of frames curated, with about 3% of frames needing curation in the worst case.

To give an impression of the amount of time that these curation statistics correspond to, in our hands, Fig. 5 shows the distribution of curation times per 100 frames across the growth-channels we analyzed. For each growth-channel, the total number of curated frames was extracted from the serialized file of curation interactions that MoMA saves. The inset of Fig. 5 shows that curation times are generally correlated with the fraction of frames that required curation.

All the curating with MoMA was performed on a MacBookPro (2.4 GHz Intel Core i7, 8 GB of memory). On this setup loading, initialization and the first round of optimization of a data set with 480 frames with two channels typically takes around 1 min. After curating the data the export step takes another 30 s.

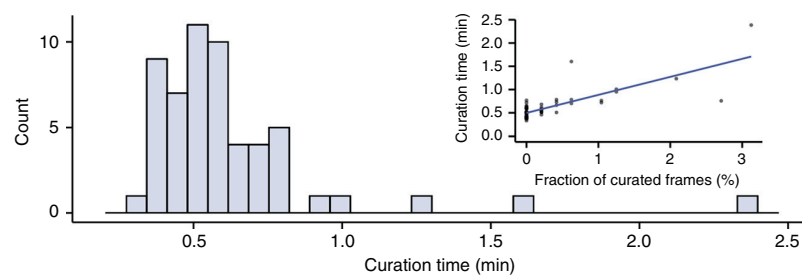


Fig. 5 Histogram of the curation times per 100 frames for a representative set of growth-channels. The inset shows a scatter plot of curation times (per 100 frames) as a function of the fraction of frames that were curated. The line shows a linear regression (least squares) fit

Cell size and growth rate estimation. From the imaging data we obtain, for each cell, pictures for each time point during its life-cycle. As an estimate of cell size, the software provides the dimension of the rectangular bounding-box within which the cell is contained. We have found that, both on our own data as on the data from other devices and microscopy setups, virtually all cells accurately follow simple exponential growth curves as a function of time, supporting the robustness of the estimation procedure. However, it is clear that the cell size estimation is quite coarse and we aimed to quantify the accuracy of these size estimates. This is difficult to do directly because we do not have independent measurements of cell sizes that can be used as a gold standard. However, if we find that the estimated cell size $s(t)$ accurately follows a simple exponential or linear form as a function of time t , then this suggests the errors in cell size are at most as large as the fluctuations of $s(t)$ away from the simple exponential or linear growth law.

Let $s(t)$ be the estimated size of the cell at time t and $x(t) = \log[s(t)]$. We used the data sets from the constant environments and used all cells which were monitored from birth to division, corresponding to 4016 cell cycles in glucose and 3387 cell cycles in lactose. For each cell cycle, we calculated the Pearson correlation between $x(t)$ and t across the cell cycle, as well as the Pearson correlation between $s(t)$ and t . Figure 6 shows the cumulative distributions of the squared Pearson correlation of the growth curves with exponential (black) and linear (orange) functions for cells grown in lactose (Fig. 6a) and in glucose (Fig. 6b).

We see that the growth curves are very well described by exponential functions of time, i.e. the median squared correlation coefficient is approximately 0.99 and almost all cells have correlation coefficients larger than 0.98. Correlation coefficients are substantially lower for fits to linear growth curves. Note that, whereas correlation coefficients are still very high for the linear growth fits, the log-likelihood for a growth-curve with squared correlation r^2 and T time points scales as $-T \log[1 - r^2]$. Thus, for a typical cell-cycle with $T = 30$ time points, the likelihood ratio between a fit with $r^2 = 0.99$ and one with $r^2 = 0.98$ is $\exp(20.8) \approx 10^9$. That is, the differences in the qualities of the linear and exponential fits are highly significant.

Since the elongation of cells is very well described by an exponential model, we can estimate the measurement error by studying the residuals of these fits. These residuals represent an upper bound on length measurement errors since they also include biological fluctuations around constant exponential growth. For each cell size observation in each cell cycle, we calculate the squared residual from the exponential fit, and obtained a squared relative error by dividing by the square of the estimated size. We then stratified the errors according to size and calculated, for each size class, the means and standard deviations of the squared relative errors. Taking the square-roots of these values we finally obtain the relative errors of the size measurements as a function of estimated size (Fig. 1g). We find that the measurement error on size is between 2 and 3%, and approximately independent of the length itself.

To estimate the average growth rate of an individual cell cycle we use linear regression of the log-sizes $x(t) = \log[s(t)]$ across the time points t in the cell-cycle, i.e. assuming all deviations from a perfect linear relationship $x(t) = a(t - t_0) + x_0$ are due to errors in the log-size estimates $x(t)$. Marginalizing over the cell-size x_0 at the time of birth t_0 , we find that the standard-deviation of the posterior distribution over growth-rate a is given by

$$\sigma(a) = \sqrt{\frac{\text{var}(x)(1 - r^2)}{(T - 1)\text{var}(t)}}, \quad (7)$$

where $\text{var}(x)$ and $\text{var}(t)$ are the variances of the log-sizes $x(t)$ and measurement times t , T is the number of measurements in the cell cycle, and r is the Pearson-correlation of the linear fit. The relative error on the estimated slope $a = \text{cov}(x, t) / \text{var}(t)$ is given by the ratio $\sigma(a)/a$.

Figure 6c shows the distribution of growth rates that we observe in constant glucose and lactose, and Fig. 6d shows the distribution of relative errors on growth rate. For the large majority of cell cycles, the error on the estimate of the growth rate is between 1 and 3%. The average growth rate is a bit higher in glucose (0.75 doublings per hour) than in lactose (0.69 doublings per hour). Notably, the variation in the growth rates of individual cell cycles is much larger than the measurement errors on these growth rates, indicating that growth rates vary

considerably across single cells. We find that growth rates vary by about 17% in both glucose and lactose (i.e. one standard deviation), and we observe cell cycles that differ by more than twofold in their growth rates.

We also investigated whether growth rates during the switching conditions vary systematically from growth rates in the corresponding constant conditions. Figure 6 shows the distribution of growth rates for individual cell cycle separately for the first, second, and third time segment in both glucose (Fig. 6e) and lactose (Fig. 6f) during the switching conditions.

We see that the growth rate distributions during individual time segments in the switching experiments are very similar to the distributions in the corresponding constant conditions. The only exception is the slightly higher growth rates in the first time segment in glucose during the switching conditions. Although we have not investigated the origin of the slightly higher growth rates in this time segment in detail, we believe that it results from a combination of two effects. First, we note that the growth rates in glucose are slightly higher in all three segments during the switching conditions than in the constant conditions. This suggests that a subtle change in the conditions on the day of the experiment may have caused slightly increased growth rates during the switching conditions. Second, when fluorescence measurements are taken, the light from the illumination causes some small stress to the cells, which is reflected in slightly lower growth rates compared to conditions where no fluorescence measurements are taken. As a consequence, we observe that cells slightly lower their growth rates during the first hours of the experiment. To correct for this systematic effect we only start recording measurements in each experiment, after 2 h in conditions with illumination. We hypothesize that during the first glucose segment in the switching experiments, the cells had not yet fully adapted to the illumination conditions.

Cell fluorescence estimation. To estimate the GFP content of each cell, we post-process the fluorescence data as follows. The raw data consist of fluorescence intensities for all pixels within the segment of the picture containing the cell. This segment is 100 pixels wide, with the growth-channel covering approximately 13 pixels in the center of the picture. We first obtain column-sums c_i by summing the pixel intensities of all pixels in each of the 100 columns i . Note that we assume that these column sums are dominated by the fluorescence coming from the cell in question, i.e. that the fluorescence coming from neighboring cells above and below the cell are negligible. We find that this is a good approximation when cells in a given growth-channel all have similar fluorescences but note that, in conditions where neighboring cells may have fluorescences that differ by orders of magnitude, this assumption may break down. Figure 7 shows the profiles of column sums c_i for a cell at three time points of its cell cycle while growing in lactose (top three panels) and for a cell growing in glucose (bottom three panels). From prior biological knowledge, we know that GFP is highly expressed during the growth on lactose, and that it is very lowly expressed during the growth on glucose.

Remarkably, the growth-channel (central 13 pixel positions in the figures) is not detectable at all in the fluorescence curves. Instead, the fluorescence signal shows a long-tailed peak centered in the middle of the growth-channel, extending far beyond the width of the growth-channel, and reaching a minimum at positions halfway between neighboring growth-channels, i.e. near the left and right ends of the profiles in Fig. 7. As the cell grows, i.e. from the leftmost to rightmost panel, the length of the segment grows and the column-sums grow proportionally to the segment length. Notably, the minimal fluorescence level is almost twice as high when growing in lactose compared to when growing in glucose. We conclude from these observations that the fluorescence from each cell spreads over significant distances across the image and that this also causes background levels to depend on the overall expression levels in neighboring growth-channels. Therefore, to properly estimate the amount of fluorescence emerging from the cell we need to fit the background intensity within each segment and we need a mathematical model for the long-tailed shape of the peak.

We found that the shape of the peak is very well described by a Cauchy (or Lorentzian) distribution, giving an overall form of the fluorescence profile:

$$c_i = \text{noise} + B + \frac{A}{1 + \left(\frac{i - i_{\text{mid}}}{w}\right)^2}, \quad (8)$$

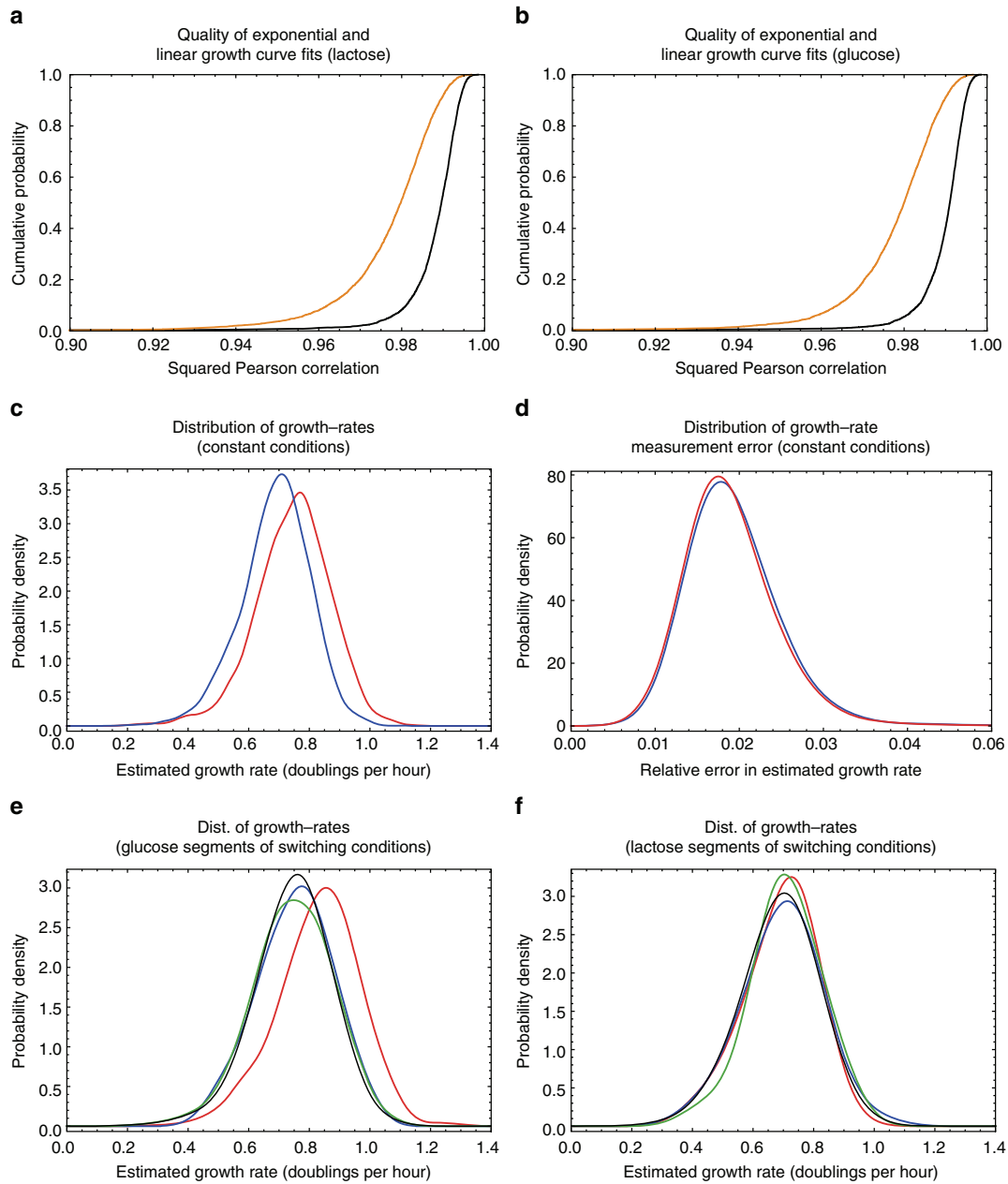


Fig. 6 Exponential growth curves and the distribution of growth rates. **a** Cumulative distributions of the squared Pearson correlations between the estimated sizes of the cells and time (orange), or estimated log-sizes and time (black) for the cells grown in lactose. **b** As in panel **a** but for cells grown in glucose. **c** Distribution of growth rates of individual cell cycles in glucose (blue) and lactose (red). **d** Distribution of relative errors of the growth rate estimates in glucose (blue) and lactose (red). **e** Distribution of growth rates of individual cell cycles during the first (red), second (blue), and third (green) time segments in glucose of the experiments with switching conditions. For comparison, the black curve shows the growth rate distribution in constant glucose. **f** As in panel **e** but for the time segments in lactose

where i is the horizontal position, i_{mid} is the center of the peak, w its width, A the amplitude of the signal, 'noise' is the measurement noise, and B the background fluorescence. Assuming that the measurement noise is Gaussian distributed, it is straight forward to fit this model using expectation maximization. We find that, systematically, the center $i_{\text{mid}} \approx 50$ –52, and the width $w \approx 5$ –6 pixels. We interpret the amplitude A to be proportional to the total number of GFP molecules in the cell, and the background B to correspond to the combined effects of the camera offset, the auto-fluorescence of the microfluidic chip and the media, and stray fluorescence from neighboring cells. The expectation maximization procedure for fitting the fluorescence profile is

1. Find the maximum and minimal fluorescence column-sums c_{max} and c_{min} across the profile.
2. Initialize B to c_{min} , A to $c_{\text{max}} - c_{\text{min}}$, w to 5.5 and i_{mid} to 50, i.e. in the middle of the profile.
3. Calculate a theoretical profile:

$$\rho_i = \left[1 + \left(\frac{i - i_{\text{mid}}}{w} \right)^2 \right]^{-1}, \quad (9)$$

and its integral $\rho = \sum_{i=1}^N \rho_i$.

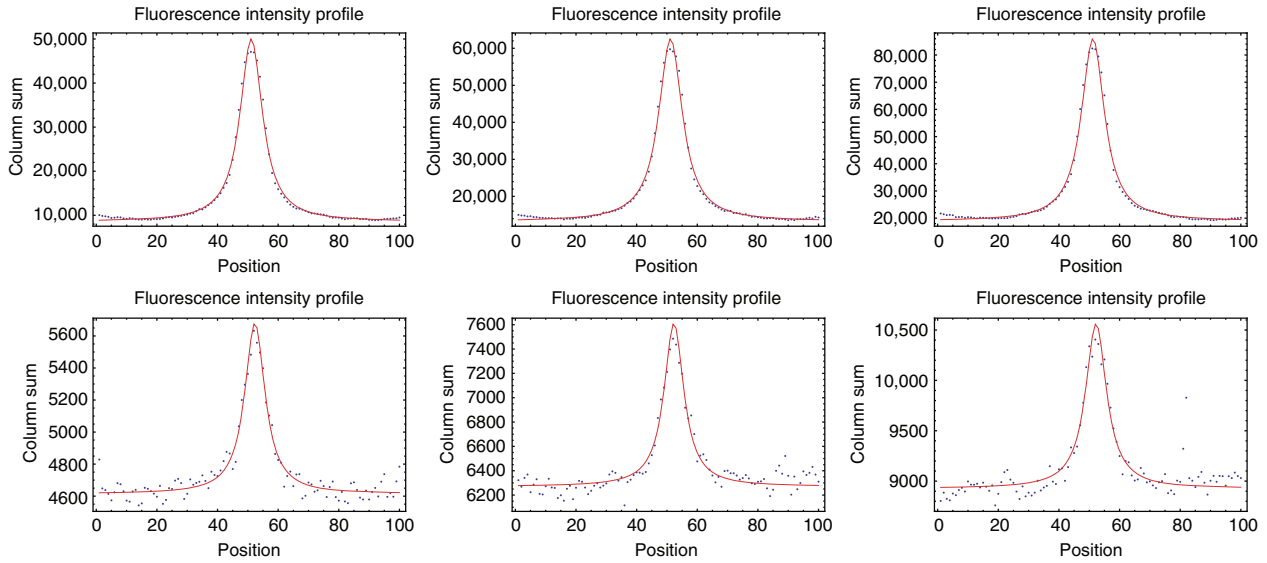


Fig. 7 Examples of the horizontal profiles of column sums c_i of fluorescence intensities (blue dots) for three cells during growth on lactose (top panels) and during growth on glucose (bottom panels). The three panels show, from left to right, the profiles at the start, the middle, and at the end of the cell cycle of two example cells. The red curves show the fits to the mixture model of a fixed background plus a Cauchy-distributed signal

4. Set a new value of the background B' :

$$B' = \frac{1}{N} \sum_{i=1}^N \frac{Bc_i}{B + A\rho_i}. \quad (10)$$

5. Set a new value for the amplitude A' :

$$A' = \frac{1}{\rho} \sum_{i=1}^N \frac{Ac_i\rho_i}{B + A\rho_i}. \quad (11)$$

6. Using the updated values A' and B' , calculate an updated profile ρ_i and optimize i_{mid} by finding the zero of the derivative:

$$\sum_{i=1}^N (i - i_{\text{mid}}) \rho_i^2 \left(-1 + \frac{c_i}{B' + A'\rho_i} \right). \quad (12)$$

7. Using the updated values A' , B' , and i_{mid} , optimize w by finding the zero of the derivative:

$$\sum_{i=1}^N \rho_i (1 - \rho_i) \left(-1 + \frac{c_i}{B' + A'\rho_i} \right). \quad (13)$$

The accuracy of this method to estimate the total fluorescence of the cell can be quantified by taking advantage of the precise environment control allowed by our new setup, as discussed in the next section. We distribute a post-processing script with the MoMA code that allows users to apply this fluorescence amplitude estimation to exported output files from MoMA.

Cell auto-fluorescence estimation. In addition to the background fluorescence of the PDMS and stray fluorescence from nearby cells that are corrected for by the methods described in the previous section, there is background fluorescence coming from the auto-fluorescence of the cells and media. To estimate this auto-fluorescence, we measured the wild-type strain of *E. coli* MG1655, i.e. without the fluorescent reporter, in the conditions where we switch between glucose and lactose. We observed that the estimated total fluorescence, i.e. the amplitude A from the previous section, correlates well with the size of the cells during their cell cycle. That is, fitting a linear relationship $A = aS + b$ of the fluorescence A as a function of the estimated cell size S typically yields Pearson correlation coefficients of $r \approx 0.9$. Moreover, we observed that the vast majority of fits were consistent with $b = 0$, i.e. the total fluorescence being directly proportional to cell size, supporting that this signal corresponds to the auto-fluorescence of the cell. Note that any uniform fluorescence coming from the growth medium would also be accounted for by this procedure (in the parameter a).

To fit the auto-fluorescence a (per micrometre of cell length) we selected all cells that were observed for a full cell cycle, who never got within 100 pixels of the end of the growth-channel during their cell cycle, and whose length as a function of time was well fit by a simple exponential growth curve ($r^2 \geq 0.99$). This latter restriction mainly serves to remove cells that had a transient stop in growth after the first switch to lactose. In total there were 284 cells that passed all these criteria. For each of these cells we replaced the directly estimated sizes S_t at each time point t , with the sizes \tilde{S}_t estimated from the exponential fit of S_t as a function of time (reasoning that these estimates are more accurate than the direct measurements). For each cell we then fit a function $A_t = a\tilde{S}_t$, assuming Gaussian measurement noise of unknown variance.

That is, for a single cell we write

$$P(D|a, \sigma) \propto \sigma^{-T} \exp \left[-\sum_t \frac{(A_t - a\tilde{S}_t)^2}{2\sigma^2} \right]. \quad (14)$$

Using a scale prior on σ of the form $P(\sigma) \propto 1/\sigma$, and integrating over σ we obtain

$$P(a|D) \propto \left[\langle \tilde{S}^2 \rangle \left(a - \frac{\langle A\tilde{S} \rangle}{\langle \tilde{S}^2 \rangle} \right)^2 + \langle A^2 \rangle - \frac{\langle A\tilde{S} \rangle^2}{\langle \tilde{S}^2 \rangle} \right]^{-T/2}, \quad (15)$$

where T is the number of time points in the cell cycle and the averages are over the time points in the cell cycle.

The optimal value of a is given by

$$a_* = \frac{\langle A\tilde{S} \rangle}{\langle \tilde{S}^2 \rangle}. \quad (16)$$

Approximating the posterior by a Gaussian we obtain for the standard deviation of the estimated a

$$\sigma_a = \sqrt{\frac{1}{T} \left[\frac{\langle A^2 \rangle}{\langle \tilde{S}^2 \rangle} - a_*^2 \right]}. \quad (17)$$

Figure 8a shows the estimated value a_* and its error bar σ_a for each of the 284 cells. Note that, although most cells have fluorescence values between 400 and 500 per μm , there are some outliers at higher fluorescence. This is also evident from the combined probability density of a values (Fig. 8b).

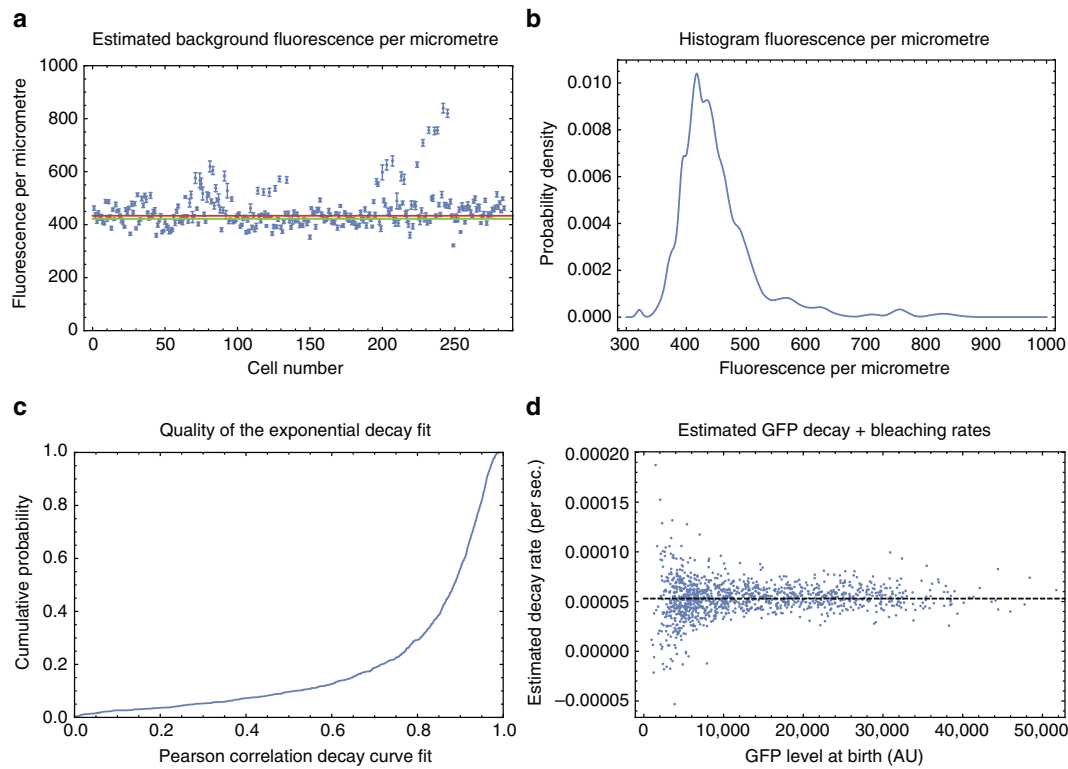


Fig. 8 Estimating auto-fluorescence and fluorescence decay. **a** Estimated auto-fluorescence per micrometre cell length a , and its error bar σ_a for 284 cells for which fluorescence was fit as a function of cell size. The red line ($a = 433.5$) is the fit obtained when all cells are assumed to have a common fluorescence per micrometre a . The green line is obtained with a mixture model that allows for “outliers” from a uniform distribution ($a = 421.8$). **b** The joint probability density of a given by the mixture of Gaussian distributions for all 284 cells. **c** GFP decay (including bleaching) was estimated by fitting observations of decreasing GFP levels during the glucose phases in the switching experiments, when no GFP is synthesized, to an exponential function of time. The panel shows the cumulative distribution of the Pearson correlations between the estimated log-GFP levels and time. **d** Estimated decay rates for individual cells (vertical axis) as a function of absolute GFP level of the cell (horizontal axis). The dashed line shows the overall estimated rate used in subsequent analysis

If we assume there is one common background fluorescence per micrometre α for all cells, then the probability of the data given α is given by

$$P(D|\alpha) = \prod_c \frac{1}{\sigma_a(c)} \exp \left[-\frac{(a_s(c) - \alpha)^2}{2\sigma_a(c)^2} \right], \quad (18)$$

where the product is over the 284 cells c .

Maximizing this function with respect to α yields

$$\alpha = \sum_c \frac{a_s(c)}{\sigma_a(c)^2} \left[\sum_c \frac{1}{\sigma_a(c)^2} \right]^{-1} = 433.5. \quad (19)$$

If we allow that there are some ‘outlier’ cells whose value of a is described by a uniform distribution of width $W = a_{\max} - a_{\min}$, then the likelihood of the data as a function of α and the fraction of non-outlier measurements ρ is given by

$$P(D|\alpha, \rho) = \prod_c \left[\frac{1 - \rho}{W} + \frac{\rho \exp \left(-\frac{(\alpha - a_s(c))^2}{2\sigma_a(c)^2} \right)}{\sqrt{2\pi}\sigma_a(c)} \right]. \quad (20)$$

Maximizing this function with respect to α and ρ yields $\alpha = 421.8$ and $\rho = 0.31$. In the following we will use this latter value of α for the auto-fluorescence per micrometre of cell length. For each cell with estimated size S and total fluorescence A , we thus obtain an auto-fluorescence corrected fluorescence level $\bar{A} = A - \alpha S$.

Estimating GFP’s bleaching and degradation. As shown in Fig. 2a, while the *lac* operon is induced in the lactose phases, GFP production ceases during the glucose phases. In this regime, the total cell fluorescence slowly decreases during the cell cycle, and approximately divides in half at each cell division. We reasoned that the slow continuous decay of fluorescence during the cell cycles is the result of GFP bleaching and, potentially, also some GFP degradation. Inspection of the data indeed shows that the total fluorescence decrease is captured well by an exponential

model. For this analysis, we consider only observations between 30 min after the switch to glucose and before the next switch to lactose, and to cells with at least 10 points in this time window. This corresponded to 33,052 independent cell observations over 1220 cells.

As shown in Fig. 8c, the GFP degradation across time is well fit by an exponential model for most cells. Assuming that a cell undergoes bleaching+GFP degradation at a rate μ per second, we estimated μ for each cell from a linear regression of $\log(\text{GFP level})$ against time (Fig. 8d). Combining information from the estimates of individual rates for each cell and their standard deviations, we estimate the overall rate μ_{tot} to be equal to $5.3 \times 10^{-5} \pm 5 \times 10^{-7}$ (mean \pm s.d.) per second. Note that this corresponds to a loss of about 18% of the GFP signal per hour due to bleaching and GFP decay.

Accuracy of the fluorescence estimation. We also took advantage of our unique ability to study the growth regime where no GFP is produced to quantify the measurement errors on the total GFP. Since, in the glucose phases of the switching experiments, the GFP dynamics is dominated by bleaching and degradation, and well described by an exponential decay model, we computed the squared residuals (normalized by the squared value) from the independent fits of $\log(\text{GFP})$ as a function of time for each cell. As for the analysis of measurement errors on length, residuals are stratified into bins based on total GFP, and the means and standard errors of the normalized squared residuals (i.e. relative to total GFP) are computed for each bin (Fig. 2g). We find that the squared relative error on the GFP measurement scales inversely with the total GFP level (i.e. a power-law fit has exponent 1.01), which indicates that, as in shot noise, the squared error is inversely proportional to total GFP level. In practice, the absolute error is around 20 molecules when the cell has 200 GFP molecules (i.e. 10%), and around 80 molecules when the total is 4000 GFP molecules (i.e. 2%).

Estimating the fluorescence per GFP molecule. To estimate the conversion factor between the background-corrected total fluorescence \bar{A} and the number of GFP molecules, we will use data on the fluctuations in fluorescence levels of newborn sibling pairs. To avoid confounding effects from GFP production, we collected division events from the glucose phases in our switching experiments,

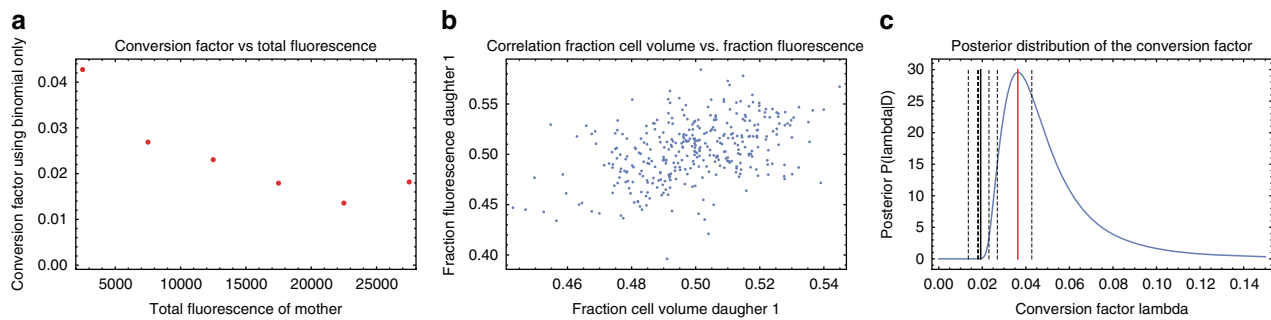


Fig. 9 Estimating the conversion factor between fluorescence intensity and number of GFP molecules. **a** Estimated conversion factor λ , using the “naive” method which assumes there are only binomial fluctuations, as a function of total fluorescence. Division events were divided into bins based on the mother’s total fluorescence. Bin size was 5000. **b** Correlation between fluctuations in fluorescence and cytoplasm size of two sibling cells immediately after birth. Each dot corresponds to a pair of sister cells with the horizontal axes showing the fraction of the total cytoplasm of the mother and the vertical axis showing the fraction of the total fluorescence of the mother taken up by the first sister. **c** Posterior distribution $P(\lambda|D)$ of the conversion factor λ given the data on our sibling pairs (blue curve). The red line shows the maximum likelihood value $\lambda_* = 0.0361$. The black lines show the estimated conversion factors that are obtained assuming binomial noise only. The solid line results from using all data, and the dashed lines result from different subsets at different absolute fluorescence values as in the left panel

when GFP production has ceased. Collecting division events from these phases has the added advantage that absolute GFP levels vary over a considerable range across cells during these phases, allowing us to quantify the size of fluctuations in sibling fluorescence as a function of total fluorescence. Our observations consist of fluorescence levels at birth (x_i, y_i) for sibling pairs of daughters, where i runs from 1 to N , with N the total number of such sibling pairs. Using the same criteria as in the decay analysis for mothers and daughters, we collected $N = 357$ sibling pairs. GFP levels at birth were estimated in each daughter cell as average of the levels at all time points corrected for the previously estimated decay. Assuming that the GFP molecules in the mother cell are distributed randomly between the daughters, the fluctuations in the numbers of GFP molecules going to each daughter should be binomially distributed, and this has been used previously to infer a conversion factor between GFP molecule numbers and fluorescence levels²⁴. In particular, assuming binomial fluctuations, the expectation of the square of the difference $\langle (n_i - m_i)^2 \rangle$ should be equal to the total count $n_i + m_i$. Given a conversion factor λ , such that the GFP molecule counts correspond to $(n_i, m_i) = \lambda(x_i, y_i)$, one can estimate λ by observing

$$1 = \left\langle \frac{(n_i - m_i)^2}{n_i + m_i} \right\rangle = \lambda \left\langle \frac{(x_i - y_i)^2}{x_i + y_i} \right\rangle. \quad (21)$$

However, using this “naive” approach, we find that the conversion factor λ systematically decreases with total fluorescence (Fig. 9a), changing by as much as fourfold depending on whether division events with low or high absolute fluorescence are used. This result implies that the variance of fluorescence fluctuations grows faster than linear with total fluorescence, suggesting that there are additional fluctuations with variance proportional to total fluorescence squared. Inspection of the data strongly suggests that these additional fluctuations derive from fluctuations in the cell size of the daughters. That is, in addition to the binomial fluctuations there are fluctuations caused by the daughters having unequal size. Practically, cell size at birth is estimated in each daughter cell by extrapolating the linear fit of $\log(\text{length})$ as a function of time. Indeed, we observe a substantial correlation between the relative sizes of the siblings and the relative amounts of fluorescence each sibling receives (Fig. 9b, Pearson correlation $r = 0.44$).

We thus developed a more sophisticated model, which takes into account fluctuations in the cell sizes, the binomial fluctuations, as well as measurement noise. For a given division event i , let ρ_i denote the measured fraction of the cytoplasm that went to the first daughter, and let $q_i = x_i/(x_i + y_i)$ be the measured fraction of the fluorescence that went to the first daughter. We will assume that q_i is a noisy measurement of the true fraction of molecules $q = n_i/(n_i + m_i)$ that went to the first daughter, and that ρ_i is a noisy measurement of the true fraction of the mother’s cytoplasm ρ that went to the first daughter. Given ρ and a total number of molecules $n = (n_i + m_i)$, the molecule numbers (n_i, m_i) will show binomial fluctuations and the fraction q will have a variance $\text{var}(q) = \rho(1-\rho)/n$. In addition to this variance we will assume there is a total measurement noise of variance v , so that the total expected square-deviation between the measurements q_i and ρ_i should be $v + \rho(1-\rho)/n$. We will assume that the sum of these fluctuations due to the binomial noise and measurement noise is approximately Gaussian distributed. Finally, we will assume that the binomial variance $\rho(1-\rho)/n$ is well approximated by the measured values $\rho_i(1-\rho_i)/(\lambda(x_i + y_i))$.

Under this model, the probability of observing the fraction q_i given the measured volume fraction ρ_i , the conversion factor λ , and the total measurement

noise v is given by

$$P(q_i|\rho_i, \lambda, v) = \left(v + \frac{\rho_i(1-\rho_i)}{\lambda(x_i + y_i)} \right)^{-1/2} \exp \left[-\frac{(q_i - \rho_i)^2}{2 \left(v + \frac{\rho_i(1-\rho_i)}{\lambda(x_i + y_i)} \right)} \right]. \quad (22)$$

The log-likelihood of λ and v is now given by a sum over the N division events:

$$L(\lambda, v) = -\frac{1}{2} \sum_{i=1}^N \frac{(q_i - \rho_i)^2}{\left(v + \frac{\rho_i(1-\rho_i)}{\lambda(x_i + y_i)} \right)} + \log \left[v + \frac{\rho_i(1-\rho_i)}{\lambda(x_i + y_i)} \right]. \quad (23)$$

To obtain the posterior probability of λ we marginalize over the unknown variance v (using a uniform prior). That is, we calculate $L(\lambda) = \log \int \exp[L(\lambda, v)] dv$, performing the integral numerically. Using this model, the maximal likelihood value of λ is given by

$$\lambda_* = 0.0361, \quad (24)$$

and the symmetric 95% posterior probability interval is given by $\lambda \in [0.026, 0.112]$.

Figure 9c shows the posterior distribution $P(\lambda|D)$ obtained with our model. For comparison, Fig. 9c also shows the conversion factors that would be obtained with the naive method that assumes there is only binomial noise, i.e. using all data the number of molecules would be underestimated by almost twofold.

Data availability.

- The designs of the DIMM device, as well as a handbook with detailed experimental methods, are available from Metafluidics web repository at <https://metafluidics.org/devices/dual-input-mother-machine/>.
- The MoMA software and source code is available on Github: <https://github.com/fjug/MoMA>. For end users, MoMA is also available as a Fiji plugin at <http://sites.imagej.net/MoMA>.
- Extensive documentation is provided as a Wiki page containing information about MoMA’s installation and use, as well as tutorial videos: <https://github.com/fjug/MoMA/wiki>.
- Raw image data of the analyzed growth-channels as well as processed data (estimated cell sizes and fluorescence levels) for all experiments presented in the paper are available from Zenodo at <https://doi.org/10.5281/zenodo.746230>. A README file is provided with detailed explanation as to which file corresponds to which experiment, and the file format of the processed data files.
- A movie from a time lapse experiment in which *E. coli* ASC622 cells grow in the DIMM under conditions that switch (every 4 h) between glucose and lactose as a carbon source is available on Youtube: <https://www.youtube.com/watch?v=2Tzm868fmc>. This movie is also available as Supplementary Movie 1.

Received: 24 October 2017 Accepted: 6 December 2017

Published online: 15 January 2018

References

- Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
- Taniguchi, Y. et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
- Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
- Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008).
- Kiviet, D. J. et al. Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).
- Ducet, A. et al. A microscope automated fluidic system to study bacterial processes in real time. *PLoS ONE* **4**, e7282 (2009).
- Robert, L. et al. Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch. *Mol. Syst. Biol.* **6**, 357 (2010).
- Boulineau, S. et al. Single-cell dynamics reveals sustained growth during diauxic shifts. *PLoS ONE* **8**, e61686 (2013).
- Wang, P. et al. Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
- Ullman, G. et al. High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **368**, 20120025 (2012).
- Lambert, G. & Kussell, E. Memory and fitness optimization of bacteria under fluctuating environments. *PLoS Genet.* **10**, e1004556 (2014).
- Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503**, 481–486 (2013).
- Long, Z. et al. Measuring bacterial adaptation dynamics at the single-cell level using a microfluidic chemostat and time-lapse fluorescence microscopy. *Analyst* **139**, 5254–5262 (2014).
- Slusarenko, O., Heinritz, J., Emonet, T. & Jacobs-Wagner, C. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.* **80**, 612–627 (2011).
- Young, J. W. et al. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88 (2012).
- Paintdakhi, A. et al. Outfit: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.* **99**, 767–777 (2016).
- Tanouchi, Y. et al. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* **523**, 357–360 (2015).
- Taheri-Araghi, S. et al. Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391 (2015).
- Hashimoto, M. et al. Noise-driven growth rate gain in clonal cellular populations. *Proc. Natl. Acad. Sci. USA* **113**, 3251–3256 (2016).
- Ferry, M. S., Razinkov, I. A. & Hasty, J. in *Methods in Enzymology*, Vol. 497 of Synthetic Biology, Part A (ed. Voigt, C.) 295–372 (Academic Press, Cambridge, MA, 2011).
- Jug, F. et al. in *Bayesian and Graphical Models for Biomedical Imaging* (eds Jorge Cardoso, M., Simpson, I., Arbel, T., Precup, D. & Ribbens, A.) (Springer, Cambridge, MA, 2014).
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
- Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA* **43**, 553–566 (1957).
- Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004).
- Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008).
- Bhogale, P. M., Sorg, R. A., Veening, J. W. & Berg, J. What makes the lac-pathway switch: identifying the fluctuations that trigger phenotype switching in gene regulatory systems. *Nucleic Acids Res.* **42**, 11321–11328 (2014).
- Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
- Jug, F. Moma—the mothermachine analyzer. <https://github.com/fjug/MoMA> (2016).
- Metafluidics. Open repository for fluidic systems. <https://metafluidics.org/devices/dual-input-mother-machine/> (2017).
- Stroock, A. D. et al. Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).
- Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μ Manager. *Curr. Protoc. Mol. Biol.* **Unit 14.20**, 1–17 (2010).
- Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
- Bar-Shalom, Y. *Multitarget-Multisensor Tracking: Advanced Applications* Vol. 391 (Artech House, Norwood, MA, 1990).
- Chenouard, N., Bloch, I. & Olivo-Marin, J.-C. Multiple hypothesis tracking for cluttered biological image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 2736–2750 (2013).
- Jiang, H., Fels, S. & Little, J. J. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1–8 <https://doi.org/10.1109/cvpr.2007.383180.24> (2007).
- Kausler, B. X. et al. in *ECCV'12: Proceedings of the 12th European Conference on Computer Vision* (eds Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y. & Schmid, C.) 144–157 (Springer-Verlag, Berlin, Heidelberg, 2012).
- Funke, J., Anders, B., Hamprecht, F. A., Cardona, A. & Cook, M. in *CVPR '12: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1004–1011 (IEEE Computer Society, Washington, DC, 2012).
- Schiegg, M., Hanslovsky, P., Kausler, B. X. & Hufnagel, L. in *Proceedings of the IEEE Conference on Computer Vision ICCV 2013*, 2928–2935 (IEEE, Sydney, 2013).
- Padfield, D., Rittscher, J. & Roysam, B. in *IPMI '09: Proceedings of the 21st International Conference on Information Processing in Medical Imaging* (Springer-Verlag, Berlin, Heidelberg, 2009).
- Schrijver, A. *Theory of Linear and Integer Programming* (John Wiley & Sons, New York, NY, 1998).
- Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
- Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
- Pietzsch, T., Preibisch, S., Tomancak, P. & Saalfeld, S. ImgLib2—generic image processing in Java. *Bioinformatics* **28**, 3009–3011 (2012).
- Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).

Acknowledgements

We thank Benjamin Sellner, Urs Jenal, Christian Schwall, James Locke, and Lydia Robert for sharing their raw data with us, to allow us to demonstrate MoMA's general applicability on data sets from different labs and setups.

Author contributions

O.S., G.M., and E.v.N. designed the research. M.K., O.S., S.D., and T.P. designed and prototyped the DIMM device. M.K. and T.J. designed and performed the experiments. F. J. designed and wrote the MoMA software. T.J. and E.v.N. developed and applied data processing methods. M.K., T.J., and E.v.N. analyzed the data. M.K., F.J., T.J., O.S., G.M., and E.v.N. wrote the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-017-02505-0>.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Chapter 4

Summary and Future Prospects

In the previous chapters the development and application of a microfluidic device called the Dual Input Mother Machine for studying single-cell transcription regulatory dynamics were presented. The Dual Input Mother Machine for the first time combines the potential of long-term observations of growth and gene expression phenotypes of single bacterial cells with precise automated environmental control.

While another study used a Y-junction to manually switch between inputs in the Mother Machine [28], the environmental control in the Dual Input Mother Machine is completely automated using programmable syringe pumps. In addition the Dual Input Mother Machine allows more precise timing of the switches because the DAW-junction is integrated into the device, which reduces the volume that has to flow through the device after the junction before the switch is effective. Like the Dual Input Mother Machine, also the chemoflux microfluidic device presented by Lambert and Kussel 2014 allows growing bacteria over long time-scales, while the environment can be changed by switching between two media with a junction integrated in the device [61]. But unlike

the Dual Input Mother Machine, the chemoflux does not allow for mixing the two media. An additional difference between the devices is that in the chemoflux device, the cells are growing in dead end channels (like in the Mother Machine) which are several micrometers wide. Because the cells that are growing in such channels can move between each other, it is harder to follow them and there is no guarantee that the same cells always stay at the bottom of the growth channel where they could be monitored over their whole lifetime. Therefore, with its ability to both switch and mix inputs while following single cells essentially over their whole lifetime, the Dual Input Mother Machine is well suited to address a wide range of biological questions and we are convinced that this setup will be an important new tool to get to a better understanding of the mechanisms of gene regulation in single cells in the future.

As a first example this setup was used to study the regulation of the expression of the *lac* operon to understand variability in single cells and how differential regulation correlates with the fitness of single cells in an environment in which the carbon source alternates between glucose and lactose every 4h. The results of this study, presented in Chapter 3, demonstrate nicely how genetically identical cells that are confronted with the same environmental changes can behave very differently because of the stochastic processes involved in transcriptional regulation. After the first switch from glucose to lactose, the time until the induction of the expression of the *lac* operon was measured and found to be highly variable between single cells. When glucose is present the expression of the *lac* operon is repressed and only sporadically is there a burst of expression, probably when LacI stochas-

tically dissociates from the promoter for a short time [18]. Likely the differences observed in the waiting times until the induction can be explained by cells ending up with varying amounts of LacY and LacZ because of the stochastic expression bursts in the repressed state before the switch. The cells that happen to have little or no LacY at the switch will import lactose very slowly, or fail to import lactose completely. In addition, the lactose that gets into such cells is not converted very efficiently into allolactose, because of their low LacZ levels. After 4h in lactose the cells were switched to glucose again and expression of the *lac* operon was observed to stop in glucose. Therefore the LacZ, LacY and LacA molecules present in the cell at this time were observed to be diluted out by division in the glucose phase. When the cells were switched from glucose to lactose for the second time, the waiting times until induction of the expression of the *lac* operon were found to be much shorter and less variable compared to the first glucose to lactose switch. This finding indicates that the levels of the Lac proteins were not completely diluted back to the level observed before the first induction of the system and the remaining molecules provided a phenotypic memory that facilitated induction of the expression in the second switch to lactose. Similar results were presented before by Lambert and Kussel 2014 [61]. They grew their cells in the chemoflux device mentioned above and also observed the lag phase after the first switch to lactose, which was gone in the second switch to lactose with a 4h growth phase in glucose in between. By changing the time grown in glucose between the switches to lactose they determined the time-scale over which the memory persists and demonstrated that overexpression of *lacZ*, and to a lesser extent of *lacY*,

can maintain the cells in an induced state. In addition they present a second memory mechanism and dissect the determinants of the lag phases that are observed after switches from glucose to lactose [61]. While the experiments in the study of Lambert and Kussel 2014 [61] lack single cell resolution, the Dual Input Mother Machine not only allows measurement of how the times until induction change for different switches but also allows quantification of the heterogeneity between cells at each switch, which is very important to understand possible functional consequences of stochasticity in gene regulation, e.g. bet hedging strategies. Therefore the experimental setup presented here opens new perspectives to study various regulatory switches on the single cell level and to understand aspects that are hidden in experiments without single cell resolution.

While only one experimental example of the use of the dual input Mother Machine was presented here, the applications are manifold. Based on the results observed in Chapter 3 it would be interesting to repeat some of the experiments done in Lambert and Kussel 2014 [61] and use the setup to study how long the phenotypic memory can last and whether it is possible to evolve the duration of the memory in an experimental evolution setup. The Dual Input Mother Machine could be used to characterize the strains coming out of such selection experiments and to thereby link genetic changes to changes in the regulation of the expression of the *lac* operon in single cells. Another interesting approach to learn more about how genetic changes can lead to changes in the regulation of a system like the *lac* operon would be to look at the behaviour of natural isolates of *E. coli* (for which we have the genome sequences) in experiments where the expression

of the *lac* operon is measured while the carbon source is alternated between glucose and lactose. The observed differences in the regulation of the expression of the *lac* operon in single cells could again be linked to genetic differences between the strains. One benefit of the dual Mother Machine is also that one can easily study the correlation between the gene expression dynamics and the growth phenotype of single cells in changing environments. Therefore it would be interesting to label the natural isolate strains that behave differently at the first switch from glucose to lactose and load them together in the dual input Mother Machine to check their fitness in experiments that differ in the duration of the glucose and lactose phases, or in the sugar concentrations, to learn about how selection can act on the noise in gene regulation.

References

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318–356 (1961).
2. Stevens, A. Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B. *Biochemical and Biophysical Research Communications* **3**, 92–96 (1960).
3. Ippen, K., Miller, J. H., Scaife, J. & Beckwith, J. New controlling element in the *lac* operon of *E. coli*. *Nature* **217**, 825–827 (1968).
4. Eron, L. & Block, R. Mechanism of initiation and repression of in vitro transcription of the *lac* operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 1828–1832 (1971).
5. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
6. Burgess, R. R., Travers, A. A., Dunn John J., J. J. & Bautz Ekkehard K.F., E. K. F. Factor stimulating transcription by RNA polymerase. *Nature* **221**, 43–46 (1969).
7. Travers, A. A. & Burgess, R. R. Cyclic re-use of the RNA polymerase sigma factor. *Nature* **222**, 537–540 (1969).
8. Dunn, J. J. & Bautz, E. K. DNA-dependent RNA polymerase from *E. coli*: Studies on the role of σ in chain initiation. *Biochemical and Biophysical Research Communications* **36**, 925–930 (1969).
9. Maeda, H., Fujita, N. & Ishihama, A. Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research* **28**, 3497–3503 (2000).
10. Hirschman, J., Wong, P. K., Sei, K., Keener, J. & Kustu, S. Products of nitrogen regulatory genes *ntr* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product

- is a sigma factor. *Proceedings of the National Academy of Sciences* **82**, 7525–7529 (1985).
11. Hunt, T. P. & Magasanik, B. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. *Proceedings of the National Academy of Sciences* **82**, 8453–8457 (1985).
 12. Paget, M. S. Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules* **5**, 1245–1265 (2015).
 13. Englesberg, E., Irr, J., Power, J. & Lee, N. Positive control of enzyme synthesis by gene C in the L-arabinose system. *Journal of bacteriology* **90**, 946–957 (1965).
 14. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 57–65 (2004).
 15. McAdams, H. & Arkin, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 814–819 (1997).
 16. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
 17. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **31**, 69–73 (2002).
 18. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
 19. Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics* **40**, 471–475 (2008).
 20. Rotem, E. *et al.* Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences* **107**, 12541–12546 (2010).
 21. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology* **13**, 497–508 (2015).
 22. Young, J. W. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols* **7**, 80–88 (2012).

23. Ducret, A. *et al.* A microscope automated fluidic system to study bacterial processes in real time. *PLoS ONE* **4**, e7282 (2009).
24. Robert, L. *et al.* Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch. *Molecular Systems Biology* **6**, 357 (2010).
25. Boulineau, S. *et al.* Single-cell dynamics reveals sustained growth during diauxic shifts. *PLoS ONE* **8**, e61686 (2013).
26. Wang, P. *et al.* Robust growth of *Escherichia coli*. *Current Biology* **20**, 1099–1103 (2010).
27. Ullman, G. *et al.* High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **368** (2012).
28. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503**, 481–486 (2013).
29. Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron : groupe de gènes à expression coordonnée par un opérateur. *Comptes Rendus de l'Académie des Sciences, Paris* **250**, 1727–1729 (1960).
30. Jacob, F. & Monod, J. *On the regulation of gene activity in Cold Spring Harbor Symposia on Quantitative Biology* **26** (Cold Spring Harbor Laboratory Press, 1961), 193–211.
31. Cohn, M. & Monod, J. Purification et propriétés de la β -galactosidase (lactase) d'*Escherichia coli*. *Biochimica et Biophysica Acta* **7**, 153–174 (1951).
32. Burstein, C., Cohn, M., Kepes, A. & Monod, J. Rôle du lactose et de ses produits métaboliques dans l'induction de l'opéron lactose chez *Escherichia coli*. *Biochimica et Biophysica Acta - Nucleic Acids and Protein Synthesis* **95**, 634–639 (1965).
33. Jobe, A. & Bourgeois, S. *lac* repressor-operator interaction. *Journal of Molecular Biology* **69**, 397–408 (1972).
34. Huber, R. E., Wallenfels, K. & Kurz, G. The action of β -galactosidase (*Escherichia coli*) on allolactose. *Canadian Journal of Biochemistry* **53**, 1035–1038 (1975).

35. Kaczorowski, G. J., Robertson, D. E. & Kaback, H. R. Mechanism of lactose translocation in membrane vesicles from *Escherichia coli*. 2. Effect of imposed $\Delta\Psi$, ΔpH , and $\Delta\mu_{\text{H}^+}$. *Biochemistry* **18**, 3697–3704 (1979).
36. Kaczorowski, G. J. & Kaback, H. R. Mechanism of lactose translocation in membrane vesicles from *Escherichia coli*. 1. Effect of pH on efflux, exchange, and counterflow. *Biochemistry* **18**, 3691–3697 (1979).
37. Patel, L., Garcia, M. L. & Kaback, H. R. Direct measurement of lactose/proton symport in *Escherichia coli* membrane vesicles: further evidence for the involvement of a histidine residue(s). *Biochemistry* **21**, 5805–5810 (1982).
38. Andrews, K. J. & Lin, E. C. Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification. *Journal of Bacteriology* **128**, 510–513 (1976).
39. Schlux, P. J., Capp, M. W. & Record Jr, T. M. Inhibition of transcription initiation by *lac* repressor. *Journal of Molecular Biology* **245**, 331–350 (1995).
40. Oehler, S., Eismann, E. R., Krämer, H. & Müller-Hill, B. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* **9**, 973–979 (1990).
41. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Müller-Hill, B. Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. *The EMBO Journal* **13**, 3348–3355 (1994).
42. Lewis, M. *et al.* Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254 (1996).
43. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nature Structural and Molecular Biology* **7**, 209–214 (2000).
44. Monod, J. Recherches sur la croissance des cultures bactériennes. *Paris: Hermann et Cie* (1942).
45. Monod, J. The phenomenon of enzymatic adaptation. *Selected Papers in Molecular Biology by Jacques Monod*, 68 (2012).

46. Epstein, W., Naono, S. & Gros, F. Synthesis of enzymes of the lactose operon during diauxic growth of *Escherichia coli*. *Biochemical and Biophysical Research Communications* **24**, 588–592 (1966).
47. Makman, R. S. & Sutherland, E. W. Adenosine 3',5'-Phosphate in *Escherichia coli*. *Journal of Biological Chemistry* **240**, 1309–1314 (1965).
48. Ullmann, A. & Monod, J. Cyclic AMP as an antagonist of catabolite repression in *Escherichia coli*. *FEBS Letters* **2**, 57–60 (1968).
49. Perlman, R. & Pastan, I. Cyclic 3'5'-AMP: stimulation of beta-galactosidase and tryptophanase induction in *E. coli*. *Biochemical and biophysical research communications* **30**, 656–664 (1968).
50. Silverstone, A. E., Magasanik, B., Reznikoff, W. S., Miller, J. H. & Beckwith, J. R. Catabolite sensitive site of the *lac operon*. *Nature* **221**, 1012–1014 (1969).
51. Perlman, R. L., De Crombrugghe, B. & Pastan, I. Cyclic AMP regulates catabolite and transient repression in *E. coli*. *Nature* **223**, 810–812 (1969).
52. Zubay, G., Schwartz, D. & Beckwith, J. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proceedings of the National Academy of Sciences of the United States of America* **66**, 104–110 (1970).
53. Schwartz, D. & Beckwith, J. R. Mutants missing a factor necessary for the expression of catabolite-sensitive operons in *E. coli*. *The lactose operon*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 417–422 (1970).
54. Emmer, M., deCrombrugghe, B., Pastan, I. & Perlman, R. Cyclic AMP receptor protein of *E. coli*: its role in the synthesis of inducible enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **66**, 480–487 (1970).
55. Beckwith, J., Grodzicker, T. & Arditti, R. Evidence for two sites in the *lac* promoter region. *Journal of Molecular Biology* **69**, 155–160 (1972).
56. Hopkins, J. D. A new class of promoter mutations in the lactose operon of *Escherichia coli*. *Journal of Molecular Biology* **87**, 715–724 (1974).

57. Inada, T., Kimata, K. & Aiba, H. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes to Cells* **1**, 293–301 (1996).
58. Deutscher, J. *et al.* The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiology and molecular biology reviews : MMBR* **78**, 231–256 (2014).
59. Osumi, T. & Saier, M. Regulation of lactose permease activity by the phosphoenolpyruvate:sugar phosphotransferase system: evidence for direct binding of the glucose-specific enzyme III to the lactose permease. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 1457–1461 (1982).
60. Sondej, M., Weinglass, A. B., Peterkofsky, A. & Kaback, H. R. Binding of enzyme IIA^{Glc}, a component of the phosphoenolpyruvate:sugar phosphotransferase system, to the *Escherichia coli* lactose permease. *Biochemistry* **41**, 5556–5565 (2002).
61. Lambert, G. & Kussell, E. Memory and fitness optimization of bacteria under fluctuating environments. *PLoS Genetics* **10**, e1004556 (2014).
62. Duffy, D. C., McDonald, J. C., Schueller, O. J. A. & Whitesides, G. M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Analytical Chemistry* **70**, 4974–4984 (1998).
63. Madou, M. J. *Fundamentals of microfabrication: the science of miniaturization* (CRC press, 2002).
64. Ferry, M. S., Razinkov, I. A. & Hasty, J. in *Synthetic Biology, Part A* (ed Voigt, C.) 295–372 (Academic Press, 2011).
65. Brody, J. P., Yager, P., Goldstein, R. E. & Austin, R. H. Biotechnology at low Reynolds numbers. *Biophysical Journal* **71**, 3430–3441 (1996).
66. Stroock, A. D. *et al.* Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).

Acknowledgements

During my doctoral study I was lucky to work with great people that helped me progress in my project. Especially I would like to thank my PhD advisor Erik van Nimwegen for giving me the chance to do my thesis in his group and for inspiring me with the many scientific ideas he contributed. I would also like to thank Olin Silander and Thomas Julou who directly supervised me in the wet lab and Chris Field for his help with the thesis manuscript. A special thanks also goes to my collaboration partners Florian Jug, Eugene Myers, Siddharth Deshpande and Thomas Pfohl. The collaborations were a great experience for me and allowed me to get an insight into other fields. I am very grateful to have had nice colleagues in both the wetlab and the computational part of the van Nimwegen group that always supported me and created a good atmosphere. But I did not only get support on the scientific side. I always knew that my family and friends were behind me, supported me a lot and helped me through difficult times in my doctoral study. I really appreciate this support and will always be grateful.

Curriculum Vitae

Matthias Kaiser

PERSONAL INFORMATION

Address: Elfenastrasse 33, 6005 Luzern

Phone: +41 76 386 26 81

E-Mail: m_kaiser@sunrise.ch

Date of birth: 15.08.1986

Languages: German (Native), English (Fluent), French (Basic)

ACADEMIC EDUCATION

2001-2005 Kantonsschule Solothurn, Switzerland

2005-2008 University of Basel, Switzerland, Bachelor of Science in Biology
Major in Molecular Biology

2008-2010 University of Basel, Switzerland, Master of Science in Molecular Biology
Master Thesis: "Second messenger-mediated adjustment of bacterial swimming velocity"
Supervised by Prof. Dr. Urs Jenal

2011-2016 University of Basel, Switzerland, PhD in Bioinformatics
PhD Thesis: "A Microfluidic Setup for Quantifying Single-Cell Transcription Regulatory Dynamics"
Supervised by Prof. Dr. Erik van Nimwegen

PUBLICATIONS

Boehm, A., Kaiser, M., Li, H., Spangler, C., Kasper, C. A., Ackermann, M., Kaever, V., Sourjik, V., Roth, V., Jenal, U. Second Messenger-Mediated Adjustment of Bacterial Swimming Velocity. *Cell* **141**, 107-116 (2010).

Jug, F., Pietzsch, T., Kainmüller, D., Funke, J., Kaiser, M., van Nimwegen, E., Rother, C., Myers, G. in *Bayesian and Graphical Models for Biomedical Imaging* (eds Jorge Cardoso, M., Simpson, I., Arbel, T., Precup, D. & Ribbens, A.) (Springer, Cambridge, MA, 2014).

Kaiser, M., Jug, F., Julou, T., Deshpande, S., Pfohl, T., Silander, O. K., Myers, G., van Nimwegen, E. Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software. *Nature Communications* **9**,212 (2018).

FELLOWSHIPS

International PhD Program "Opportunities for Excellence" of the Werner Siemens Foundation and the Biozentrum of the University of Basel