SINGLE MAMMALIAN CELL GENE EXPRESSION ANALYSIS USING MICROFLUIDICS

Thesis by

Joshua S. Marcus

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, CA

2006

(Defended April 11, 2006)

Acknowledgements

I thank my advisor Steve Quake for above all, his unwavering support throughout my graduate career. Steve is not only brilliant, but also a master motivator, leader, and role model. To me, he is the best kind of leader–one who lets his passion, commitment, results, and work ethic, do the talking. He has allowed me to innovate and develop my work independently, all while still being an integral part of the research. I hope that the work presented here has left the lab in a better position than when I started, and will not be delegated to the dustbins of history.

I thank Steve Mayo for making my transition to graduate school smooth and affording me the opportunity to train in his lab. I am also grateful to Steve for being someone I could turn to for advice and counsel. Seeing his scientific expertise as well as how he supported me, it is easy to see Steve is a first rate advisor. Steve had Premal Shaw train me in the lab; and I thank Premal for teaching me many of the molecular biology and biochemistry techniques necessary to succeed in the field. Premal has become a friend as well as a colleague and I am fortunate that our paths crossed while at Caltech.

I thank the members of my thesis committee – Judy Campbell, Jackie Barton, Jim Heath and Steve Mayo. All have been supportive of my research and goals and I am better off for having them as a part of my graduate career. I am sincerely grateful to Jim as well as Mel Simon, for opening their respective labs to me and allowing me to use their equipment. Without Jim and Mel, the research would not have progressed at the rate it did. Working in the Heath and Simon labs provided me the opportunity to gain unique

perspectives on my work through the eyes of their respective members. In particular, I thank John Nagarah, Sangdun Choi, and Misook Chang for helpful and stimulating discussions.

I thank French Anderson for his wisdom, dedication, enthusiasm, and resolve. French has been a part of the work since the inception and has treated me like one of his own students. I am fortunate to have worked with him. French has an unparalleled lab manager in Kathy Burke and I thank her for all the help and advice she has given me. Kathy has been a great outlet to talk to about everything from cell culture to qPCR to life in general. I also thank Jiang (John) Zhong, another member of French's group. John sorted the cells used for the study presented in Chapter 5, and has been a valuable resource for all things biology.

I have been extremely fortunate to work with many skilled scientists in my time at Caltech. I am most grateful to Carl Hansen, for his advice, friendship, and example. Carl was the catalyst behind the technology innovation in Chapter 5. Although the device is nothing like our first design, seeing Carl's confidence in a possible solution to the problem pushed me to find a correct one. I hope our friendship continues to grow, and look forward to working with him on the intellectual property portion of his research. Sebastian Maerkl has also been a vital resource in the lab. I am grateful for the insight he has imparted to me and conversations we have had over the years. I am grateful to the remaining members of the Quake group and support staff for their scientific and engineering skills, logistical support, and their friendship. In particular I thank Jong

Wook Hong, Luigi Warren, Emil Kartalov, and Chris Lacenere for useful discussions and advice. I also thank Chris Lacenere for his help with video editing, being someone to vent to about the state of the world, and for looking after my dog on occasion.

I thank the microfluidic foundry for fabricating some of the devices used throughout my research. In particular, I am grateful to Alejandra Torres for device fabrication, putting up with my perfectionist attitude towards fabrication, and for being a very good friend. Your friendship and your devotion to my dog are incredible. I also thank Ali Ghaffari for help and discussions concerning optical lithography issues.

I thank my dog and best buddy, Akili, for his unconditional love. We have been a team for seven years and it's hard to imagine my life without him. Akili is stricken with hereditary ataxia, an ailment that has progressively left him in little control of his muscles and movements. Seeing the fight in Akili and how he still loves living life, regardless of how difficult it is to get around, is astounding. I hope his resolve and care free attitude have rubbed off on me.

I am eternally grateful to my parents for their unconditional love, support, guidance, and knowledge that whatever comes of my life, they will be by my side. I thank my twin brother, Brad, who is also my best friend. I would do anything for him and am fortunate to have his love, friendship, and counsel. I look forward to the next stage of our lives, when we will be able to see more of each other. I am grateful to Sarit Bulmash for being above all, my best friend. Sarit and I have known each other eight years, and she is truly

remarkable. I hope I have matched a fraction of the passion, love, dedication, and determination you show everyday. I know our friendship will continue to grow over the years. The following work could not have been realized without the support of you four.

Abstract

Single cell gene expression studies hold great promise for deciphering the ubiquitous heterogeneity present in biological organisms. Although much progress has been made in the field, tools to study gene expression (specific and global) in single cells are generally lacking. This thesis describes the development of novel microfluidic technologies and processes capable of processing single cells to first strand cDNA in a parallel fashion, thereby filling a void in the single cell biology field. The author then utilizes the technology to probe for transcriptional noise in ubiquitous genes present in single mammalian cells. The noise measured far exceeds any measurement reported to this date, and was shown to be attenuated during the G2 stage of the cell cycle. The work presented here is first hand proof that technological innovation is a key component in undertaking novel science.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Overview

1.1 Introduction

Vast cellular heterogeneity is a common theme in all biological organisms, and the ability to interrogate cells in an individual manner could elicit the presence of subpopulations and provide insight into specific processes, such as stem cell differentiation events, which are masked at the population level. It has become increasingly clear that studying cell populations, regardless of what is being assayed (e.g. mRNA, protein, or small molecule levels) is only sufficient for obtaining average values over the particular populations or many subpopulations. Major cellular decisions concerning growth, death and differentiation are reflected in altered patterns of gene expression in individual cells. The ability to quantify transcription levels of individual genes is a necessity in understanding these processes.[1] Although an assessment of mRNA or protein expression does not directly address function, the knowledge of when and where a gene is expressed can provide information about potential roles of the respective gene; and has been historically used for gene discovery.[2] While questions pertaining to what genes and proteins are expressed in single cells, to how individual cells communicate and respond to different stimuli are all active research areas, tools to answer these questions are generally lacking. The main reason for the lack of technologies to study the single cell can be attributed to the poor yields encountered when subjecting these precious samples to multi-step processes.

The state of the art in global gene expression studies in single cells has utilized single cell reverse transcriptase-polymerase chain reaction (RT-PCR)[3, 4] or T7 RNA amplification[5] coupled to product hybridization to oligonucleotide microarrays.

Although thousands of genes could be probed in parallel from single cells, there were inherent problems with the assays and consequently only a minimal amount could be learned from the data obtained. The amplification procedure was non-linear for the PCR studies (50 cycles), so it is safe to assume products generated were not an accurate reflection of the starting material present in each cell. Only 16 cells[4] and 60 cells[3] were analyzed for each study, so the stochasticity of gene expression could not be fully explored. In the case of Dulac's group, 90 cells were isolated by laser capture microdissection (LCM), but only 16 exhibited robust expression of the ubiquitous transcript glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Although the T7 amplification procedure is linear in nature, LCM was utilized to isolate cells, which may have damaged the integrity of the samples, as was most likely the case with Dulac's study.[4] The isolation method also limits the number of samples that can be processed.

Quantitative gene expression studies on one or a limited number of genes can be performed with qPCR (discussed further in chapter 5). Although multiplex qPCR is common for bulk samples, with a PubMed literature search yielding hundreds of papers, qPCR, multiplex or otherwise, is much rarer for single cells, due to the sample processing issues stated above[2, 6-15]. The multiplex single cell qPCR studies presented in this thesis are the first of their kind, with detection of three or four distinct gene segments in the initial log phase of the PCRs.

Clearly, there is room for single cell gene expression assay improvement. Although one can query for the presence of thousands of genes in parallel with microarrays, there is a limitation to a finite number of genes, so generating a full mRNA expression profile is not possible. Also, if an interesting result is found, one does not

have the ability to isolate the gene sequence of interest and make the corresponding protein. Compounding these issues, variation in results from researcher to researcher is also a problem because of non-genetic influences such as quality of spotted DNA, surface chemistries, and analysis methods.[16] Furthermore, when a PCR is implemented, a true reflection of initial levels of gene expression cannot be achieved on a regular basis unless detection is employed during the initial log phase of the reaction.

Ideally, one would like to eliminate PCR and employ an *a priori* approach such as cDNA library construction to analyze single cells, so as not to introduce biases towards certain sequences, limit oneself to a finite number of genes, and so protein can be made from individual clones, if desired. Because of the *a priori* nature of library construction, it facilitates the discovery of genes important for cellular processes and phenomena. When amplification becomes necessary to generate sufficient amounts of cDNA, global linear amplification methods utilizing an RNA polymerase are preferred, to keep an accurate reflection of the starting material. To date, 10^4-10^5 cells are the minimum number needed to construct a cDNA library.[17] Because of all the advantages cDNA library construction has over current global gene expression analysis methods, the technique would be extremely useful if applied to single cells.

Microfluidic assays overcome some of the limitations of conventional single cell gene expression studies. The typical channel dimensions found in microfluidic devices (10 μm–100 μm in x,y and z) and the ability to manipulate nanoliters of reagents on-chip have made the devices encouraging platforms for the analysis of single cells.[18] Furthermore, the economy-of-scale benefits along with the ability to parallelize and automate processes are significant advantages not found with conventional biological

assays. Whole tissue or cultured cells can be placed in liquid suspension and manipulated on-chip. Active valves[19] make it possible to individually address compartments of the chip[20] and precisely mix and meter an array of reagents with different properties.[21, 22] Furthermore, the ability to build affinity columns directly on-chip makes it possible to obtain a pure nucleic acid sample from whole cell lysate.[23, 24] The work presented in this thesis encompasses the first five steps of cDNA library construction, applied microfluidically, and can be built on by future students to realize the full process. Furthermore, the technology disclosed in this thesis is mature enough to perform various quantitative single cell gene expression assays. Here, the author discloses the first reported measurements of gene expression noise in mammalian cells, as well as the first studies to measure noise directly at the transcription level.

1.2 Context

The technologies in this thesis utilize and build upon the multilayer soft lithography (MSL) process[19] developed in the Quake group. At the inception of this work, MSL had been applied to numerous biological questions, including but not limited to protein crystallography,[22, 25] nucleic acid processing,[24] FACS,[26, 27] enzyme assays,[20] and PCR.[28, 29] Although nucleic acid processing had been demonstrated previously, this work provides a more robust implementation with quantitative calibrations, and provides a building block for integration of downstream steps to eventually synthesize cDNA libraries from single cells. The thesis describes the first application of MSL to the quantitative study of gene expression in single cells. In addition, the integrated process of cell capture/lysis → mRNA isolation/purification → cDNA synthesis/purification is

implemented for the first time on a microfluidic device. In all, four different devices are presented throughout the work, with three pertaining to single cell processing for gene expression analysis. The fourth device is a PCR device intended to carry out picoliter RT-PCRs in a parallel fashion. The RT-PCR device provides substantial improvements, as compared to the state of the art in microfluidic RT-PCR, in detection sensitivity, volumes utilized, and degree of parallelization.

The three remaining devices are similar, in that they all mechanically isolate single cells, purify the mRNA inside the cells, synthesize and purify the corresponding first strand cDNA. It will become obvious to the reader that technological innovations and need for sample parallelization and 100% reactor addressability pushed the device to a third generation, which provides the most robust implementation of the aforementioned process. The first reported gene expression noise measurements on mammalian cells are performed on single cell cDNA samples processed with the second generation cells-to-cDNA device. The single mammalian cell noise measurements are the culmination of the thesis work and demonstrate firsthand how technological innovation is crucial for performing novel science.

1.3 Organization

Chapter 2 describes the first cells-to-cDNA device, in which samples were processed in a serial fashion. The chapter describes early proof-of-principle studies aimed at validating the steps in the microfluidic process of cell lysis/mRNA isolation/cDNA synthesis/product recovery.

Chapter 3 is an extension of the work presented in chapter 2, but with some modifications. The manuscript was published as an article in *Analytical Chemistry*. The chip presented in the manuscript is capable of processing four samples in parallel and was utilized to perform quantitative calibrations of the microfluidic cells-to-cDNA method. Quantitative single cell gene expression measurements were also performed with samples processed on-chip.

Chapter 4 describes qPCR in detail, along with multiplex primer/probe design for efficient and reproducible qPCR. The state of the art in single cell qPCR is discussed, along with considerations for robust assays. Primer and probe validation assays are presented for both multiplex qPCR studies on murine genes, as well as bacterial primer design for single cell spike experiments.

Chapter 5 presents the first reported measurements of gene expression noise at the transcriptional level, in single mammalian cells. Highly parallel microfluidic cells-to-cDNA devices used for these measurements are discussed, along with process refinements and calibrations.

Chapter 6 is a technical note published in *Analytical Chemistry*[30] presenting a device designed to carry out picoliter RT-qPCRs or qPCRs in a highly parallel fashion. Results on RNA templates are offered.

Chapter 7 discusses MSL in general, along with technical advancements that were made by the author during the course of his thesis research.

Appendix A presents the major protocols utilized throughout the thesis research. qPCR protocols for microfluidic and conventional samples, as well as cell labeling for

FACS analysis, photolithography processes and MSL processes are all discussed in detail.

Appendix B is the author's contribution to a book chapter entitled "Progress in nanofluidics for cell biology" to be published in the second edition of the *CRC Handbook for nanoscience, technology, and engineering*.

Appendix C is a manuscript entitled "Prudent modeling of core polar residues in computational protein design" published in the *Journal of Molecular Biology*. The majority of the work was performed by Daniel Bolon, a former Caltech graduate student of Stephen Mayo's. My contribution to this work included synthesizing one of the gene constructs by site directed mutagenesis, expressing the corresponding protein, and performing various biochemical assays on the construct, including circular dichroism (CD) spectroscopy and NMR spectroscopy (along with Scott Ross).

Chapter 2 - Proof of principle mRNA isolation / cDNA synthesis studies

Introduction

Previous work[24] in the Quake laboratory provided a starting point for the design of a robust microfluidic single cell nucleic acid analysis device. The following chapter describes refinements to the first generation cell lysis/mRNA purification device (developed by Vincent Studer) and describes the second generation chip along with preliminary results with it. End point PCR validation assays on microfluidic samples are presented to confirm steps in the process. In addition, proof-of-principle qPCR studies are demonstrated to show the feasibility of quantifying microfluidic samples by qPCR. Preliminary results pertaining to the integration of a downstream first strand cDNA synthesis step are also presented.

2.1 Chip refinements

The author's first task was to revise the chip design in order to achieve a more robust device. Lysis by diffusion took at least 15 minutes, and beads clumped after the diffusion step. The clustering of beads and the inability to release to output was taken to mean the lysis step did not go to completion. Accordingly, the author introduced a rotary mixer into the chip design to lyse cells in an active manner. Fig. 2.1 shows the column portion of a device after mRNA affinity capture and bead release. Cells were subjected to lysis by free interface diffusion in Fig. 2.1A and rotary lysis in Fig 2.1B. Although all mixing on microfluidic length scales is eventually accomplished by diffusion, the rotary mixer speeds up the process considerably by increasing diffusion length scales.

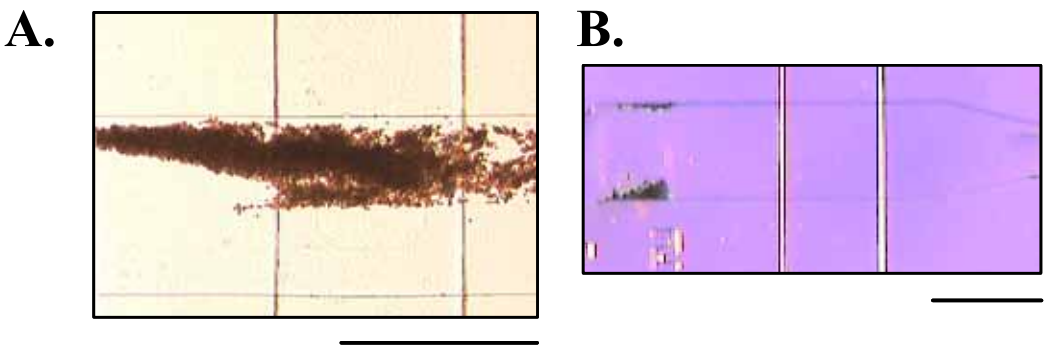


Fig. 2.1. Optical micrographs of column construction area after bead release. Beads were released by pneumatic pressure to an output port after cells-to-cDNA process was complete. **Left**: Cell was lysed by free interface diffusion. **Right**: Cell was lysed in a rotary mixer. Scale bars are 200 μm.

The author also refined the way in which the affinity purification column is built. The implementation presented here and in later chapters provides a significant upgrade, providing digital control over column construction, as compared to the analog control in the first generation device.[24] The column of oligo$(dT)_{25}$ derivatized paramagnetic beads are stacked against a sieve valve (Figure 2.2, blue valve in micrograph), present on a 13 μm high flow structure with a rectangular cross section, molded out of SU8-2015. The previous column design used a slightly opened valve (present on a flow channel with semi-circular cross section) to allow fluid flow while not allowing the 2.8 μm beads to pass through.[24] Parameter space sufficient to build both types of columns, as well as fabrication considerations, are given in chapters 3 and 7, respectively.

The second generation chip is presented in Figure 2.2. Each 1" X 1" chip is able to carry out two isolations, which is one more than its first generation counterpart.

Furthermore, when the isolation is complete, the beads are sent to the output port and collected with a pipette, as opposed to cutting the chip to retrieve products. In the author's scheme, the chip can be cleaned and reused, or disposed of. The ability to process > 1 sample/chip decreases fabrication time by half when multiple measurements are necessary.

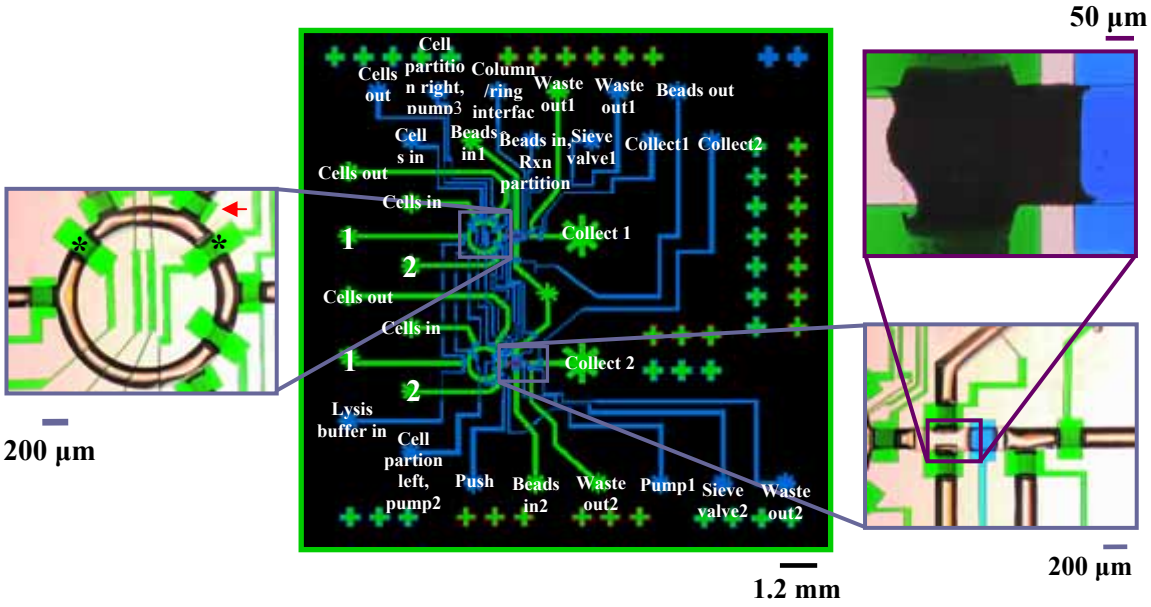


Figure 2.2. Second generation cell lysis/mRNA purification chip. Inputs for control and flow channels are labeled according to function. **Middle**: AutoCAD drawing of the chip. Control channels are in blue and flow channels are in green. The column flow structure is depicted in red. **Left:** Optical micrograph of the lysis ring. **Right:** Optical micrograph of the column area and a stacked affinity column.

2.2 Experimental design and chip layout

As mentioned above, each chip is able to carry out two mRNA purifications. Purified NIH/3T3 mRNA (Ambion) or NIH/3T3 cell suspension is geometrically metered

into the mixing ring between the valves marked with asterisks (Fig. 2.2 left). The remainder of the ring is dead end filled with lysis buffer to the valves marked with asterisks, through flow inlet 1 (Fig 2.2). The contents of the ring are mixed, leading to cell lysis in a matter of seconds (visualized by optical microscopy). The short mixing time is possible because of the parabolic flow profile taken on by the fluid slugs. Dispersion stretches the fluid streams, thereby decreasing the diffusion length by maximizing the interface between the fluid streams.[21, 22]

Once the ring is mixed, a bolus of air is applied through flow inlet 2 (Fig 2.2 middle), pushing the lysate over the stacked beads at 10-20 μm/s to allow for extended interaction between the affinity column and mRNA transcripts. A column wash step is then performed to deal with non-specific interactors that may be present. The beads are collected in reverse transcriptase-polymerase chain reaction (RT-PCR) buffer and frozen at -80°C for future analysis.

2.3 On-chip mRNA isolation results

The first experiments the author carried out were with purified NIH/3T3 mRNA (Ambion) and NIH/3T3 cells as templates for lysis and subsequent mRNA affinity capture. In order to gauge the starting point sensitivity (e.g. without any optimization or procedural refinements) of the bead column, various dilutions of mRNA were loaded in the top portion of the rotary mixer (Fig. 2.2). For the cell capture/lysis → mRNA isolation/purification experiments, murine NIH/3T3 cells were loaded into the region of the ring where mRNA was loaded. The valve marked with a red arrow in Figure 2.2 was opened and closed until the desired number of cells were present in the ring. After

loading the remainder of the ring with lysis buffer, the ring contents were mixed for 10 seconds and pushed via pneumatic pressure-driven flow (through inlet 2, Fig. 2.2) over the oligo$(dT)_{25}$ column. The beads were then washed in a salt solution and subsequently collected in RT-PCR buffer. Conventional RT-PCR (35 cycles, 50 μL reactions) or real-time RT-PCR (20 μL reactions) on a 150 base pair (bp) segment of the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene was then carried out on the mRNA bound to the beads.

2.3.1 End point RT-PCR analysis: *GAPDH*

Results from conventional RT-PCR experiments from recovered on-chip sample templates are shown in Figure 2.3. A clear, distinct band was seen for initial mRNA templates ranging from 50→1 pg. Although there was no distinct band for the initial template of 100 fg, the author was able to retrieve the mRNA from the chip, as seen by the smear of DNA present on the gel (Fig 2.3). Since all the beads were retrieved, it is likely either there were not enough initial copies of the gene to amplify efficiently, or primer-dimer amplification competed for PCR reagents. Judging by the bands in the noise of the gel, the latter is the case. Future studies will utilize primers with less of a tendency to form non-specific products. Results for the on-chip cell capture/lysis → mRNA isolation/purification experiments are also shown in Figure 2.3. Distinct bands were visualized for all cellular templates and single cell resolution was achieved. Negative control experiments (PBS from cell suspension) confirmed the absence of contamination.

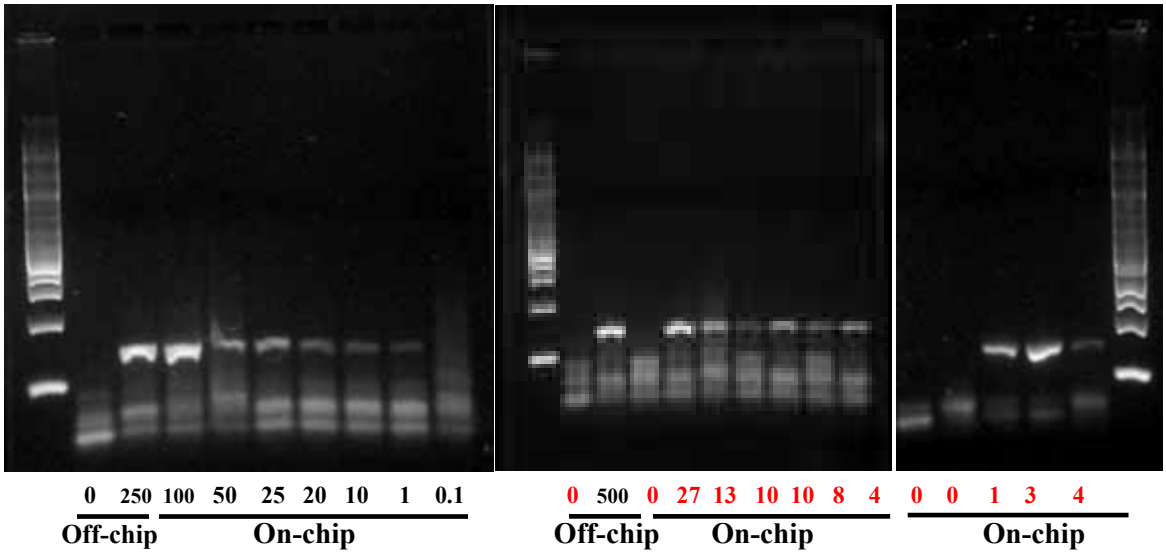


Figure 2.3. Agarose gel electrophoresis on microfluidic samples. 1% agarose gel was run after 35 cycles of an RT-PCR on purified NIH/3T3 mRNA and NIH/3T3 cells. Black numbers are pg NIH/3T3 mRNA and red numbers are NIH/3T3 cells. The band is a 132 bp portion of GAPDH.

2.3.2 Genomic DNA signals

Because the GAPDH primers do not span introns, the author needed to check for genomic DNA contamination (gDNA) in the mRNA preparations. Six cells and 19 cells were lysed on-chip and the lysate was driven over the affinity column. The beads were then manipulated as previously described and collected in RT-PCR buffer in the absence of enzymes. Each reaction was then split into two tubes. The beads in each tube were then pelleted by centrifugation and resuspended in either RT-PCR buffer with reverse transcriptase (RT) or without RT. An RT-PCR was then run to probe for genomic DNA signal, with results from this experiment shown in Figure 2.4. A faint band for the six cell mRNA isolation in the absence of RT was observed. However, no band was

observed for the 19 cell preparation in the absence of RT. These results indicate the contamination was from an outside source and was not a genomic signal arising from the NIH/3T3 cells. The author then repeated the experiment with β-actin primers that spanned introns, and did not split the samples in half (Fig. 2.4). Both no RT reactions were negative for the presence of DNA while both reactions with RT displayed a signal (Figure 2.4). These results indicate the chip itself is not a source of contamination and sample manipulation off-chip is the main contributor to false positive signals. Future studies will control for contamination by having all primers span introns, as well as controlling for sequence homology with human DNA (not done with GAPDH primers).

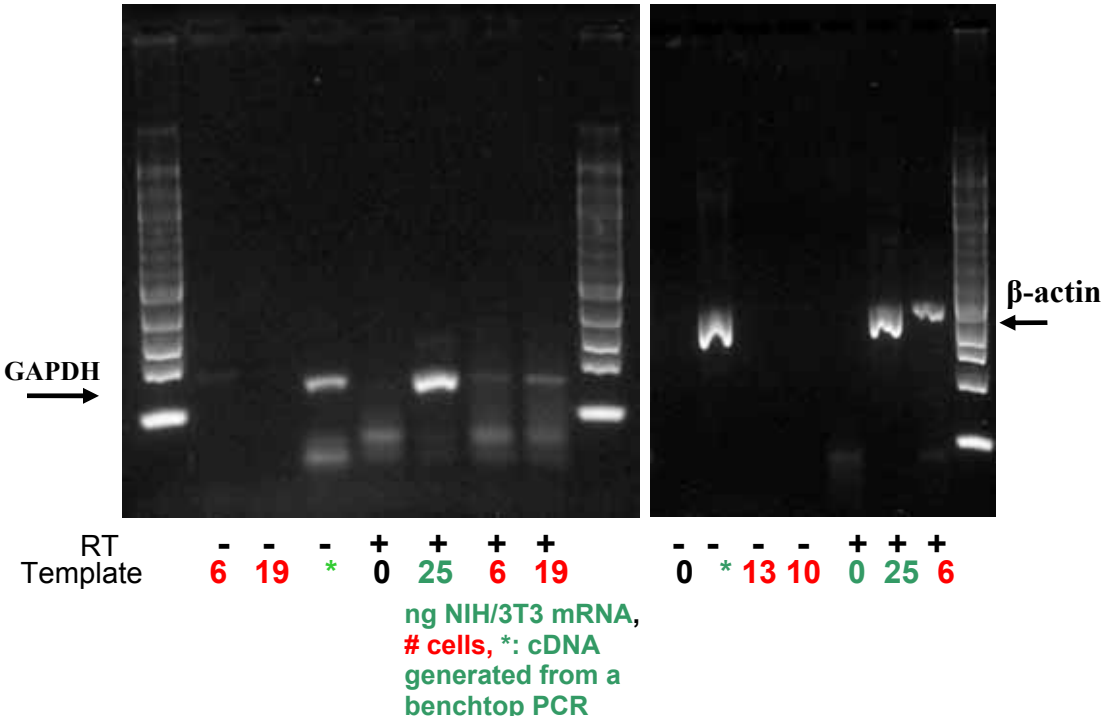


Figure 2.4. Agarose gel electrophoresis on genomic DNA control samples. mRNA samples were purified on-chip (and/or lysed beforehand) followed by product collection and an RT-PCR and gel electrophoresis of the PCR products.

2.3.3 End point RT-PCR analysis: *OZF*

In the next set of mRNA isolations, the author probed for a 399 bp portion of the transcription factor OZF, which is expressed in much lower abundance than *GAPDH*. Results from these isolations are shown in Figure 2.5. The equivalent of roughly the amount of mRNA present in one cell[31] (1 pg) was able to be recovered and amplified to obtain a signal.

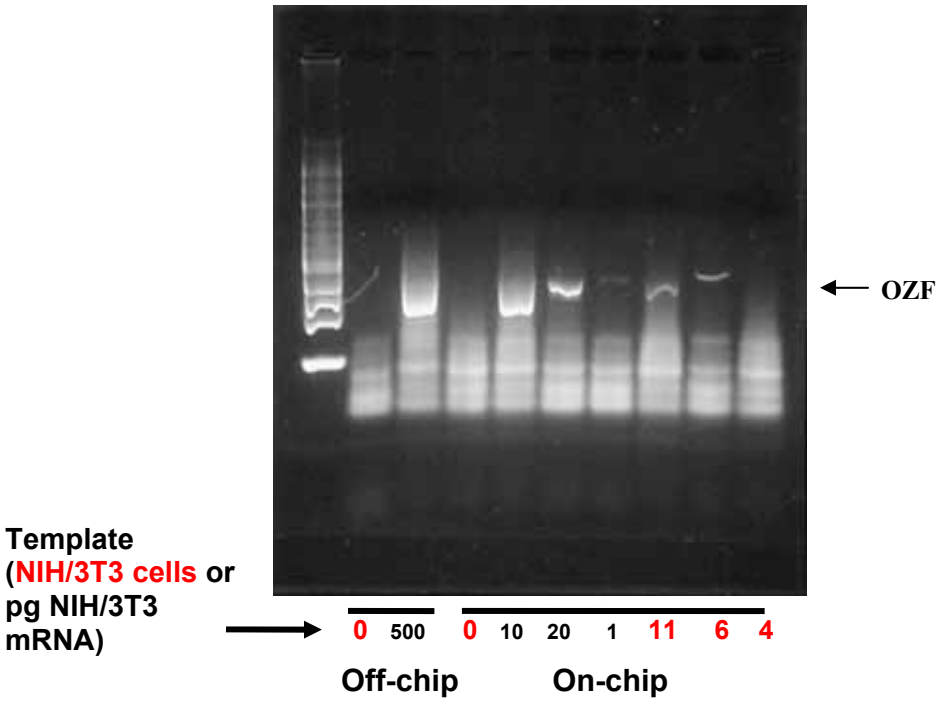


Figure 2.5. Low copy gene detection from microfluidic samples. Agarose (2%) gel electrophoresis run after a RT-PCR, in which the author probed for a 399 bp segment from the OZF gene.

2.3.4 Real-time RT-PCR results

2.3.4.1 Feasibility Studies

Because µL amounts of the beads (Dynal) used to construct mRNA affinity columns fluoresce under green light, it was imperative to see if the ~3 nL column would contribute to a SYBR Green I signal during real-time detection of product formation. Accordingly, on-chip isolations with various templates were carried out, followed by real-time RT-PCR analysis (RT-qPCR utilized interchangeably throughout thesis; discussed in chapter 4 in greater detail). The recovered column from the no-template purification yielded no detectable fluorescence signal when detection was enabled at the appropriate temperature (Fig 2.6). These results allow for subsequent samples to be analyzed in real-time to obtain quantitative gene expression data, unattainable when utilizing end-point analysis.

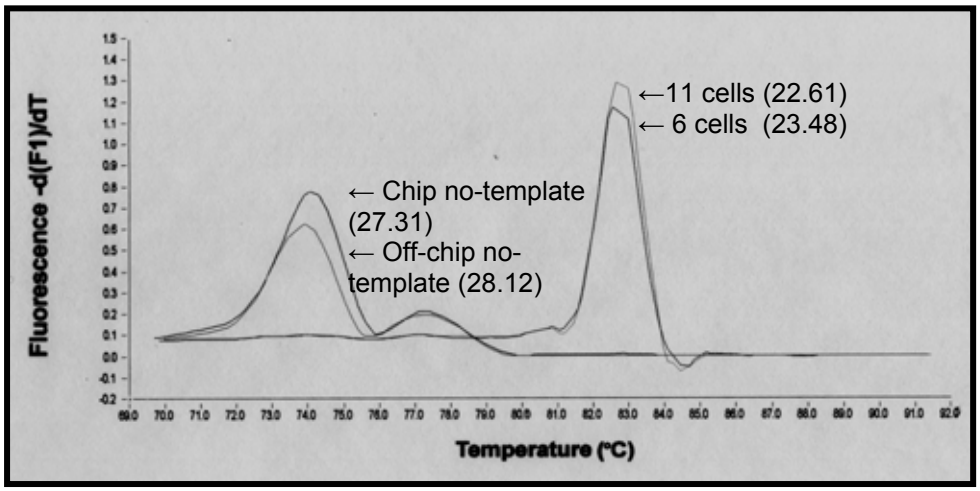


Figure 2.6. Melting peak analysis on microfluidic samples subjected to real-time PCR. Threshold cycle values are in parenthesis. Primer-dimer peaks are present at 74°C, while

product peaks are present at 83°C. During cycling, detection was enabled at 79°C to eliminate signal from primer-dimers.

2.3.4.2 Serial mRNA isolation calibration

In order to determine how reliable and robust the microfluidic cells→mRNA method is, the author generated standard curves with known NIH/3T3 mRNA templates loaded onto the chip. These experiments give an indication of the column's dynamic range of capacity as well as recovery efficiency. Known NIH/3T3 mRNA templates were loaded into the lysis ring, mixed with lysis buffer, and subsequently pushed via pneumatic pressure over the columns. The columns were then washed as described previously, and collected by pipetting. The templates were then subjected to real-time RT-PCRs. Results from two separate experiments are plotted together and shown in Figure 2.6. As the figure shows, the method of isolation and recovery is quantitative and reproducible over three orders of magnitude (R = -0.99). The process efficiency was then extrapolated by plotting the microfluidic threshold cycles onto a GAPDH standard curve generated with off-chip NIH/3T3 mRNA standards (Fig. 2.6B). The average efficiency for all samples processed was 47% that of an off-chip RT-PCR with benchtop standards.

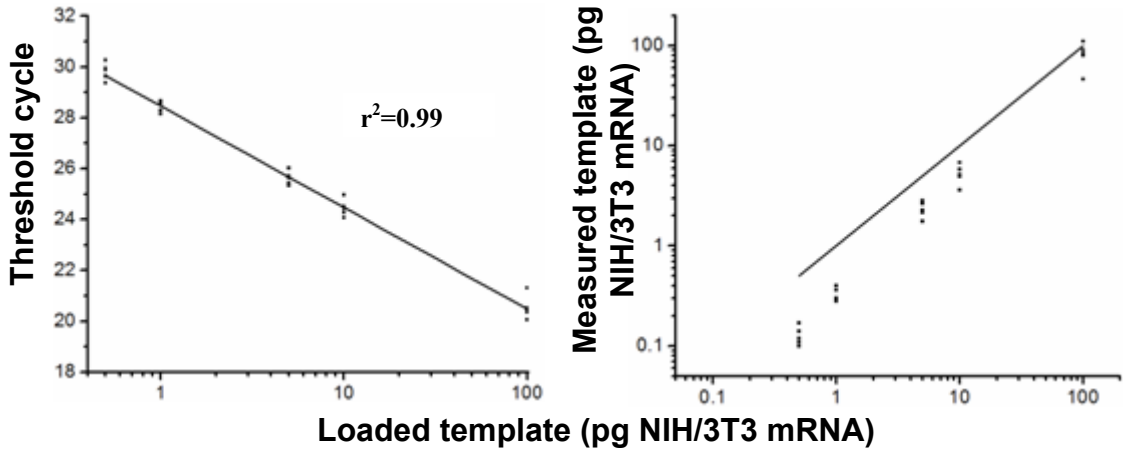


Figure 2.7. Microfluidic mRNA isolation calibration and efficiency measurements. Graphs are of data pooled from two independent experiments. **A.** mRNA isolation standard curve. Threshold cycle corresponds to the PCR cycle when fluorescence measured is significantly higher than background levels. **B.** The line's slope (m=1) indicates 100% efficiency. Measured template values were calculated by plotting the Cts of the microfluidic samples onto an off-chip GAPDH standard curve generated with identical off-chip template concentrations.

2.5 On-chip first strand cDNA synthesis integration

After mRNA isolation, cDNA synthesis is the next step in the process of creating a cDNA library. Furthermore, the preference is to work with cDNA rather than mRNA because of ambient RNases and subsequent degradation issues. Because of these two factors, it would be well served to integrate a cDNA synthesis step with the existing microfluidic process. The author chose to integrate the on-chip cDNA synthesis step by implementing a solid phase synthesis strategy. The oligo$(dT)_{25}$ sequences present on the beads were used both to capture the mRNA and as primers for first strand synthesis. The chip was heated to 40°C on a thermal microscope stage and reverse transcriptase along with dNTPs were allowed to interact with the column at 10-20 μm/s for a defined amount of time.

The author chose 45 minutes as the time for cDNA synthesis. The manufacturer of the reverse transcriptase enzyme (Qiagen) recommended 1hr. for first strand synthesis by benchtop methods, but because of the smaller reaction volume on-chip, 45 minutes was reasoned to be more than sufficient for the reaction to go to completion. Results for the first experiments integrating the two step process of mRNA isolation/purification →

cDNA synthesis/purification and the three step process of cell capture/lysis → mRNA isolation/purification → cDNA synthesis/purification are shown in Fig 2.8. A positive signal was obtained for all cell and mRNA templates; and all negative controls confirmed the absence of DNA contamination. As with the mRNA isolation experiments, single cell resolution was achieved when probing for a portion of *GAPDH*.

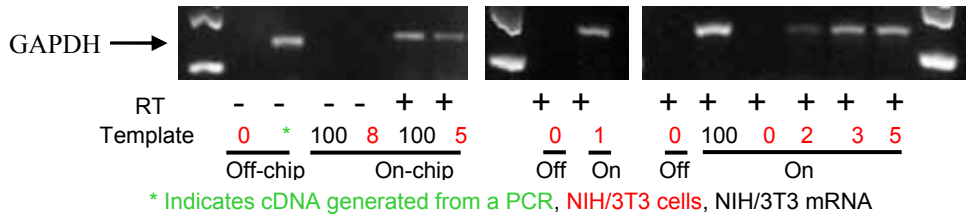


Figure 2.8. First strand cDNA synthesis on-chip. Agarose gel electrophoresis after PCR for microfluidic cDNA synthesis samples. Black signifies NIH/3T3 mRNA templates (pg) and red are NIH/3T3 cell templates. The green star is cDNA generated from an off-chip PCR.

Chapter 3:

Microfluidic Single Cell mRNA Isolation and Analysis

Microfluidic Single-Cell mRNA Isolation and Analysis

Joshua S. Marcus,[†,‡] W. French Anderson,[‡,§] and Stephen R. Quake*[,‡,∥]

Biochemistry and Molecular Biophysics, and Applied Physics and Physics, California Institute of Technology, MS 128-95, Pasadena, California 91125, and Gene Therapy Laboratories, Keck School of Medicine, University of Southern California, Los Angeles, California 90033

Single-cell gene expression analysis holds great promise for studying diverse biological systems, but methodology to process these precious samples in a reproducible, quantitative, and parallel fashion remains challenging. Here, we utilize microfluidics to isolate picogram and subpicogram mRNA templates, as well as to synthesize cDNA from these templates. We demonstrate single-cell mRNA isolation and cDNA synthesis, provide quantitative calibrations for each step in the process, and measure gene expression in individual cells. The techniques presented here form the foundation for highly parallel single-cell gene expression studies.

It is in general challenging to study mRNA levels in single cells. Techniques that are commonplace for population studies, such as differential display, subtractive cloning, and serial analysis of gene expression, are not yet amenable to the study of individual cells.[1] Currently, the most widely used assay for obtaining information about specific gene expression levels is cDNA synthesis from a cellular mRNA template using the RNA-dependent DNA polymerase reverse transcriptase (RT) coupled to the polymerase chain reaction (PCR).[2] This technique, analyzed either with end point gel electrophoresis or with real-time fluorescence, has in a few instances, been applied to single cells.[3−7] Probing specific gene expression with fluorescence in situ hybridization has also been accomplished in fixed cells by utilizing computational fluorescence microscopy.[8] A few groups have applied RT-PCR and T7 mRNA amplification approaches to prepare cDNA for single-cell microarray analysis.[9−12] Although these researchers were able to ask questions about global gene expression in single cells, each group was limited in their ability to process a large number of cells. For example, in one study,[12] 90 cells were isolated by laser capture microdissection, but only 16 exhibited robust expression of the ubiquitous transcript glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Single-cell cDNA library construction has been utilized sparingly and is limited to medium- and high-abundance transcripts.[13−16]

The paucity of single-cell RT-PCR studies has been attributed to the laborious and difficult task of purifying mRNA from individual cells and subsequently synthesizing and purifying total single-cell cDNA.[17] The difficulty is due to the loss of material during the subsequent steps of cell isolation, lysis, mRNA isolation, and cDNA synthesis. Reasons for the loss include mRNA degradation due to RNases or damage, nonspecific adhesion to the reaction vessel, and reverse transcription not going to completion.

Because of the small reaction volumes and the ability for automation, microfluidic devices are the ideal platform to analyze single cells in a high-throughput, highly parallel fashion. Previous work in our laboratory demonstrated the feasibility of using microfluidic devices with integrated micromechanical valves to isolate mRNA and genomic DNA from small numbers of mammalian and bacterial cells, respectively.[18] In this report, we implement a more robust isolation and purification scheme, provide the first quantitative calibrations for microfluidic mRNA isolation and cDNA synthesis, and demonstrate sensitivity sufficient to detect medium and low copy number transcripts in single cells. All five steps (cell capture, cell lysis, mRNA purification, cDNA synthesis, cDNA purification) were implemented in a microfluidic assay on one integrated device.

* To whom correspondence should be addressed. E-mail: quake@stanford.edu.
† Biochemistry and Molecular Biophysics, California Institute of Technology.
‡ Applied Physics and Physics, California Institute of Technology.
§ University of Southern California.
∥ Current address: Bioengineering Department, Stanford University, Clark Center E350Q, Stanford, CA 94305.

(1) Li, B. *Electron. J. Biotechnol.* **2005**, *8*, 71−81.
(2) Bustin, S. A. *J. Mol. Endocrinol.* **2000**, *25*, 169−193.
(3) Bengtsson, M.; Stahlberg, A.; Rorsman, P.; Kubista, M. *Genome Res.* **2005**, *15*, 1388−1392.
(4) Dalik, M.; Puma, C.; Quirion, R.; Williams, S. *J. Neurosci. Res.* **2003**, *74*, 286−295.
(5) Henne, J.; Pottering, S.; Jeserich, G. *J. Neurosci. Res.* **2000**, *62*, 629−637.
(6) Lindqvist, N.; Vidal-Sanz, M.; Hallbook, F. *Brain Res. Brain Res. Protoc.* **2002**, *10*, 75−83.
(7) Peixoto, A.; Monteiro, M.; Rocha, B.; Veiga-Fernandes, H. *Genome Res.* **2004**, *14*, 1938−1947.
(8) Levsky, J. M.; Shenoy, S. M.; Pezo, R. C.; Singer, R. H. *Science* **2002**, *297*, 836−840.
(9) Chiang, M. K.; Melton, D. A. *Dev. Cell* **2003**, *4*, 383−393.
(10) Kamme, F.; Salunga, R.; Yu, J.; Tran, D. T.; Zhu, J.; Luo, L.; Bittner, A.; Guo, H. Q.; Miller, N.; Wan, J.; Erlander, M. *J. Neurosci.* **2003**, *23*, 3607−3615.
(11) Luo, L.; Salunga, R. C.; Guo, H.; Bittner, A.; Joy, K. C.; Galindo, J. E.; Xiao, H.; Rogers, K. E.; Wan, J. S.; Jackson, M. R.; Erlander, M. G. *Nat. Med.* **1999**, *5*, 117−122.
(12) Tietjen, I.; Rihel, J. M.; Cao, Y.; Koentges, G.; Zakhary, L.; Dulac, C. *Neuron* **2003**, *38*, 161−175.
(13) Bishop, J. O.; Morton, J. G.; Rosbash, M.; Richardson, M. *Nature* **1974**, *250*, 199−204.
(14) Dulac, C.; Axel, R. *Cell* **1995**, *83*, 195−206.
(15) Matsunami, H.; Buck, L. B. *Cell* **1997**, *90*, 775−784.
(16) Pantages, E.; Dulac, C. *Neuron* **2000**, *28*, 835−845.
(17) Liss, B. *Nucleic Acids Res.* **2002**, *30*, e89.
(18) Hong, J. W.; Studer, V.; Hang, G.; Anderson, W. F.; Quake, S. R. *Nat. Biotechnol.* **2004**, *22*, 435−439.

10.1021/ac0519460 CCC: $33.50
Published on Web 03/31/2006

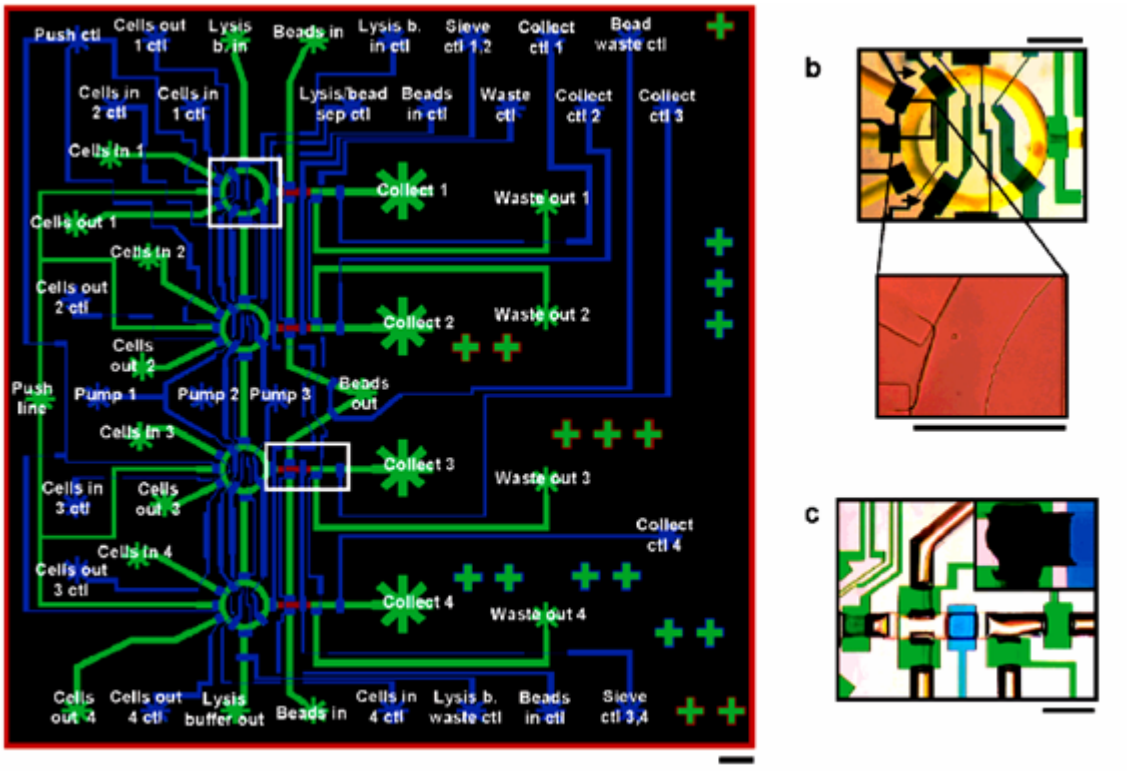


Figure 1. 4plex mRNA isolation/first strand synthesis device. (a) AutoCAD drawing of a device with inputs and outputs labeled according to function. Rounded flow channels are depicted in green and control channels are shown in blue. Unrounded (rectangular profile) flow channels for affinity column construction are shown in red. Portions of drawing in white boxes are shown in (b) and (c), respectively. (b) Optical micrographs of the lysis ring and an NIH/3T3 cell captured in the ring. (c) Optical micrographs of the affinity column construction area and a stacked column. Scale bars are 400 μm.

EXPERIMENTAL SECTION

Device Fabrication. All devices were fabricated using the process of multilayer soft lithography.[19] Devices are composed of three layers of the silicone elastomer poly(dimethylsiloxane) (General Electric) bonded to a RCA cleaned No. 1.5 glass coverslip and were fabricated as previously described,[20] but with slight modifications.[21] Negative master molds were fabricated out of photoresist by standard optical lithography and patterned with 20 000 dpi transparency masks (CAD/Art Services) drafted with AutoCAD software (Autodesk). The flow layer masks were sized to 101.5% of the control layer masks to compensate for shrinking of features during the first elastomer curing step. The flow master molds were fabricated out of 40-μm AZ-100XT/13-μm SU8-2015 photoresists (Clariant/Microchem), and the control molds were cast from 24-μm SU8-2025 (Microchem). The procedures for master mold and device fabrication are described in detail in the Supporting Information.

General Device Operation. The on−off valves within each device are controlled by individual pressure sources (Fluidigm) and are interfaced via 23-gauge pins (New England Small Tube) and Tygon tubing (VWR). A NI-DAQ card and Labview interface (National Instruments) were used to actuate the pressure sources.

Reagent Preparation. Lysis and wash buffers are brought to room temperature before utilization. Paramagnetic beads derivatized with oligo $(dT)_{25}$ sequences (Dynal Biotech) are suspended in lysis buffer to 3× the manufacturer's concentration. The first strand synthesis reaction is prepared according to the manufacturer's protocol (Qiagen), with enzymes left out until the reaction is ready to be performed on-chip. PCR collection buffer is prepared by bringing the respective PCR buffer to 1× concentration in H_2O.

Microfluidic mRNA Isolation and cDNA Synthesis Process Flow. Nucleic acid processing was performed in one of two ways: single reactions in a serial manner or four reactions in parallel with the 4plex device. An autoCAD drawing of a 4plex device, as well as optical micrographs of sections of a device are pictured in Figure 1. The devices contain modules for (1) cell lysis (Figure 1b), (2) mRNA isolation/cDNA synthesis/purification (Figure 1c), and (3) product collection. We performed single isolations serially for initial proof-of-principle studies, utilizing gel electrophoresis for analysis. The samples processed in parallel were analyzed with qPCR.

Lysis buffer is loaded into all flow lines in the device through the specified inlet (Figure 1a), except for the lines utilized for

(19) Unger, M. A.; Chou, H. P.; Thorsen, T.; Scherer, A.; Quake, S. R. *Science* **2000**, *288*, 113−116.
(20) Fu, A. Y.; Spence, C.; Scherer, A.; Arnold, F. H.; Quake, S. R. *Nat. Biotechnol.* **1999**, *17*, 1109−1111.
(21) Studer, V. H., G.; Pandolfi, A.; Ortiz, M.; Anderson, W. F.; Quake, S. R. *J. Appl. Phys.* **2004**, *95*, 393−398.

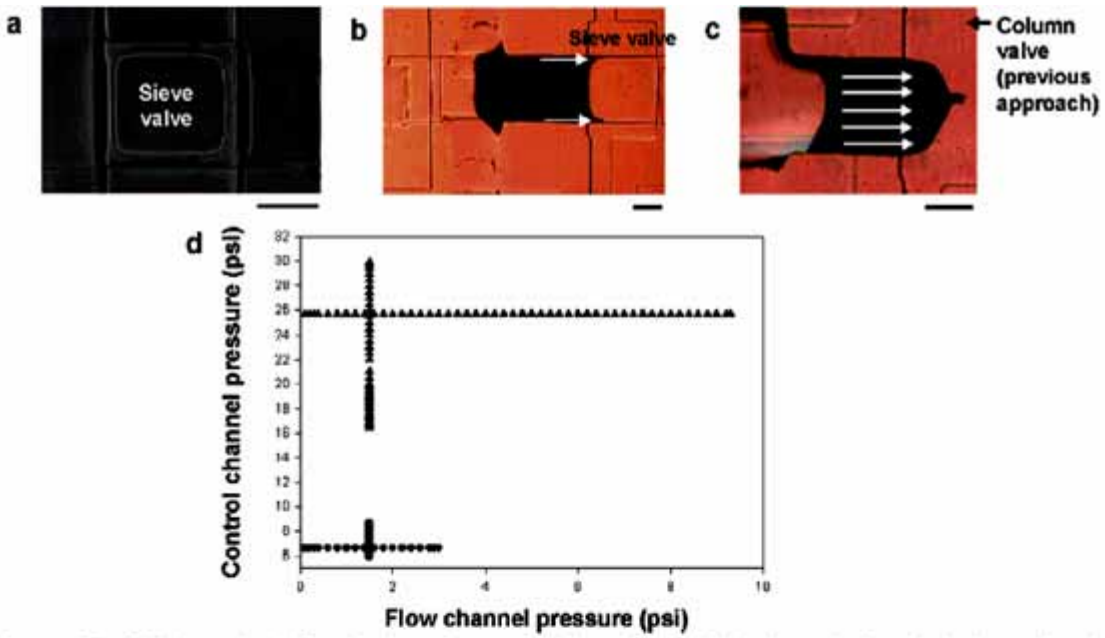


Figure 2. Microfluidic sieve valve. (a) Top-down image of an actuated sieve valve. (b) Optical micrograph of a stacked column using a sieve valve. Arrows indicate where beads have the capacity to escape. (c) Optical micrograph of a stacked column utilizing the previous approach. Arrows indicate where beads have the potential to escape. Scale bars are 100 μm. (d) Sampling of parameter space sufficient to build functional bead columns. Triangles represent the sieve valve method, and circles represent the previous approach.[18] The flow channel pressure is the pressure applied to the bead inlet and the column pressure is the pressure applied to the column valve's inlet.

cell loading and capture. This is accomplished by keeping pump valves 1 and 2 closed (Figure 1a). The beads are then loaded onto the chip in a north−south south−north fashion (Figure 1a), depending on which reactor is used, and subsequently stacked against sieve valves (Figures 1c and 2).[22] When loading the beads, the waste valves are open and the collection and lysis ring/bead separation control channels are closed, so as not to contaminate other areas of the device with beads (Figure 1a, c).

Sieve valves are present on a portion of a 200-μm-wide, 10-μm-high flow channel with a rectangular profile. A top-down view of a closed sieve valve (Figure 2a, b) shows when the valve is actuated, flow is restricted to the top and bottom ~10 μm of the channel. Previous microfluidic column designs in our group used a slightly opened valve present on a flow channel with a semicircular profile to allow fluid flow, while not allowing beads to pass (Figure 2c).[18] A sampling of parameter space sufficient to build functional bead columns by the two methods is given in Figure 2d. We either applied 1.5 psi pneumatic pressure to the bead inlet while varying the column valve's pressure or kept the column valve's pressure constant while varying the pressure applied to the flow inlet. Using the sieve approach, flow pressure can be varied by 1 order of magnitude more than the first-generation column, measured by whether the beads escape to waste (Figure 2d). Similarly, when applying constant pressure to the bead inlet, the pressure applied to the sieve valve used to stack the beads can be adjusted 7-fold more than its first-generation counterpart. Because of the sieve valve's functionality over a breadth of parameter space, it enables columns to be constructed in a digital manner. This is contrasted to its analog predecessor,[18] where pressures needed to be finely tuned in order to obtain a functional column. The digital and analog nature of the respective columns is evident when visualizing the flow profile through stacked columns, and where in the channel beads can potentially escape (arrows in Figure 2a and b). If the applied flow pressure over a column built by the previous method becomes too great, beads can escape at any point in the channel, whereas beads can only escape at the top and bottom portions (top-down view) of the channel if a column is built against a sieve valve.

After affinity column construction, NIH/3T3 cells or purified NIH/3T3 mRNA (Ambion) are loaded into a 4-nL portion of the lysis ring between the valves for pumps 1 and 2 (Figure 1a, b). The contents of the ring are subsequently mixed[23,24] and pressure-driven over the affinity columns. Once mRNA capture is complete, a wash step is implemented. Then, the bead−mRNA complexes are collected by opening the sieve valve, closing the waste lines, and sending the beads to the output ports in RT-PCR buffer. Alternatively, solid-phase cDNA synthesis is performed directly on the affinity columns. To do this, the columns are flushed with RT and dNTPs in a first strand reaction buffer for 45 min, followed by heating the chips to 40°C on a thermal microscope stage to activate the polymerase. The dT sequences on the beads are used as primers. When cDNA synthesis is complete, the bead−cDNA complexes are sent to the output ports in PCR buffer, collected with gel loading tips (Promega), and frozen, pending future analysis.

(22) Sieve valve designed in collaboration with Carl Hansen.

(23) Chou, H. P.; Unger, M. A.; Quake, S. R. *Biomed. Microdevices* **2001**, *3*, 323−330.

(24) Hansen, C. L.; Sommer, M. O.; Quake, S. R. *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 14431−14436.

NIH/3T3 Cell Preparation. NIH/3T3 cells were grown to ~70% confluency in Dulbecco's modified Eagle medium/10% calf serum (Hyclone) in 12.5-cm² vacuum-gas plasma-treated polystyrene culture flasks (BD Falcon). Cells were then washed in 1× PBS, trypsinized, and further diluted in culture medium to stop the trypsinization reaction. After counting the cells, the cell suspension was washed in 1× PBS. The cells in the suspension were then pelleted at 2040*g* for 5 min and suspended in 1× PBS to a concentration of 10^6 cells/mL.

Conventional and Real-Time qPCR. A 132-bp portion of GAPDH was amplified from either the mRNA or 1st strand cDNA template using the following primers: GAPDH sense 5′-CCTGGAGAAACCTGCCAAGTATG-3′; GAPDH antisense 5′-AGAGTGGGAGTTGCTTTGAAGTC-3′.

A 294-bp portion of HPRT was amplified with the following primers: hypoxanthine phosphoribosyl transferase (HPRT) sense 5′-GCTGGTGAAAAGGACCTCT-3′; HPRT antisense 5′-CACAGGACTAGAACACCTGC-3′.

For real-time qPCR cycling, SYBR green was used as a probe in a Roche Lightcycler 1.5 according to the following thermal cycling protocol: 95 °C for 15 min and 35 cycles of 94 °C for 15 s, 55 °C for 20 s, and 72 °C for 15 s. Detection was enabled at 79 °C for GAPDH and 72 °C for HPRT. Postrun melt curve analysis confirmed all products formed were full length and not primer dimers.

Conventional thermal cycling was carried out according to the following protocol: 95 °C for 15 min and 30 cycles of 94 °C for 30 s, 56 °C for 1 min, and 72 °C for 1 min, followed by a final extension phase at 72 °C for 10 min. PCR products were visualized by SYBR green fluorescence on 1% agarose gels.

RESULTS AND DISCUSSION

Microfluidic mRNA Isolation. We first studied the efficiency and reproducibility of the microfluidic mRNA affinity purification step with purified mRNA from NIH/3T3 cells. These samples were loaded onto the chip shown in Figure 1, captured on bead columns, followed by product−bead recovery and analysis with benchtop RT-qPCR. We measured the medium-abundance mRNA GAPDH (~1800 copies/cell, Figure S-1a, Supporting Information) and the low copy number mRNA HPRT (5−10 copies/cell[25]). For the two linear regression lines fitted to the threshold cycles (Ct) of the on-chip standards, R values of 0.98−0.99 were obtained (Figure S-1b). The measured sensitivity and dynamic range for GAPDH probed samples are 0.3−100 pg of NIH/3T3 mRNA, suitable for assays on 1−100 cells (Figure 3). We achieved an average isolation/recovery efficiency of 80% for the samples interrogated for GAPDH (Figure 3). For the purposes of this discussion, 100% efficiency is defined as what one achieves from an RT-qPCR utilizing purified templates subjected to no isolation procedure (e.g., a benchtop standard curve with defined mRNA standards). We were also able to detect HPRT from NIH/3T3 mRNA templates ranging from 1 to 100 pg, both on- and off-chip (Figure S-1b, c), but observed that the amplification efficiency for the on-chip samples differed from that of the off-chip samples. This complicates efforts[26] to calculate a reliable isolation/recovery efficiency using HPRT as a marker. At this point, we are only able

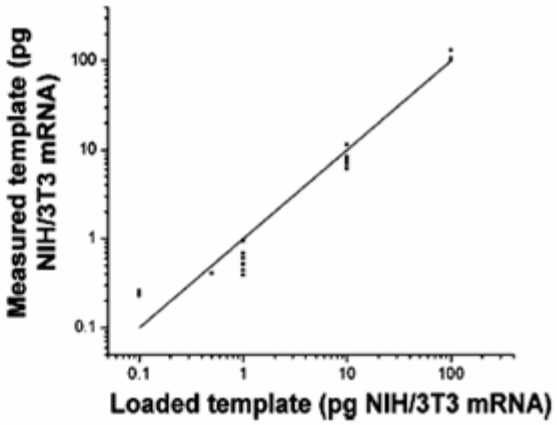


Figure 3. 4plex mRNA isolation calibration and efficiency measurements. A 132-bp portion of GAPDH was amplified and detected. Measured values for 4plex microfluidic standards were extrapolated utilizing an off-chip standard curve (Figure S-2b) generated with known quantities of NIH/3T3 mRNA. The line represents 100% efficiency ($m = 1$).

to determine relative amounts of HPRT in a sample, without absolute quantification. The coefficient of deviation (experimental noise) for the GAPDH efficiency measurements was 14%. This noise can be attributed mainly to pipetting error in the downstream qRT-PCR reaction setups, different reagent lots for the on- and off-chip standard curve experiments, and loss of product during recovery from the devices.

Microfluidic cDNA Synthesis. Solid-phase cDNA synthesis was implemented on-chip with purified NIH/3T3 mRNA by utilizing the beads as both primers and a support. Gel electrophoresis results from PCR on templates processed microfluidically are shown in Figure 4a. When the reverse transcriptase enzyme was withheld as a negative control, no signal was detected by PCR with 100 pg of NIH/3T3 mRNA or with eight NIH/3T3 cells, demonstrating that DNA contamination within our devices is not an issue. As with the mRNA isolation process, single-cell sensitivity was realized. To calibrate the cDNA synthesis step, we generated a qPCR standard curve fitted to Ct values of known NIH/3T3 mRNA templates subjected to the solid-support first strand cDNA synthesis on the 4plex platform (Figure S-1d). The qPCR standard curve generated from microfluidic samples spans 4 orders of magnitude with a highly reliable correlation coefficient ($R = 0.99$). As with the mRNA isolation procedure, we extrapolated measured values for our known on-chip templates by plotting the Ct values onto an off-chip RT-qPCR standard curve (primers specific to GAPDH, Figure 4b). We obtained an average efficiency of 44% for the coupled mRNA isolation/cDNA synthesis/recovery steps. Because we performed total cDNA synthesis with oligo(dT) primers and then utilized specific primers for the PCR, the on-chip measurements are less efficient than if we were to introduce specific primers into the first strand reaction. The average experimental variability (coefficient of deviation) for the coupled mRNA isolation/cDNA synthesis steps was 19%. This noise can be attributed to the same reasons that were found for the mRNA isolation variability, as well as the aforementioned differences in primers.

Cellular Measurements. Next, we sought to study gene expression in small numbers of NIH/3T3 cells. We first utilized

(25) Pannetier, C.; Delassus, S.; Darche, S.; Saucier, C.; Kourilsky, P. *Nucleic Acids Res.* **1993**, *21*, 577−583.
(26) Rutledge, R. G.; Cote, C. *Nucleic Acids Res.* **2003**, *31*, e93.

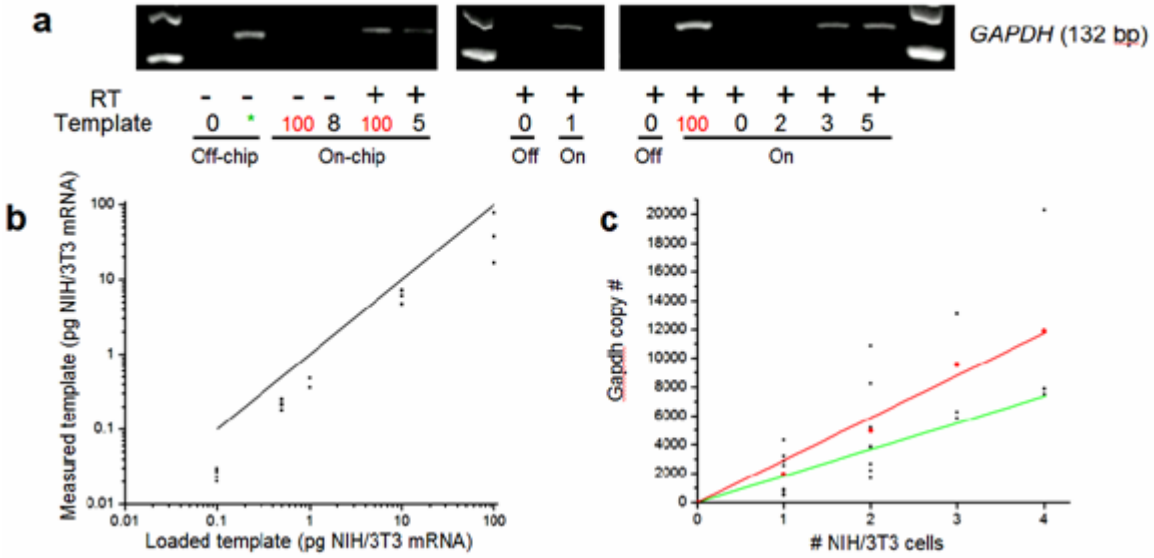


Figure 4. On-chip cDNA synthesis calibration and gene expression measurements in various numbers of cells. (a) Gel electrophoresis results from samples subjected to on-chip cDNA synthesis. Each lane on the 1% agarose gel contains 1 μL of the respective 50-μL PCR reaction. The black and red numbers are number of NIH/3T3 cells and NIH/3T3 mRNA templates in picograms, respectively. The * represents 1 μL of unpurified GAPDH PCR product, utilized as a positive control. (b) 4plex mRNA isolation/cDNA synthesis efficiency data. An off-chip qRT-PCR standard curve (Figure S-2c) was utilized to extrapolate measured values for loaded microfluidic templates (black squares). The line represents 100% efficiency ($m = 1$). (c) Varying levels of GAPDH gene expression in individual cells. The average copy number per NIH/3T3 cell obtained from a bulk experiment was plotted as a green line, $Y = 1836X$. The microfluidic samples were similarly run against a standard curve of known quantities of GAPDH cDNA (Figure S-2a, red) to determine copy number for each sample (black circles). The average copy number for each sample group, adjusted for efficiency of the process, are plotted as red circles, fitted to $Y = 2938X$, $R^2 = 0.98$.

4plex devices to carry out the process of cell capture–cell lysis–mRNA isolation on individual cells ($n = 5$). By comparing the Cts of the NIH/3T3 cellular samples to the off-chip NIH/3T3 mRNA standard curve (Figure S-1c), we obtained a raw estimate of 0.8 pg of mRNA/cell. Upon correcting for the 80% measured efficiency of the process, we estimate an average value of 1 pg mRNA/cell, consistent with published values[27] and slightly better than most commercial kits (Table S-1, Supporting Information).

We then carried out the five-step microfluidic process (cell capture–lysis–mRNA isolation–cDNA synthesis–purification) on various populations of cells, ranging from one to four cells. We subsequently ran off-chip GAPDH cDNA standards (Figure S-1a) to calibrate the GAPDH copy number in each cellular sample (Figure 4c, black circles), as well as to obtain an average copy number for each sample size (Figure 4c, red circles). Based on the raw GAPDH copy number of 1293 copies/cell obtained in the microfluidic cell samples, and after accounting for the 44% efficiency of the method, we estimate an average of 2938 ± 140 GAPDH transcripts per cell. This result is within a factor of 2 of the average of 1837 ± 146 transcripts obtained from bulk measurements (Figure 4c, green line). The bulk average was determined by running mRNA templates isolated from 2500 NIH/3T3 cells against a GAPDH cDNA standard curve (Figure S-1a). For the single-cell samples, the cell-to-cell variation in copy number is rather large, with a coefficient of variation of 0.75 ($n = 8$). This is significantly greater than the random error introduced by processing, which we estimate at 0.19 ($n = 25$). These differences can be most likely attributed to inherent cell-to-cell variation, asynchronization in the cell cycle, and varying metabolic states.

A key feature of the microfluidic process is the ability to generate and recover concentrated mRNA or cDNA products. Because the beads are paramagnetic and therefore can be reconcentrated, they can be flushed out of the device in microliter amounts of buffer without irreversibly diluting the products. The product concentration after mRNA isolation/first strand cDNA synthesis is crucial for the integration of downstream on-chip steps and off-chip validation assays. A byproduct of the column's broad functionality is the versatility gained by the device user in choosing whether to make cDNA after mRNA isolation is complete. Furthermore, the present strategy to build affinity columns shows marked improvement in speed and robustness over the previous implementation[18] (Figure 2d).

The mRNA isolation and cDNA synthesis calibrations show that this technique is reliable for templates that span 4 orders of magnitude in copy number. Notably, we were able to detect the presence of GAPDH and HPRT mRNAs from initial templates of 0.1 and 1 pg of mRNA, respectively, the former being roughly 10–20% of a single cell's mRNA. These results show promise for the detection of low-abundance transcription factors present in single cells. We were also able to show that cell lysis did not have an abnormal effect on subsequent mRNA capture. The mRNA content of single cells was determined by qRT-PCR to fall within 0.5–2.5 pg (Table S-1), values consistent with published values.[27] We also demonstrated the significance of studying single cells as opposed to bulk samples. Figure 4c makes apparent the discrepancy in information attained when studying single cells versus

(27) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 4 ed.; Garland Publishing: New York, 1994.

populations. This approach could have far-reaching implications for gene expression studies in developmental and cancer biology, where individual cells within a population succumb to different fates. Information pertaining to what genes are expressed in low to medium abundance in these individual cells would be impossible to gain from bulk studies that give average values over populations.

The ability to acquire gene expression data from multiple cells, as well as single cells, should not be overlooked. In many cases, such as intercellular signaling studies, it may be of interest to study gene expression in small numbers of cells rather than individual cellular samples. Our technique and device has the dynamic range to support both types of assays. Furthermore, based on the mRNA isolation and cDNA synthesis calibration curves, it is not unreasonable to suspect our processing capabilities to be ≥ 100 cells/reactor, thus allowing automated preparation of larger samples.

CONCLUSIONS

The device and methodology presented here lay the groundwork for highly parallel economy-of-scale single-cell gene expression analysis. Although we limited the process to four reactions in parallel, the notion of realizing 50−500 reactions on one device is not far in the future. The possible number of samples one can attain with this technology is at least 1 order of magnitude larger than the current state of the art in single-cell gene expression studies. We also demonstrated the method to be a robust one, in that sample concentrations ranging 4 orders of magnitude could be retrieved and detected in a quantitative manner.

ACKNOWLEDGMENT

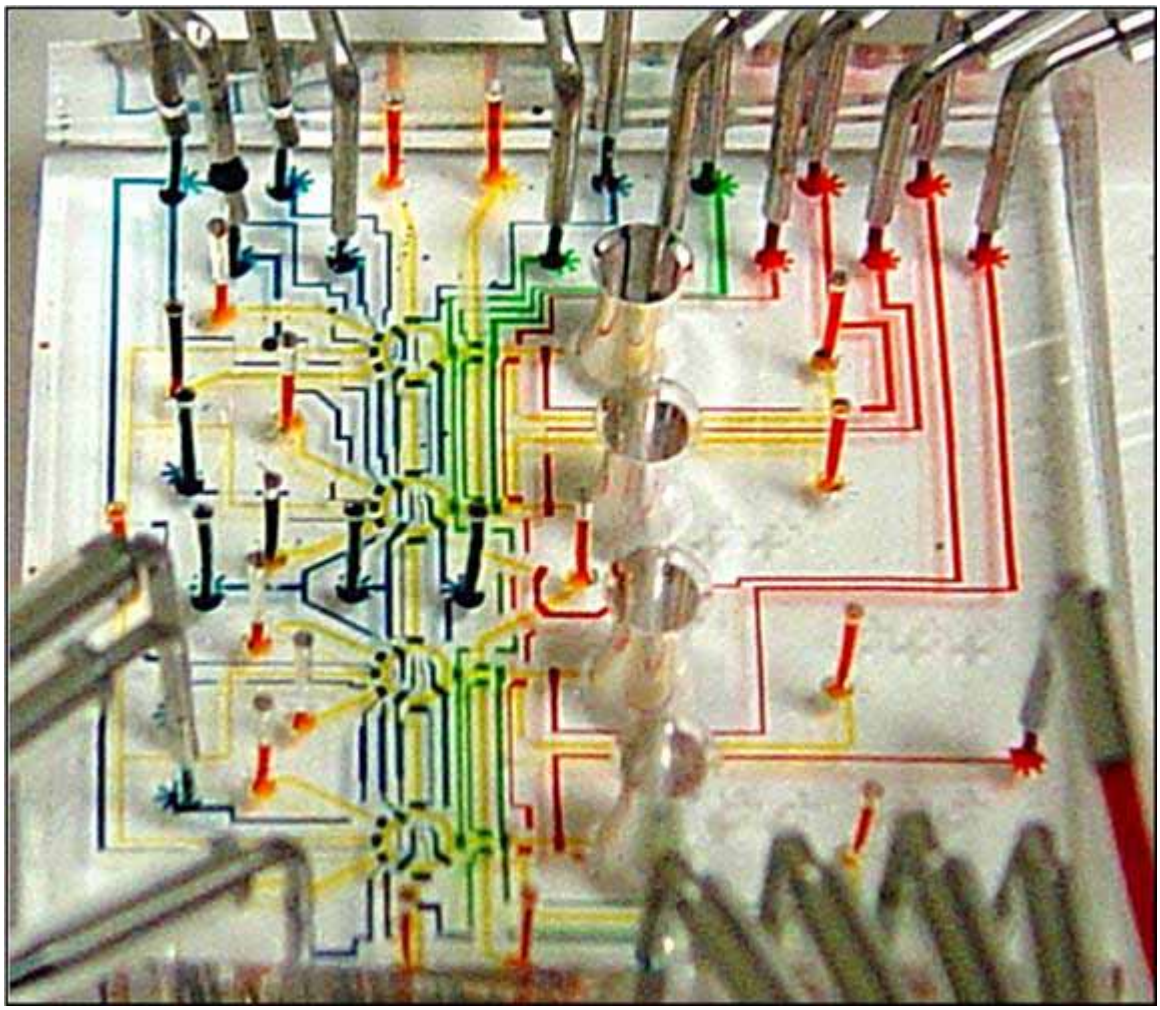
We thank Kathy Burke for assistance with cell culture and real-time PCR. We also acknowledge Alejandra Torres for assistance with device fabrication and Carl Hansen and Sebastian Maerkl for helpful discussions. This work was supported by a National Research Service Award (T32GM07616) from the National Institute of General Medical Sciences (J.S.M.) and National Institutes of Health (NIH) grant NIH 1RO1 HG002644-01A1.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review October 31, 2005. Accepted February 24, 2006.

AC0519460



Supporting information for:

Microfluidic single cell mRNA isolation and analysis

Joshua S. Marcus,[1,2] W. French Anderson,[2,3] & Stephen R. Quake[2,4,5]

[1]Biochemistry and Molecular Biophysics, [2]Applied Physics and Physics, California Institute of Technology, MS 128-95, Pasadena, CA 91125. [3]Gene Therapy Laboratories, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033

4: To whom correspondence should be addressed. Email quake@stanford.edu

5: Current address: Bioengineering Dept., Stanford University, Clark Center, E350Q, Stanford, CA 94305

Table of Contents

Supporting methods

Master Mold Fabrication

All photomasks, which define device features, were designed with AutoCAD software (AutoDesk, Sausalito, CA) and printed at a resolution of 20,000 dots per inch on transparency films. In all optical lithography processes, 3'' silicon wafers were utilized as substrates, and mold exposures were under UV light on an MJB mask aligner (7 mW/cm^2) for the indicated time.

Twenty-four-μm-high features present on control molds are fabricated with a single lithographic step. SU8-2025 (Microchem) is spun on a wafer (3,000 rpm, 45 s.), baked before exposure to evaporate excess solvent (2 min./5 min. at 65°C/95°C), exposed under a negative mask for 75 s., baked after exposure (2 min./5 min. at 65°C/95°C) to facilitate additional resist polymerization, and developed in NanoSU8 developer (Microchem). Once features are developed, the mold is baked again for 40 s. at 95°C to remove remaining solvent.

A two step lithographic process is utilized for fourplex flow mold fabrication. The first step is to define column construction flow channels (10 μm high). For these channels, SU8-2010 (Microchem) is spun onto a wafer (3,000 rpm, 45 s.), baked before exposure to evaporate excess solvent (1 min./3 min. at 65°C/95°C), exposed under a negative mask for 45 s., baked after exposure (1 min./3 min. at 65°C/95°C) to facilitate additional resist polymerization, and developed in NanoSU8 developer (Microchem). Upon visualization of developed features, the mold is subjected to a 90 min. hard baked at 150°C. When the mold cools to room temperature, the second step defining the remaining 40 μm high channels is carried out. First, to promote resist adhesion, the mold

is exposed to hexamethyldisilizane (HMDS) vapor for 2 min. AZ-50 (Clariant) is then spun onto the mold (1,600 rpm, 60 s.). The mold is subsequntly soft baked (2 min./5 min. at 65°C/115°C), aligned to the column construction flow channels, exposed under a positive mask for 4 min., and developed in 25% 2401 developer (Microchem, diluted with 18 mΩ dI H_2O). Once features are developed, the mold is annealed/hard baked for 3 hr. at 200°C for the aforementioned reasons. Additionally, if the AZ-50 resist does not undergo a chemical change (color turns from red to black), the features will crack when subjected to subsequent soft lithography baking steps.

Device Fabrication

All devices are fabricated by multilayer soft lithography (MSL) with the silicone elastomer polydimethylsiloxane (PDMS, General Electric). Each device employs push-up valve geometry and is a three layer elastomeric structure bonded to a 3'' X 1'' RCA-cleaned coverslip. Negative master molds are first exposed to chlortrimethylsilane (TMCS, Aldrich) vapor for 2 min. to promote elastomer release from molds after baking steps. Thirty g of liquid PDMS (5 parts A:1 part B) is poured onto the flow master, degassed under vacuum, and baked for 45 min. at 80°C. Liquid PDMS (20 parts A:1 part B) is spun onto the control master mold (2,000 rpm for 60 s.) and allowed to settle for 30 min. in order to obtain a uniform elastomer membrane on top of the control features. The mold is then baked for 30 min. at 80°C. Upon completion of the baking steps, the partially cured flow layer is peeled from its mold and 650 micron diameter flow channel access holes punched (Technical Innovations, part# CR0350255N20R4). The layer is then aligned to the partially cured control layer, still on the control master. The

two layer structure was then baked for 45 min. The third layer, a featureless elastomeric membrane, is fabricated by spinning liquid PDMS (20 parts A:1 part B) onto a clean silicon wafer (1,600 rpm for 60 s.) followed by baking for 30 min. at 80°C. Once baking is completed, the two layer structure is peeled from the control master, control channel access holes punched (Technical Innovations, part# CR0350255N20R4) and mounted onto the third partially cured PDMS layer. The three layer structure is then baked for 5 hr. at 80°C. The assembled three layer structure is then peeled from the clean silicon wafer, output holes punched (Technical Innovations, part #CR0830655N14R4), cut to size, and bonded to an RCA-cleaned coverslip and baked overnight at 80°C.

Figure S-1

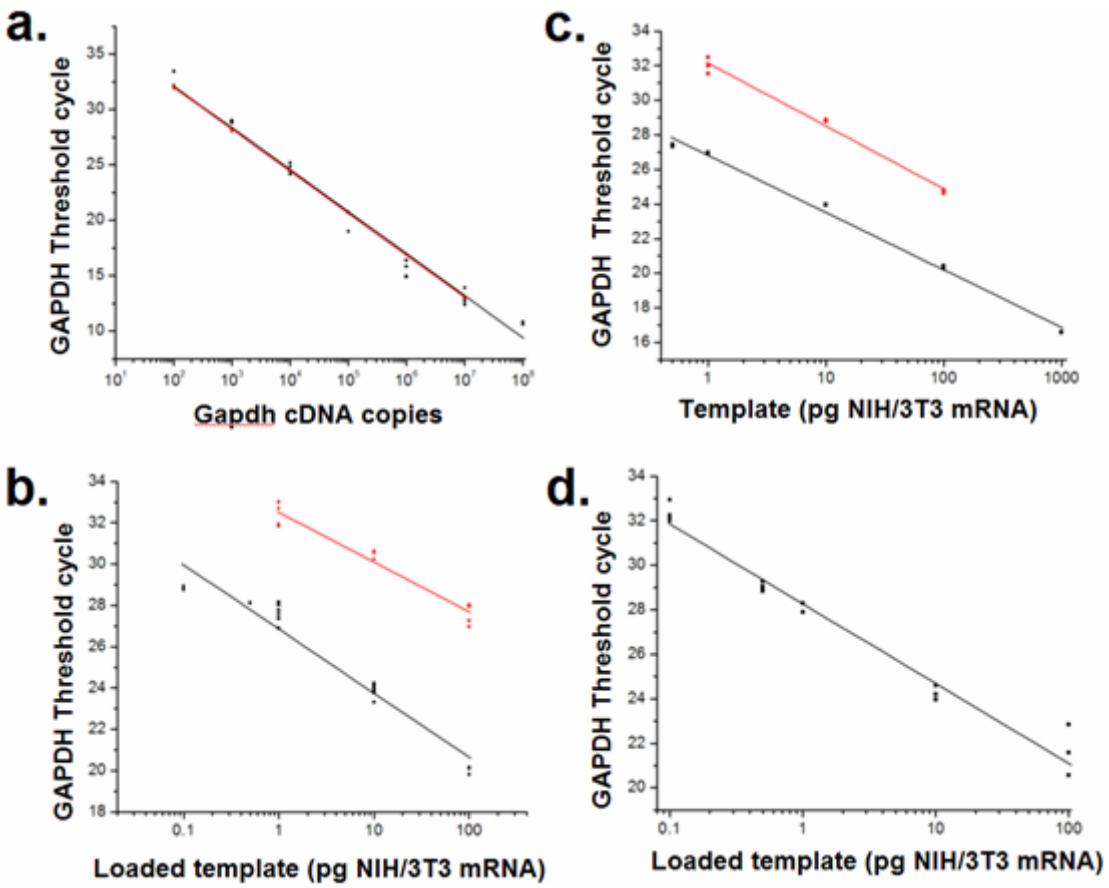

a.
GAPDH Threshold cycle
Gapdh cDNA copies
c.
GAPDH Threshold cycle
Template (pg NIH/3T3 mRNA)
b.
GAPDH Threshold cycle
Loaded template (pg NIH/3T3 mRNA)
d.
GAPDH Threshold cycle
Loaded template (pg NIH/3T3 mRNA)

Table S-1

Yields from various mRNA isolation procedures

Method	Total RNA yield per NIH/3T3 cell	mRNA yield per NIH/3T3 cell
Microfluidic fourplex	N/A	1pg
Dynal Bulk isolation	N/A	0.5pg
Promega PolyA Tract mRNA	N/A	0.33pg
Sigma GenElute Mammalian total	15pg	0.15pg-0.75pg*
Stratagene Absolutely RNA kit[3]	17.7pg	0.18pg-0.89pg*

*Range is 1-5% of total RNA

Supporting References

1. http://www.promega.com/pnotes/56/5338f/5338f_core.pdf

2. http://www.sigmaaldrich.com/sigma/bulletin/rtn10bul.pdf

3. http://www.stratagene.com/newsletter/pdf/13_3_p104-105.pdf

Supporting Figure Legends

Supporting Figure 1: qPCR and RT-qPCR standard curves. **a**. Standard curves utilized to extract GAPDH copy number from cell samples. Black: qPCR standard curve was generated with known GAPDH cDNA templates using the icycler platform ($Y = -3.78X + 39.67$. $r^2=0.99$). The regression line was utilized to extrapolate average GAPDH copy number in bulk preparations of NIH/3T3 mRNA by running RT-qPCR on seven mRNA samples, each extracted from 2,500 NIH/3T3 cells. Red: Results from the identical experiment in (a) utilizing the Roche Lightcycler ($Y = -3.79X + 39.57$. $r^2=0.99$). **b**. RT-qPCR standard curve generated with microfluidic mRNA samples (NIH/3T3 mRNA loaded on-chip) . A 132 bp portion of GAPDH (black, $Y = -3.11X + 26.86$. $r^2=0.95$), and a 294 bp portion of HPRT (red, $Y = -2.40X + 32.51$. $r^2=0.96$) were amplified and detected. **c**. RT-qPCR standard curves generated with NIH/3T3 mRNA. A 132 bp portion of GAPDH (black: $Y = -3.31X + 26.82$. $r^2=0.99$) or a 294 bp portion of HPRT (red: $Y = -3.63X + 32.16$. $r^2=0.99$) were amplified and detected. **d.** qPCR standard curve generated with microfluidic NIH/3T3 mRNA standards subjected to mRNA isolation/first strand cDNA synthesis. $Y = -3.59X + 28.27$. $r^2=0.98$.

Chapter 4: Real-time qPCR and oligonucleotide validation

4.1 Introduction

In order to achieve reliable quantification of mRNAs in biological samples, the quality and reproducibility of the system (e.g. microfluidic process and subsequent readout method) are of major importance.[32] To verify the microfluidic cells-to-cDNA process was quantitative, precise, and scalable (with regard to template concentration), the author constructed qPCR standard curves[9] with NIH/3T3 mRNA standards (Ambion) subjected to the isolation and recovery process presented in chapters 3 and 5. Process efficiency is extrapolated from stepwise process calibrations (ch. 3, 5) and is utilized as a correction factor for the single cell gene expression measurements presented in this thesis. Comparing standards along each step of the microfluidic process allows for the efficiency of each step to be resolved, as well as determination of any process efficiency differences that may exist at varying template concentrations. In this chapter, the amplification efficiency and reproducibility of qPCR will be discussed as it relates to multiple template concentrations and multiple genes, when probed for individually or in a multiplex fashion. Only after it is determined that reaction efficiencies with non-microfluidic standard templates within the dynamic range of the study are consistent and reproducible can qPCR be utilized for cellular measurements and the quantification of the microfluidic method itself.

In general, qPCR assays are characterized by a wide dynamic range of quantification, high sensitivity, and precision, making these assays a natural fit for validation of the microfluidic process.[9] Furthermore, the need for post-PCR steps is eliminated because detection of products occurs in real-time, thus avoiding the possibility

of cross-contamination when setting up a downstream assay. The majority of qPCR assays discussed in this chapter are gene specific, and take advantage of the 5' → 3' exonuclease activity of *Taq* polymerase, as well as fluorescence resonance energy transfer (FRET, Fig. 4.1). For each gene studied, three oligonucleotides (sense primer, antisense primer, dual-labeled fluorescent probe) were designed. The dual-labeled probe[33] binds a sequence upstream of the sense primer sequence, and downstream of the antisense primer sequence (or vice versa). Fluorescence of the reporter dye (5' end) is quenched by the second dye (3' end of probe) when the probe is intact. When *Taq* polymerase extends the PCR product off one of these primers, it cleaves the dual-labeled probe and eliminates fluorescence quenching caused by FRET (Fig. 4.1). Because reporter dyes become free in solution only when products are synthesized (unless DNases are present in reactions), the amount of fluorescence measured is directly proportional to amount of product generated.

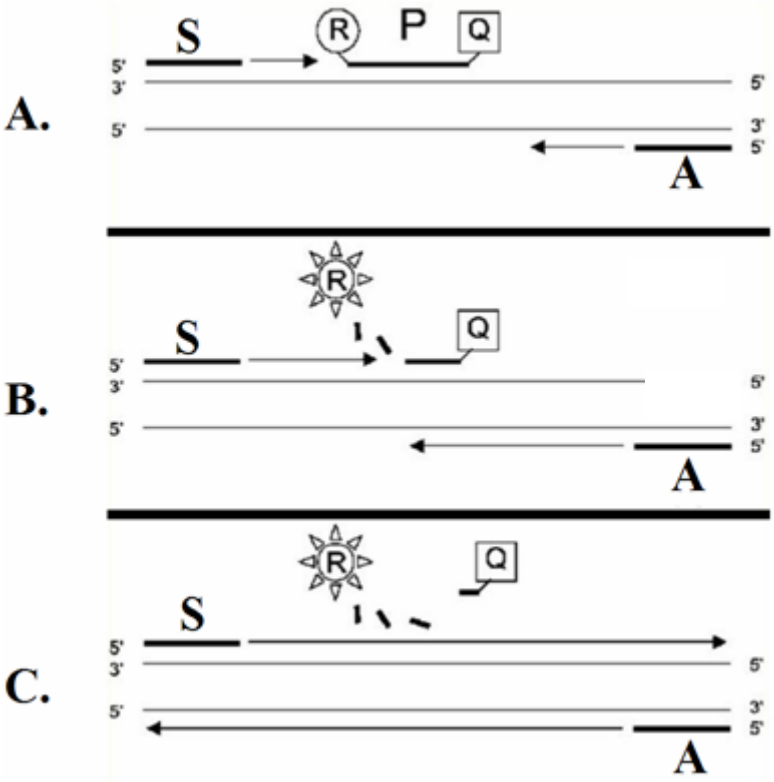


Figure 4.1. Taqman PCR scheme. S is sense primer, A is antisense primer, and P is dual-labeled probe. R and Q are the dual labeled probe's reporter and quencher dyes, respectively. **A.** Intitial state of the oligonucleotides before synthesis and extension of products. **B.** The product is extended off the sense primer and in doing so, the probe is cleaved. The reporter dye now free in solution is no longer quenched by the 3' dye. **C.** The probe is completely cleaved and the product has been fully extended. Quenching of the 5' reporter is completely eliminated.

The single cell qPCR literature is confined to studies that analyze cells harvested either by laser capture microdisection (LCM),[4, 34-36] patch-clamp,[8, 15, 37] or other micromanipulation methods.[1-4, 6, 7, 9-15, 37-46] These methods suffer from a lack of throughput and have been known to damage the integrity of the cells captured. Other groups[47, 48] have utilized aliquots from extremely dilute cell suspensions (which may or may not have a single cell) and claimed single cell resolution. Single cell multiplex PCR studies found in the literature all implement two successive rounds of PCR, with detection taking place at the end of the thermal cycling.[12, 46] Although the researchers claimed their methods maintained an accurate reflection of the starting material, it is reasonable to assume this would not be the case 100% of the time, or for all genes in a cell. Inter-reaction efficiency differences (causing different amounts of products to be generated in reactions with identical templates) can be caused by discrepancies in reagent concentrations and primer-dimer competition differences. These differences will vary from gene to gene depending on gene and primer characteristics such as secondary structure and melting temperature. For multiplex studies, in order to maintain an accurate

quantification, individual reaction efficiencies for each gene should be similar. Efficiency of a PCR on a given gene segment is stated as

$$P = T(1 + E)^n \quad (1)$$

where P is product after n cycles, T is starting template, and E is the reaction efficiency. E will vary from 0 to 1 (1 = 100% efficiency), with values of 0.7–0.9 being the most common. Because DNA generated from PCR increases exponentially after each cycle (Equation 1), small inter-reaction variability can cause dramatic differences in reagents consumed and therefore amount of product synthesized during a reaction. qPCR overcomes this problem by monitoring products generated after each PCR cycle with fluorescent probes that can either be sequence specific or non-specific. Quantification for each reaction takes place by determining the PCR cycle (threshold cycle or Ct) where fluorescence is significantly greater than background levels. The value typically corresponds to the first inflection point of an amplification curve. The log-linear relationship that exists between starting template concentration and Ct can then be utilized to quantify unknown samples. Because product is exponentially amplified in the reaction, end point analysis poses problems with quantifying the amount of mRNA present initially. A high efficiency reaction starting with a low copy number can reach the same product generation plateau as a low efficiency reaction that starts with a high copy number (Fig. 4.2, experimental observations).

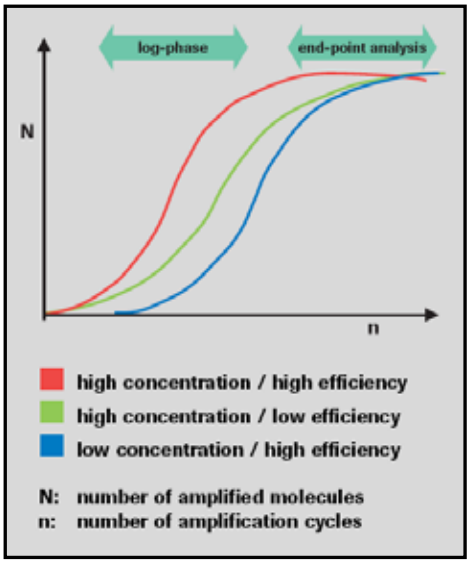


Figure 4.2. Real-time PCR amplification curves showing endpoint detection is not sufficient for quantitative analysis. Diagram copied from Roche Molecular Biochemicals technical note no. LC 8/99.

4.2 Bacterial mRNA primer and probe design

To obtain the sensitivity and dynamic range for the microfluidic cells-to-cDNA process, with respect to transcript copy number, the author employed a synthetic mRNA standard, specific to one transcript, and subjected it to the aforementioned microfluidic process, or portions of it (ch. 5). In order to quantify murine cell lysis and competition effects, the standard could not have homology to genes expressed in NIH/3T3 cells. Comparing Ct values for the spiked-in transcripts (Dap mRNA + cells processed to cDNA on-chip → qPCR) to Ct values for identically processed transcripts without cells present (Dap mRNA processed to cDNA on-chip → qPCR) gives a quantification of the cell lysis step. In addition, it is imperative to demonstrate equal priming and amplification efficiency for the synthetic standard oligonucleotides, as compared to the

efficiency of the oligonucleotides utilized in the murine study. The author designed primers for three of the four genes present in a synthetic polyA RNA control cocktail (Affymetrix) with Beacon Designer software.[49] Homology with mammalian genes was controlled for by BLASTing (Basic Local Alignment Search Tool)[50] the primer and probe sequences against mouse, human, and rat genomes. Primers and probes were designed to have T_ms of 60°C ± 3°C and T_m(primer) + 10°C, respectively. These values have been shown to be optimal for Taqman PCR assays.[1] Primer and probe specifications are given in Table 4.1 below.

Synthetic Bacterial mRNA primer & probe specifications

	Dap	**Lys**	**Thr**
Accession number	L38424	X17013	X04603
Sequence length (bases)	1,560	1,192	1,680
Probe sequence	TGT GAT GTG TAT TCC ATT CCG CTC GCC A	CCG AAA CCT CCT CCA AGA TTC AGC ACC T	TGC CTT TGC CAC AGC CAT AAC CAT TCC G
Probe position; length	908; 28	696; 28	97; 28
Probe Tm (°C); GC%	63.5; 50	63.6; 53.6	64.9; 53.6
Beacon Designer probe rating; quality	80.5; best	71.3; best	77.1; best
Sense primer sequence	ACC GGA TGT CTC GGC ATT AAT C	GGC CGG TTT TGT GTT AGC AG	CGG AAG GCG TCA ATC CTA CG
Sense primer position; length	880; 22	592; 20	33; 20
Sense primer Tm (°C); GC%	57.1; 50	57; 55	58; 60
Antisense primer sequence	GCA CAA GAA TTT CCG CAG TAC C	GCG GTT CAT CAT CTT CCG TAT AAC	CGC ACA TAA TCG TGT CAT TGC C
Antisense primer position; length	965; 22	725; 24	128; 22
Antisense primer Tm (°C); GC%	56.7; 50	55.7; 45.8	57; 50
Product length; produt Tm (°C)	86; 83.8	134; 85.9	96; 85.2

Table 4.1. Primer and probe specifications for three synthetic bacterial mRNAs.

Once the primers and probes were designed and synthesized (Integrated DNA Technologies), the author tested them against bacterial mRNA templates as well as NIH/3T3 mRNA in RT-qPCR reactions, to determine if the primer and probe sequences were unique to the synthetic bacterial mRNAs. A three step thermal cycling protocol was executed (Appendix A), with a combined annealing/extension step at 57°C for 1 min. Data collection was also enabled at this step. Amplification plots for the three genes (100,000 copies respective bacterial mRNA/reaction, Fig. 4.3A) show the primers and probe designed for Dap are the most reliable, giving strong amplification and a Ct one cycle less than its tested counterparts for the same template. This implies either better amplification efficiency or amplification of two products in parallel. To test for the latter, the author utilized the Dap oligonucleotide set and performed Sybr green qPCR followed by post run melt curve analysis (protocol given in Appendix A) on a dilution series of bacterial mRNA templates. It should be noted that since the FAM reporter (utilized for Dap probe) exhibits similar fluorescence excitation and emission properties to those of Sybr green, the Dap probe needed to be resynthesized sans fluorophores. Figure 4.3B illustrates the results of this experiment. The presence of one distinct peak for the template dilution series verifies only one product is formed during PCR and further validates the functionality of the Dap primer set. Furthermore, preliminary results showed no cross homology with NIH/3T3 mRNA for all primer sets tested (Fig. 4.3A). The author verified the Dap results in Fig. 4.3A by probing for Dap mRNA against a full dilution series of NIH/3T3 mRNA in an RT-qPCR (Fig. 4.3C). The absence of non-specific amplification was verified over the full range of NIH/3T3 templates tested (Fig. 4.3C).

Finally, the author needed to demonstrate equal priming and amplification efficiencies for the Dap product as compared to the NIH/3T3 products, in order to reliably measure absolute copies of mRNA present in single NIH/3T3 cells. Here, RT-qPCR purified products generated with the primer and probe sets utilized in chapter 5 (Dap–Table 4.1, β-actin, GAPDH, PGK–Table 4.2) were utilized to generate standard curves by qPCR (Dap) and triplex qPCR (β-actin, GAPDH, PGK). Primer specifications for the multiplex study are discussed later in the chapter. In order to create a template for triplex qPCR, so as to simulate the single cell measurements taken in chapter 5, the mammalian gene products were combined into one tube. The four standard curves generated are shown in Fig. 4.3D. The overlapping nature of the curves demonstrates equal amplification efficiencies for templates spanning 6 orders of magnitude, and allows the microfluidic process efficiency measurements generated with the synthetic mRNA Dap to be transferred to the study of the three aforementioned genes in NIH/3T3 cells.

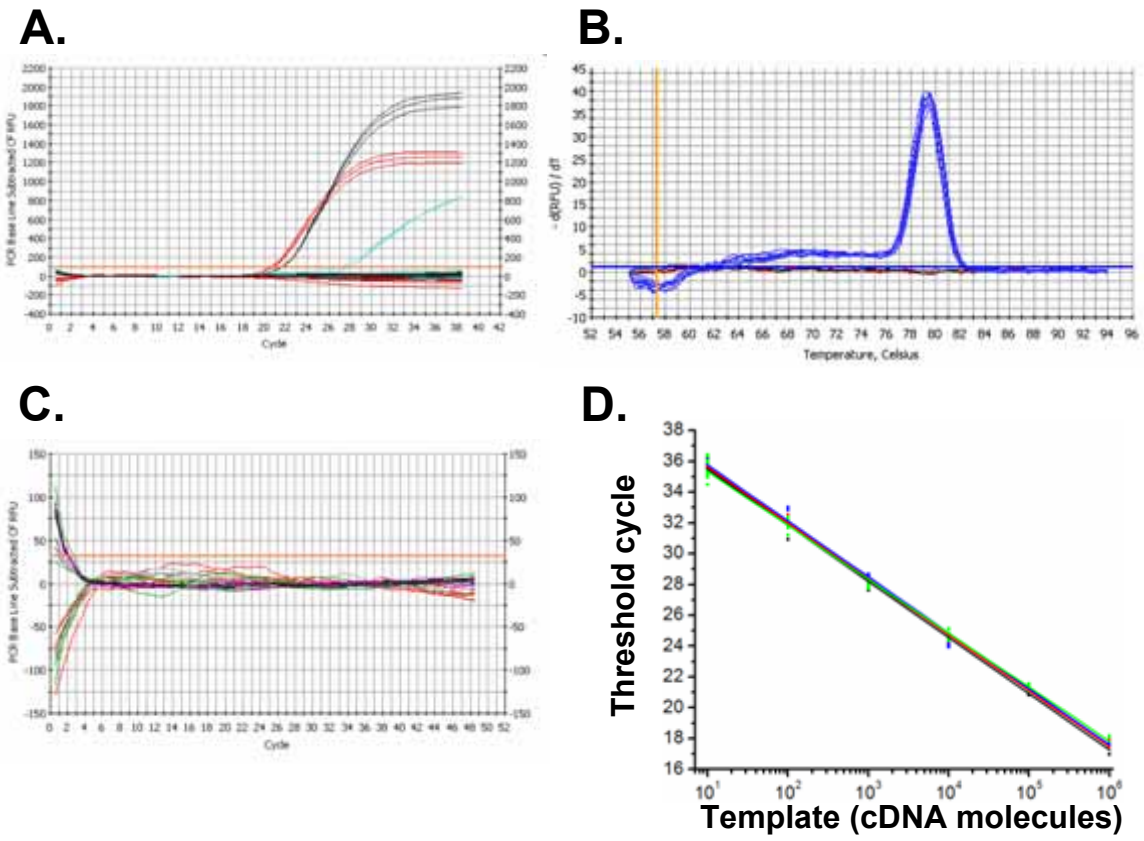


Figure 4.3. Bacterial mRNA primer validation. Orange lines in amplification plots areplaced at the fluorescence threshold, in order to determine Ct values. **A.** Threbacterial mRNA primer sets [Table 4.1, thr (black); lys (cyan); dap (red)] were tested against serially diluted NIH/3T3 mRNA (0.1 pg, 1 pg, 10 pg, 100 pg, n=3 for each template) as well as 100,000 copies of the respective bacterial mRNA template. Only the positive controls were amplified for the three genes. **B.** Post-qPCR melting peak analysis for six distinct Dap templates (10^6 copies → 10 copies Dap mRNA, serially diluted by a factor of 10, n=3 for each template) **C.** Repeat of experiment in (**A**), but with only Dap primers and probe. Five repeats for each NIH/3T3 template in the dilution series (0.1 pg, 1 pg, 10 pg, 100 pg, 1 ng) were run. **D.** qPCR standard curves generated in a triplex reaction (β-actin - red, GAPDH - black, PGK – blue, see Ch. 5 for fits) or a simplex (Dap – green, Y = -3.5(X) + 38.82, R = -0.99) reaction. Templates for all reactions were spectrophotometrically quantified PCR purified products.

4.3 Benchtop reverse transcription efficiency measurements

Once it was established the Dap primer/probe set was amenable for microfluidic efficiency measurements, the author ran an RT-qPCR dilution series experiment with 10^6 Dap mRNA molecules → 10 Dap mRNA molecules, diluting each template by an order of magnitude for a total of six template concentrations. Amplification curves for this experiment, as well as the corresponding standard curve, are given in Fig. 4.4A-B. The figure shows the RT-qPCR to be reproducible and precise, evident from the overlapping amplification plots of identical standards and the highly reliable correlation coefficient of the corresponding standard curve (R = -0.99). The standard curve's goodness of fit to the

data also implies equal amplification efficiency for all templates tested. If reaction efficiencies varied by template concentration, linear regression analysis would be problematic and therefore quantifying a range of templates present in mRNA or cell samples would also pose troublesome. Only if amplification efficiency of all genes of interest varied identically would quantification be possible in the unequal efficiency case. Purification of the RT-qPCR products in (**A**) , and subsequent quantification by spectrophotometry, enabled the author to generate a standard curve composed of varying concentrations of the 86 bp Dap cDNA fragment (Fig. 4.3D green and Fig 4.4B red). A measure of reverse transcription efficiency can then be extrapolated by plotting the Ct values from the RT-qPCR (Fig. 4.4A-B, black) onto the qPCR (Fig. 4.4B, red) standard curve. Averaging over all templates tested (10 molecules → 10^6 molecules), a reverse transcription efficiency of 77% was measured, meaning for every one mRNA molecule, 0.77 corresponding cDNA molecules were synthesized (Table 4.2). The coefficient of variation for all the measurements was 0.25. Measuring reverse transcription efficiency for individual templates shows the enzyme to be fairly consistent over six orders of magnitude. However, the variation is much higher for the lowest template concentration than for the more concentrated ones (Table 4.2). This is expected, as reverse transcriptase and *taq* polymerase are less efficient and more variable for low template concentrations because of the stochastic nature of the RT and PCR reactions. The yields obtained fall within the range of reverse transcription efficiency found in previous studies.[51, 52] However, the average efficiency values found in the literature tend to be ~0.5, but the enzyme utilized in this thesis was not tested in previous published reports.[51, 52]

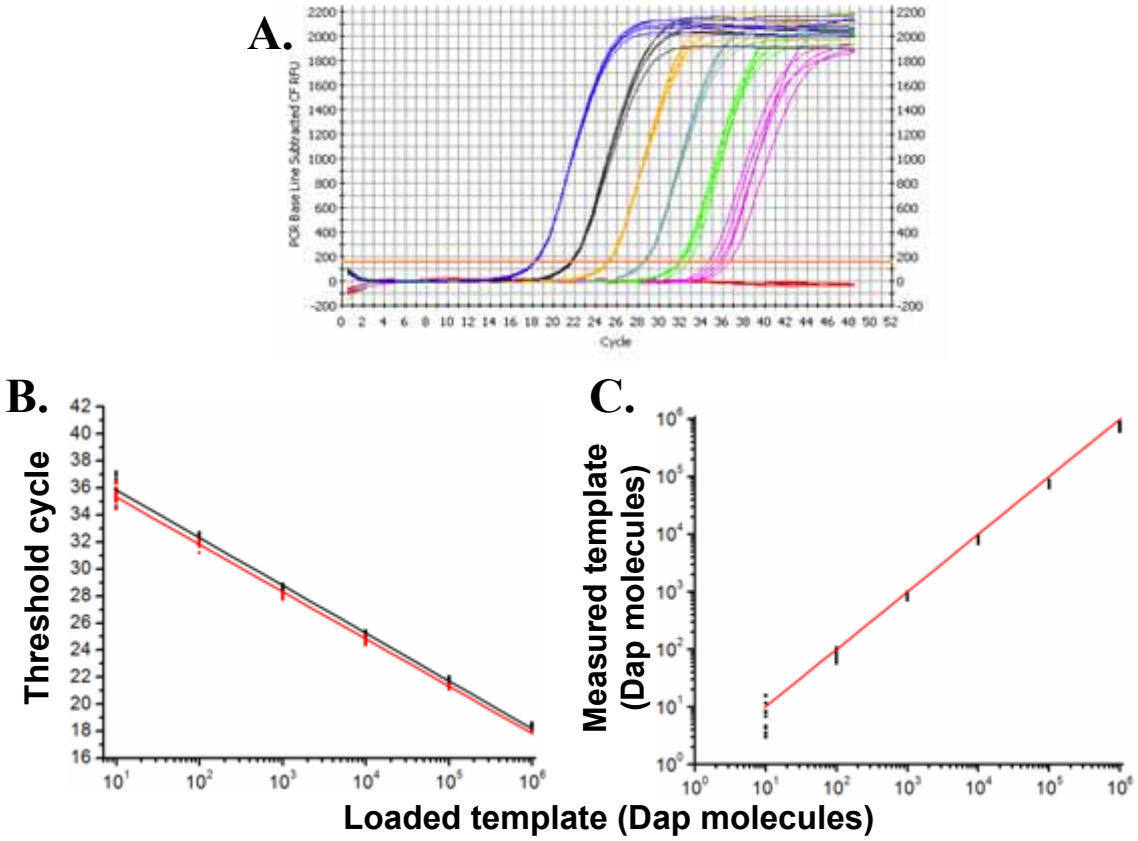


Figure 4.4. Off-chip reverse transcription efficiency experiments. **A.** Dap amplification curves for serially diluted [10^6 copies (magenta) → 10 copies (blue)] Dap mRNA. Red curves are no-template control reactions. Orange line is placed at the fluorescence threshold, which determines Ct values. **B.** Black: standard curve generated from the Ct values in (**A**). Y = -3.54(X) + 39.4, R = -0.99. n=10 for each template. Red: Standard curve generated from a qPCR with PCR purified templates. The curve is identical to the blue curve in (**Fig. 4.3C**). Y = -3.50(X) + 38.82, R = -0.99. **C.** RT efficiency measurements over six orders of magnitude. The Ct values from (**A.**) were plotted onto the red standard curve in (**B.**) to arrive at a measured value of cDNA molecules. The line indicates 100% efficiency (m=1).

RT Efficiency Measurements

mRNA copies (n=10 for each)	10	100	1000	10000	100000	1000000	all templates
average efficiency	0.68	0.83	0.81	0.8	0.77	0.73	0.77
variance	0.16	0.021	0.0059	0.0078	0.0053	0.0069	0.036
SD	0.39	0.14	0.077	0.088	0.073	0.083	0.19
c.v.	0.58	0.17	0.095	0.11	0.095	0.11	0.25

Table 4.2. RT efficiency measurements and corresponding statistics. SD is standard deviation and c.v. is coefficient of variation.

4.4 Primer/probe design for multiplex study

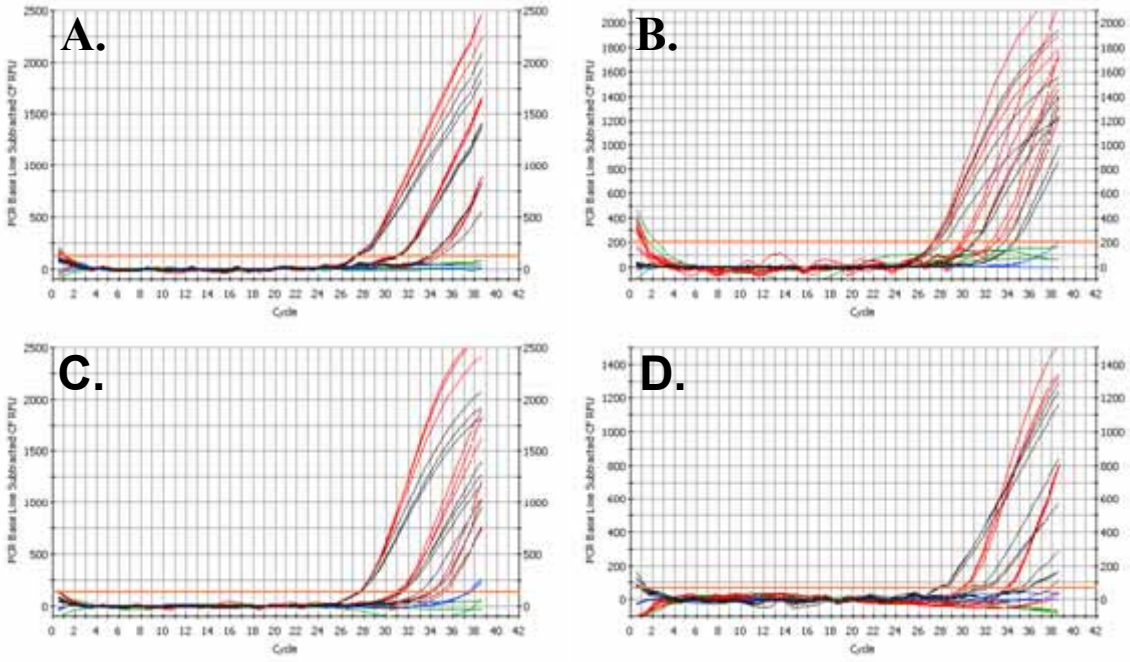
One of the scientific goals for the technologies in this thesis is to study gene expression noise at the single cell level. To do this, the author chose to examine genes expressed ubiquitously in NIH/3T3 cells. Real time detection is limited to four genes per assay because the instrument contains four distinct emission channels. Accordingly, the author opted for three genes present in the glycolysis pathway along with the cytoskeleton protein, β-actin for the initial primer validation studies. Primer and probe characteristics for the genes of interest are given in Table 4.3. As with the bacterial mRNA design principles, the T_ms of the primers and probes were designated to be 60°C ± 3°C and T_m(primer) + 10°C, respectively. Primer-dimer energies were also constrained to be ≤ -6 kcal/mol in order to limit competition of non-specific product formation during the RT and PCR reactions. To validate the functionality of the primer/probe sets, the author performed simplex and fourplex qPCR reactions in parallel, as well as simplex and

triplex qPCRs in parallel, for all four genes, or three of the four (Fig. 4.5). 10 pg/μL liver cDNA was diluted serially by factors of 10 to give three templates for the experiment: 10 pg/reaction, 1 pg/reaction, and 0.1 pg/ reaction. Each template was run in triplicate for simplex, triplex, and fourplex reactions. It should be noted the experiments in **A.-D.** were run on different days than the ones pictured in **E.-G.** Accordingly, the simplex reactions are not identical for the respective genes when comparing the black curves to the blue curves in Fig. 4.5. As shown in Fig. 4.5, Ct values for simplex and fourplex as well as simplex and triplex reactions overlap for each gene for every template tested, indicating primer-dimer formation is not an issue during the qPCR. However, there is much more variation for the Cy5 channel (PGK) in the fourplex reaction than the triplex. This result was confirmed over multiple qPCR runs and it was determined to employ three gene detection for the single cell measurements. Further optimization of the fourplex qPCR should allow for quantitative detection of the four genes in parallel (Table 4.3).

Primer and Probe Specifications for Multiplex Study

	β-actin	GAPDH	PGK	Enolase
Accession number	NM_007393	NM_001001303	M15668	NM_023119
Sequence length (bases)	1,892	1,243	1,653	1,720
Probe sequence	CCA TCC TGC GTC TGG ACC TGG CTG GC	CGT GCC GCC TGG AGA AAC CTG CCA AGT	ACC ACA GTC CAA GCC CAT CCA GCC AGC A	CCA CCT GGA TGC CCG CAC TAG CCG T
Probe position; length	601; 26	789; 27	973; 28	1,041; 25
Probe Tm (°C); GC%	68.7; 69.2	68.3; 63	68.9; 60.7	68.7; 68
Beacon Designer probe rating; quality	69.4; good	64.5; good	77.2; best	73.7; good
Sense primer sequence	CCA TCT ACG AGG GCT ATG CTC TCC	CCA ATG TGT CCG TCG TGG ATC TG	ACC TTG CCT GTT GAC TTT GTC ACT G	GTA CAA GTC CTT CGT CCA GAA CTA CC
Sense primer position; length	571; 24	765; 23	863; 25	932; 26
Sense primer Tm (°C); GC%	60.4; 58.3	62.3; 59.9	59.6; 48	58.4; 50
Antisense primer sequence	CAC GCT CGG TCA GGA TCT TCA TG	TCC TCA GTG TAG CCC AAG ATG CC	CCT CGG CAT ATT TCT TGC TGC TCT C	TTG GCA ATC CGC TTA GGG TTG G
Antisense primer position; length	669; 23	888; 23	1,001; 25	1,084, 22
Antisense primer Tm (°C); GC%	59.7; 56.5	60.9; 56.5	59.7; 52	60.2; 54.5
Product length; produt Tm (°C)	99; 90.4	124; 90.4	139; 89.3	153; 92.7

Table 4.3. Primer and probe specifications for multiplex PCR study.



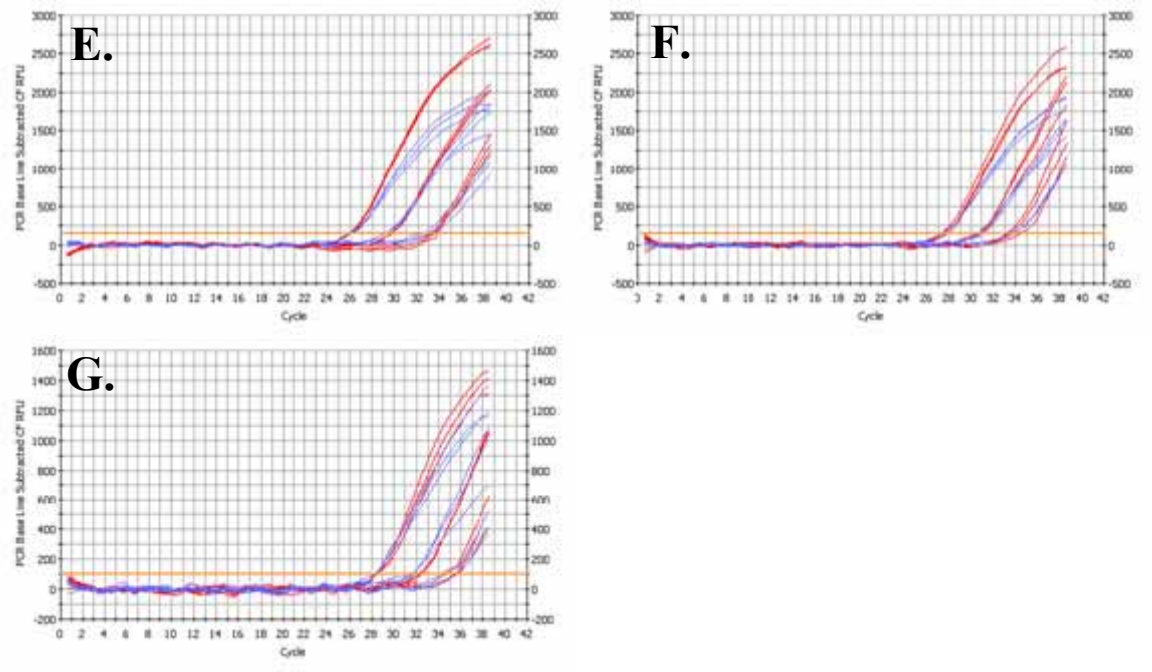


Figure 4.5. Amplification plots for fourplex, triplex, and simplex qPCRs. **A.-D.** Fourplex reactions. Negative controls for simplex and fourplex reactions are shown in green and blue, respectively. Red amplification curves are simplex reactions and black curves are fourplex reactions. **A.** Enolase (Fam) **B.** GAPDH (Texas red) **C.** β-actin (Hex) **D.** PGK (Cy5). **E.-G.** Triplex reactions. Red amplification curves are simplex reactions and blue curves are triplex reactions. Negative controls are not pictured. **E.** GAPDH (Fam) **F.** β-actin (Hex) **G.** PGK (Cy5).

4.5 Combating non-specific product formation during RT step

When the author ran RT-qPCR with multiple primer/probe sets (GAPDH, β-actin, PGK) in a single reaction, it became clear that non-specific products form during the reverse transcription step (Fig. 4.6). Multiplex qPCR over three orders of magnitude (10 pg, 1 pg, 0.1 pg liver cDNA) showed identical Ct values to those of the corresponding

simplex reactions (Fig. 4.6, column a). If primer-dimers were forming from the onset of the PCR, it would be reasonable to suspect the Ct values for the multiplex reactions to be higher than those of the respective simplex reactions, and that is not seen. However, non-specific products are forming during the exponential phase of the multiplex reactions, as shown by the amplification curves plateauing earlier than the corresponding simplex curves. The plateau is due to reagent consumption in the formation of non-specific products, thereby minimizing reagents for the main product to amplify. The non-specific product formation during PCR is not an issue for the NIH/3T3 cell measurements because Ct values are resolved for all three genes in the range necessary to detect messages from single cells. In contrast, when utilizing identical concentrations of NIH/3T3 mRNA to run multiplex RT-qPCR, only the most concentrated template (10 pg/reaction) is detected when probing GAPDH and PGK. For β-actin, 10 pg and 1 pg NIH/3T3 mRNA templates can be resolved. In order to measure transcript levels in single cells, the lowest template concentration (0.1 pg/reaction) needs to be discerned from the measurement noise. It is known that reverse transcriptase is active at room temperature and demonstrates non-specific priming. These two facts make it very difficult to perform multiplex RT-qPCR reactions in a one-tube format.

The options to combat the non-specific reverse transcription problem are as follows: (1) design new primers, (2) perform a two step reaction, both steps benchtop (first strand cDNA synthesis reaction followed by qPCR), (3) perform a two step reaction: microfluidic first strand cDNA synthesis, followed by qPCR. The strong amplification plots obtained on cDNA templates down to 0.1 pg cDNA (Fig. 4.4 column a) taken together with the characteristics of the reverse transcriptase enzyme, make it

highly unlikely that designing new primers is the answer. Option two is a viable one, although not optimal. Loss of product may become an issue if the beads are handled off-chip in additional processing steps. Option three is the best because primer-dimers and other non-specific products will be washed away after the RT step and subsequent buffer exchange.

Varying templates of NIH/3T3 mRNA (50pg, 10pg, 1pg and 0.1pg) were isolated with 20X mRNA isolation chips (ch. 5). First strand cDNA synthesis on-chip was then carried out on-chip (ch. 2, 3, 5) followed by recovery of the products. qPCR was subsequently employed on the microfluidic NIH/3T3 standards (Fig. 4.4). The author compared Ct values and amplification plots of the microfluidic samples to amplification plots generated with a RT-qPCR (Fig. 4.6). These results indicate microfluidic cDNA synthesis is a solution to the non-specific product formation during RT and subsequent inhibition of the qPCR. The plots are cleaner, the exponential phase goes for longer, and templates down to 0.1 pg are resolvable for all three genes, which is not the case when RT-qPCRs are implemented. Despite the effectiveness of the microfluidic solution, primer-dimers still are detected in the hex (β-actin) channel around cycle 35. In order to quantify β-actin in single cells with the existing primer set, it is necessary for the respective cell's mRNA amplification to occur before the non-specific amplification. This is indeed what is seen for single NIH/3T3 cells (ch. 5).

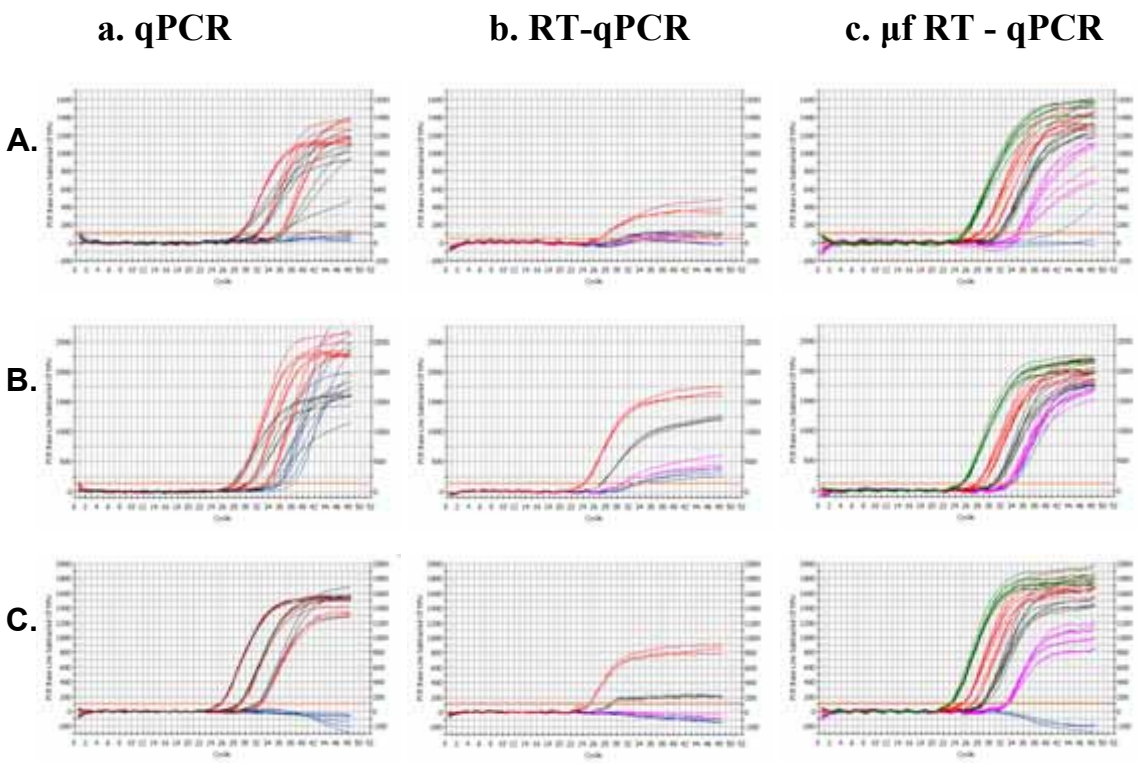


Figure 4.6. Amplification plots showing µF RT-qPCR is the preferred format. PGK (cy5), β-actin (hex), and GAPDH (fam) curves are given in **A., B. and C.**, respectively. Column **a.** shows simplex (red) and triplex (black) qPCRs for the three genes of interest. Templates are in triplicate and are either 0 (blue curves), 0.1, 1, or 10 pg liver cDNA. Column **b.** gives amplification plots for triplex RT-PCR reactions for the three genes. Templates are in triplicate and are either 0 (blue), 0.1 (magenta), 1 (black), or 10 pg (red) NIH/3T3 mRNA. Column **c.** shows amplification plots for NIH/3T3 samples subjected to microfluidic first strand synthesis, followed by qPCR. Templates are either 0 (blue), 0.1 (magenta), 1 (black), 10 pg (red), or 50 pg (green) NIH/3T3 mRNA.

Chapter 5: Measuring gene expression noise in single mammalian cells using microfluidics

5.1 Introduction

5.1.1 Noise background

It is widely accepted that clonal populations of cells exhibit considerable phenotypic variation, which is manifested by differences in molecular expression.[53] The variation, or noise, has been implicated as the underlying phenomenon necessary for cellular differentiation events and development of heterogeneous tissues or systems.[53, 54] Therefore, a better understanding of how noise is generated and the quantification of it may provide insight into processes such as stem cell differentiation events and cancer cell generation/metastasis. Variations in gene expression have been shown indirectly in single isogenic *E. coli* and yeast by measuring dual reporter protein levels.[53, 55, 56] This method incorporates GFP variants into chromosomes, so as to circumvent the extrinsic noise added into systems when utilizing plasmids free in the cellular environment.[57] Plasmid number is unable to be quantified because the concentration fluctuates randomly inside a cell, thereby introducing a source of extrinsic noise. The single plasmid method[58] is most like the system described here in that there is some extrinsic noise initially, due to fluctuating mRNA levels, inherent decay, RNases, or ribosomal translation. With the dual reporter method, the total noise (η_{tot}) of a system can be deconvoluted into two categories: (1) fluctuations of cellular components (extrinsic noise, η_{ext}) and (2) inherent stochasticity present in the biochemical process of gene expression (intrinsic noise, η_{int}).[53-55, 59] Intrinsic noise is defined by the extent of which the two reporter proteins' expressions in the single cell fail to correlate, whereas the cell to cell difference in

reporter protein levels defines the extrinsic noise.[53] To date, the dual reporter technique has yet to be applied to a mammalian system. Single reporter techniques have been used to quantify dynamic gene expression in single cells and can be conceivably utilized for the measurement of η_{tot} in individual mammalian cells.[60, 61] This technique has already been applied to the study of noise strength in single *S. cerevisiae*.[58] Results in the literature are consistent in the fact that extrinsic noise is found to be the main contributor to the total noise of the system. This is not surprising considering the many sources for this noise (e.g. molecule concentration, cell cycle stage, mRNA decay rate, environment)

Reports discussing gene expression noise measurements, to this date, have measured the corresponding protein levels to study noise in the genes expressed. This is not ideal in that translational efficiency is not directly measurable because mRNA levels are not initially quantified. Furthermore, noise has been extrapolated from arbitrary units of fluorescence in protein levels, which makes it impossible to compare measurements across laboratories unless a calibration in terms of molecules is present. This problem parallels what is faced by researchers trying to discern the microarray data across laboratories. The standard metrics to quantify noise assume cell distributions to be normal about the mean. There have been a few cases in which histograms show normal distributions for protein distribution.[58, 62] These distributions are surprising in that there are a multitude of reactions needed to go from DNA to protein expression, and all are sensitive to random fluctuations of reactant molecules. In addition, studies that show the normal protein distribution are in experimental contrast with a portion of the literature that demonstrates intracellular mRNA and protein distributions support log transformations.[15, 56, 63-65] In most studies, histograms are lacking and careful statistical

analysis is therefore overlooked. In one study,[56] the authors claim a lognormal fit describes their dataset with more certainty than a normal Gaussian distribution; however the metrics to quantify the corresponding noise are still linear in nature. Although the coefficient of variation, which is commonly used as a noise metric, for a lognormally distributed dataset is equivalent to the untransformed set; a more correct characterization of the population, if lognormal, would be to incorporate a geometric mean term. In the study presented in this chapter, the author measured gene expression noise directly at the transcriptional level in single NIH/3T3 cells, and asked to what degree cell cycle phase and pathway specific gene expression are contributors to η_{tot} in the cell population. The following is the first reported direct measurement of transcriptional noise in single cells, regardless of cell type, and the first measurement of noise in single mammalian cells.

5.1.2 Lognormal distribution background

There is increasing evidence in the literature that products of various biological processes and reactions, transcription and translation included, are distributed lognormally. The following section will describe the lognormal distribution mathematically.

Consider a lognormally distributed variable x with mean μ_x and standard deviation σ_x, denoted $LN(\mu_x, \sigma_x^2)$. Then, the variable y, where

$$y=\ln(x) \tag{1}$$

is normally distributed with mean μ_y, standard deviation σ_y, and is denoted $N(\mu_y, \sigma_y^2)$. It should be noted μ_y and σ_y^2 refer to the parameters of the log-transformed variable y, and

therefore need to be back-transformed for meaningful analysis. The probability density function f(x) (untransformed data) satisfying the lognormal distribution is stated as

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln x - \mu_y}{\sigma_y}\right]^2}, \quad (2)$$

It can be seen from the probability density function in mathematical and graphical terms (Fig. 5.1), that the lognormal distribution only holds for values of x > 0, and the distribution encompasses a tail (skewed) to the right of the mean.

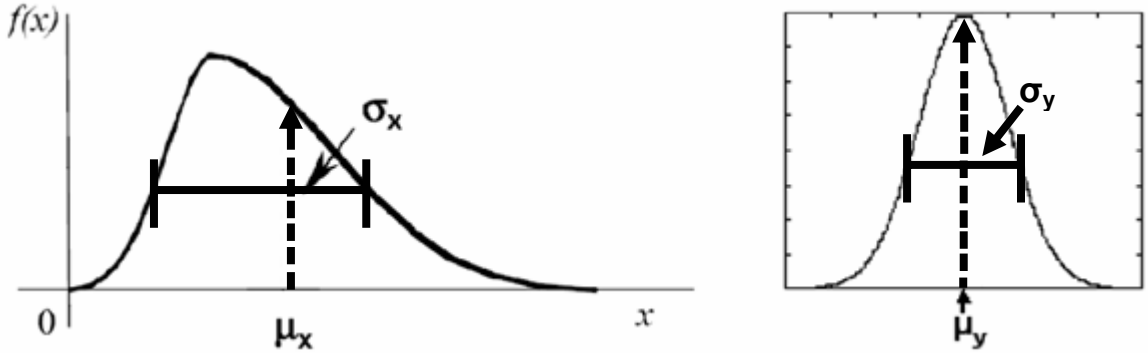


Fig 5.1. Probability density function of the lognormal distribution.[66] **Left**: Linear coordinates along the x axis. The peak of the curve corresponds to the median and the geometric mean of the dataset. **Right:** Log transformation of plot on the left, with μ_y and σ_y labeled.

The mean and variance for the untransformed dataset, μ_x and σ_x, can be discerned if μ_y and σ_y are known for the transformed set:

$$\mu_x = e^{\mu_y + \frac{1}{2}\sigma_y^2} \quad (3)$$

$$\sigma_x^2 = e^{2\mu_y + \sigma_y^2}\left(e^{\sigma_y^2} - 1\right) \quad (4)$$

If μ_y and σ_y are unknown, but μ_x and σ_x are known, it follows from (1) that

$$\mu_y = \ln\left(\frac{\mu_x^2}{\sqrt{\mu_x^2 + \sigma_x^2}}\right) \tag{5}$$

$$\sigma_y^2 = \ln\left(1 + \frac{\sigma_x^2}{\mu_x^2}\right) \tag{6}$$

It should be noted again at this point μ_x corresponds to the population mean, and not the geometric mean, which will correspond to the median of the untransformed dataset:

$$\text{Geometric Mean(x)} = \text{Median f(x)} = \exp(\mu_y) \tag{7}$$

Now, by taking the square root of equation (4), and dividing it by equation (3), an expression for the coefficient of variation of the dataset arises:

$$CV = \sqrt{\exp(\sigma_y^2) - 1} \tag{8}$$

Again, equation (8) takes the log-transformed variance as its argument. Interestingly, the c.v. is only a function of σ_y, and not of the mean, μ_y of the transformed dataset.

5.1.3 Modeling noise in transcription–considerations and differences with reporter systems

The main difference in the system employed here as opposed to elsewhere (in bacteria and yeast) is that the latter utilized inducible promoters and the former is a snapshot with background mRNA levels providing a source of extrinsic noise. This can be reconciled by either calculating or modeling mRNA decay rates.[67] Because the gene expression data for all genes tested exhibit lognormal distributions, the model should incorporate processes that are multiplicative in nature. For example, Paulsson[57] describes the probabilities of having n_1 and n_2 molecules per cell of chemical species X_1 and X_2

(for his example X_1,X_2 =mRNAs, proteins) as a birth and death Markov process with events

$$n_1 \xrightarrow{R_1^{\pm}(n_1)} n_1 \pm 1 \text{ and } n_2 \xrightarrow{R_2^{\pm}(n_1,n_2)} n_2 \pm 1 \quad .$$

These statements capture the dynamic disorder of the cell because n_1 affects rate R_2, but n_2 does not affect R_1. In the system used here, protein translation is not measured; however, these rates still hold for the intracellular environment. For example, n_1 could represent the rate-limiting component of the transcriptional machinery (including chromatin remodeling proteins), whose concentration and fluctuations would randomize mRNA transcription, to an extent.

5.2 Results

5.2.1 Single cell measurements

With each step along the microfluidic process quantified, and triplex qPCR parameters worked out, it is now possible to measure absolute mRNA copy numbers of the specified genes in single cells. NIH/3T3 cells were FACS sorted based on DNA content, and a total of 202 cells (99 G1, 103 G2) were processed microfluidically to cDNA, utilizing the 50X devices discussed later in the chapter. These samples were then subjected to qPCR, after which Ct values were plotted onto the respective standard curves, to obtain raw copy number values. Cells that contained copy numbers < 10 (before correction) for any of the three genes were immediately thrown out because 10 copies is the detection limit, in the author's hands, for triplex qPCR. A total of 10 G1 and 12 G2 cells fulfilled the above criteria, and were scrapped, leaving 89 G1 and 91 G2 cells for analysis. The author first corrected the raw copy number data by dividing the values by 0.12 (total efficiency of process, see methods). mRNA copies for each cell

were then normalized by their respective population average, and graphically represented as a three-dimensional scatter plot, either sorted based on DNA content, or unsorted (Fig. 5.2A, B). G1 and G2 populations can be distinguished from each other in this representation (Fig. 5.2A). However, when the data are unsorted, and normalized by the entire population average, only one population is observed (Fig. 5.10B). These results show cell cycle phase to be a factor in the absolute expression of the three ubiquitous housekeeping genes in single NIH/3T3 cells. The G1 and G2 populations are also distinguishable when plotting GAPDH vs. PGK expression two dimensionally (Fig. 5.2C).

Because the distributions in 5.10A and B exhibit long tails, the author transformed the data into log coordinates, subtracting the log mean copy number from each cell's transcript copy number (Fig. 5.2D). Upon doing so, the two populations are undistinguishable from each other (Fig 5.2D). Furthermore, transformed copy number values centered about the mean for all three genes in both populations, indicating lognormal distributions. The author further tested the distributions of the dataset by plotting all pairwise correlation data in two-dimensional scatter plots (Fig. 5.2E-H). It was found that the pooled G1 and G2 populations are co-linear with respect to PGK and β-actin expression (Fig. 5.2E). Furthermore, these two populations pass a Kologmorov-Smirnov test indicating the same underlying distribution (P=0.05). When the data are separated according to cell cycle phase, the underlying β-actin and PGK distributions still pass the Kologmorov-Smirnov test (P=0.2 for both, Fig. 5.2F-G). However, when normalizing expression by the mean copy number of the respective transcript, the distributions in the two genes do not match (Fig. 5.10H). Because PGK and β-actin have

different promoters and are located on different chromosomes, it is highly unlikely the distributions are identical, as the statistical test implies. Furthermore, because the distributions do not match statistically when subtracting out the mean transcript number, the author takes the distribution data in Fig. 5.2E-G to be artifactual in nature.

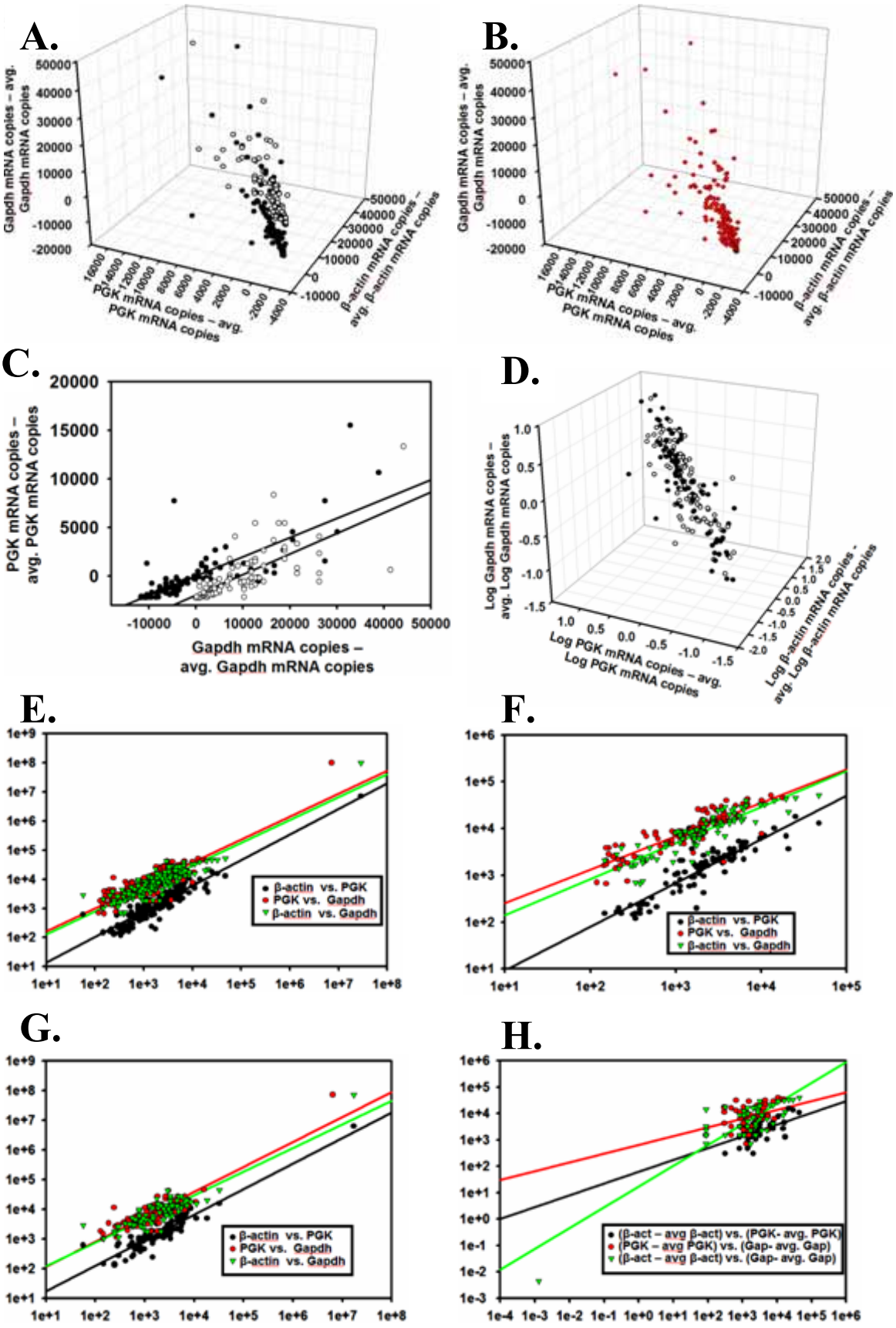
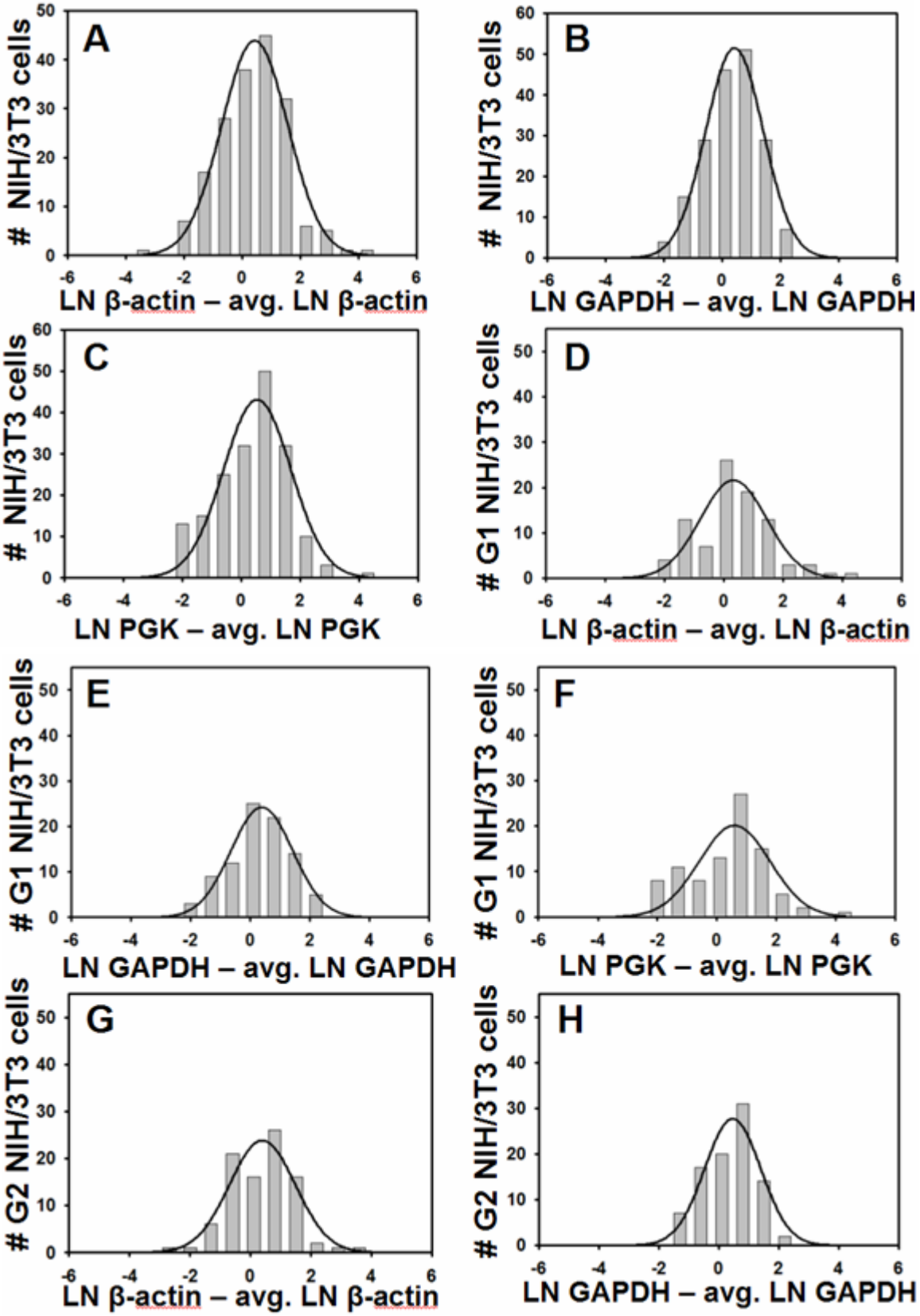

A.
Gapdh mRNA copies – avg. Gapdh mRNA copies
PGK mRNA copies – avg. PGK mRNA copies
β-actin mRNA copies – avg. β-actin mRNA copies
B.
Gapdh mRNA copies – avg. Gapdh mRNA copies
PGK mRNA copies – avg. PGK mRNA copies
β-actin mRNA copies – avg. β-actin mRNA copies
C.
PGK mRNA copies – avg. PGK mRNA copies
Gapdh mRNA copies – avg. Gapdh mRNA copies
D.
Log Gapdh mRNA copies – avg. Log Gapdh mRNA copies
Log PGK mRNA copies – avg. Log PGK mRNA copies
Log β-actin mRNA copies – avg. Log β-actin mRNA copies
E.
β-actin vs. PGK
PGK vs. Gapdh
β-actin vs. Gapdh
F.
β-actin vs. PGK
PGK vs. Gapdh
β-actin vs. Gapdh
G.
β-actin vs. PGK
PGK vs. Gapdh
β-actin vs. Gapdh
H.
(β-act – avg β-act) vs. (PGK- avg. PGK)
(PGK – avg PGK) vs. (Gap- avg. Gap)
(β-act – avg β-act) vs. (Gap- avg. Gap)

Figure 5.2. Gene expression correlation data. **A.** Three dimensional scatter plot showing correlation between β-actin, PGK, and GAPDH gene expression. Raw copy numbers were normalized by subtracting out the G1 (black) or G2 (white) population mean. **B.** Copy numbers for genes were normalized by the average copy number for the entire dataset and plotted three dimensionally. **C.** GAPDH vs. PGK expression in single G1 (black) and G2 (white) NIH/3T3 cells. **D.** Copy numbers for genes were log-transformed and normalized by the average log copy number for the respective gene (G1 –black, G2 – white). **E.-G.** Transcript correlation data without normalization by the mean. **E.** Pooled dataset. **F.** G1 cells. **G.** G2 cells. **H.** Pairwise gene expression correlation with mean normalization.

In order to test for lognormal distributions in the datasets, the author plotted histograms of ln transformed copy number values for each gene and cell cycle phase; after normalizing each cell's value to the ln mean mRNA copy number (Fig. 5.3). Upon fitting the histograms to Gaussian distributions, it was revealed that the three genes are lognormally distributed in single NIH/3T3 cells, regardless of DNA content (heterogeneous, G1, or G2; Fig. 5.3). The lognormal distribution is not surprising because the process of transcription is dependent on multiplicative effects such as concentrations of molecules required for intracellular signaling and the necessary convergence of proteins to form the transcriptional machinery. The author also showed that normalizing each gene pairwise, by taking the difference in natural logs, resulted in lognormal distributions for all ratios tested (Fig. 5.4). This result is useful because the difference of two logs, distributed normally is equivalent to the lognormal distribution of

the copy number ratios. The lognormal distributions visualized are also another validation of the microfluidic cells-to-cDNA process. If the steps along the process were not repeatable and consistent (i.e. if lysis did not go to completion), one would expect to see random distributions. These results are also consistent with gene expression data from single pancreatic cells, as well as what is found for protein expression in single *E. coli*.[15, 56]


A
NIH/3T3 cells
LN β-actin – avg. LN β-actin
B
NIH/3T3 cells
LN GAPDH – avg. LN GAPDH
C
NIH/3T3 cells
LN PGK – avg. LN PGK
D
G1 NIH/3T3 cells
LN β-actin – avg. LN β-actin
E
G1 NIH/3T3 cells
LN GAPDH – avg. LN GAPDH
F
G1 NIH/3T3 cells
LN PGK – avg. LN PGK
G
G2 NIH/3T3 cells
LN β-actin – avg. LN β-actin
H
G2 NIH/3T3 cells
LN GAPDH – avg. LN GAPDH

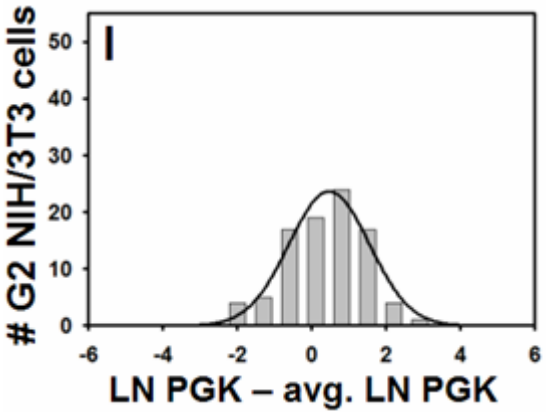


Figure 5.3. Individual genes are distributed lognormally over pooled (**A.-C.)** G1 (**D.-F.**) and G2 (**G.-I.)** populations.

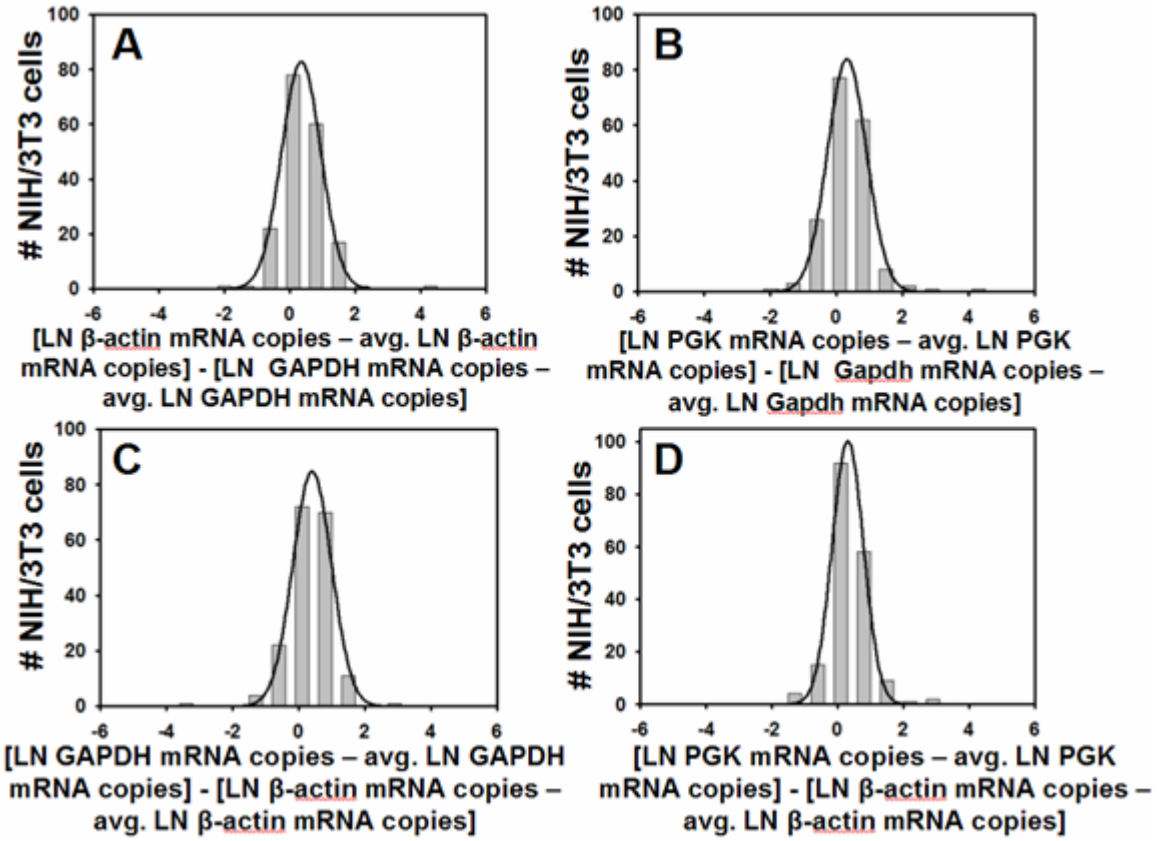


Figure 5.4. Gene expression normalized by in- and out-of-pathway genes. All combinations are distributed lognormally over pooled, G1, and G2 populations. Only pooled data are shown. **A., B.** Normalized to GAPDH. **C., D.** Normalized to β-actin.

5.2.2 Noise measurements

As stated above, mRNA distributions are distributed lognormally at the single cell level. Despite this, the author quantified noise in the untransformed dataset, in terms of coefficient of variation (c.v.: standard deviation/mean) and Fano factor (variance/mean). Because some of the noise literature is void of cell population histograms and/or indicate lognormal distributions, the author did as the authors of these studies do: quantified noise based on the previously stated linear metrics (Table 5.4). The author first determined if the noise exhibited in NIH/3T3 cells scaled as a Poisson process (Fig. 5.5A). If the noise was Poissonian, the Fano factor would equal one (line in Fig. 5.5A: m=1). As the figure shows, the mean copy number does scale with the variance, but the scaling is orders of magnitude higher than what is seen in a Poisson process. The c.v. and Fano factors obtained for the datasets are exceedingly higher than what is seen in the literature (Fig. 5.5 B, C). However, comparing datasets across studies may not be meaningful because noise is not quantified in terms of copy numbers in previous work, as is done in the present study. Nevertheless, possible explanations for the high noise include: (1) mammalian cells were used in this study as compared to yeast and bacteria in previous work, and (2) gene expression occurs on a single locus in the present study, whereas engineered cells utilizing The transcriptional noise seen in NIH/3T3 exceeds the measurement error, in some cases by five times; and therefore is a real affect.

Controlling for the extrinsic variable of DNA content, the author demonstrated noise values decrease for the three genes in the G2 stage of the cell cycle, but remain relatively unaffected in G1, as compared to a heterogenous population (Fig. 5.5, Table 5.1). These results seem to implicate regulatory control of the glycolysis enzymes before cell division. This is not surprising in that the daughter cells require some minimal concentration of these enzymes to survive. Interestingly, when controlling for the basic cellular environment, by normalizing copy numbers of the glycolysis genes to β-actin gene expression, noise is seen to increase for the gene ratios as compared to the raw copy number values. However, the noise in the ratio groups decreases in G1 as compared to G2 and the heterogenous sample (Fig. 5.5). Furthermore in all three groups, the ratios' coefficients of variation are higher than the noise measured for the glycolysis genes alone. Although the total noise in all three genes goes down in G2, the ratios of gene expression increase most likely due to the genes being active in distinct cellular processes, with intracellular variability non-deleterious if copy numbers stay above a certain threshold. The cell probably makes more than enough of the three genes studied and employs regulation before divison. This mechanism could decrease noise at the protein level (e.g. high copy numbers are the limiting case for translational noise) for the housekeeping genes, with excess transcripts always available for translation, thus obviating stochastic affects. The next generation of measurements, coupling protein expression (utilizing fluorescence microscopy before lysis) to mRNA transcript expression, should resolve this issue.

Gene Expression Noise Based on Raw Copy Numbers

	g1 β actin	g1 pgk	g1 GAPDH
mean	3833.18	2339.11	12365.96
geo mean	1921.7	1255.2	7993.2
std.dev	6313.28	2825.16	11200.06
var	4.03E+07	8.07E+06	1.27E+08
c.v.	1.65	1.21	0.91
Fano	10516.16	3450.98	10259.35

	g2 β actin	g2 pgk	g2 GAPDH
mean	3034.12	2375.62	10255.84
geo mean	1726.2	1458.8	7359.8
std.dev	4096.02	2513.21	8494.96
var	4.03E+07	8.07E+06	1.27E+08
c.v.	1.37	1.1	0.83
Fano	13285.7	3397.93	12370.2

	Pooled β actin	pooled pgk	pooled GAPDH
mean	3512.54	2382.8	11694.9
geo mean	1326.2	1104.6	5173.7
std.dev	5883	2823.4	12463.3
var	34609888.9	7971642.2	155334247.6
c.v.	1.67	1.18	1.07
Fano	9853.23	3345.52	13282.25

Table 5.1. Gene expression noise measurements calculated based on raw mRNA transcript values and the log-transformed data. The c.v. values are identical whether in the log or linear scale (see discussions on lognormal distribution).

The noise strength (Fano factor) for the copy number ratios is exceedingly lower than the ones found when analyzing the raw dataset. This demonstrates the need to stay in original units when quantifying noise strength. Normalization, in most cases, will decrease variance, and therefore decrease the noise strength.

Pairwise Normalization of Gene Expression Noise

	g1 actin/pgk	g1 gap/pgk	g1 pgk/actin	g1 gap/actin	g1 actin/gap	g1 pgk/gap
mean	1.77	7.60	0.78	4.78	0.28	0.20
Std.dev	1.14	4.78	0.75	2.72	0.17	0.23
var	1.30	22.81	0.56	7.42	0.03	0.05
c.v.	0.65	0.63	0.96	0.57	0.62	1.16
Fano	0.74	3.00	0.72	1.55	0.11	0.27
	g2 actin/pgk	**g2 gap/pgk**	**g2 pgk/actin**	**g2 gap/actin**	**g2 actin/gap**	**g2 pgk/gap**
mean	1.40	6.19	1.04	5.27	0.29	0.24
Std.dev	0.81	5.46	1.21	5.48	0.19	0.16
var	0.66	29.83	1.47	30.08	0.03	0.02
c.v.	0.58	0.88	1.17	1.04	0.65	0.66
Fano	0.47	4.82	1.41	5.71	0.12	0.10
	pooled actin/pgk	**pooled gap/pgk**	**pooled pgk/actin**	**pooled gap/actin**	**pooled actin/gap**	**pooled pgk/gap**
mean	1.58	6.90	0.91	5.04	0.28	0.22
Std.dev	1.00	5.14	1.01	4.31	0.18	0.20
var	1.00	26.42	1.02	18.60	0.03	0.04
c.v.	0.63	0.74	1.11	0.86	0.63	0.91
Fano	0.63	3.83	1.12	3.69	0.11	0.18

Table 5.2. Gene expression noise measurements calculated based on pairwise normalization of genes. Because these data are distributed lognormally when transformed (e.g. ln actin – ln pgk = log normal relationship), the author divided the raw copy numbers to arrive at these values

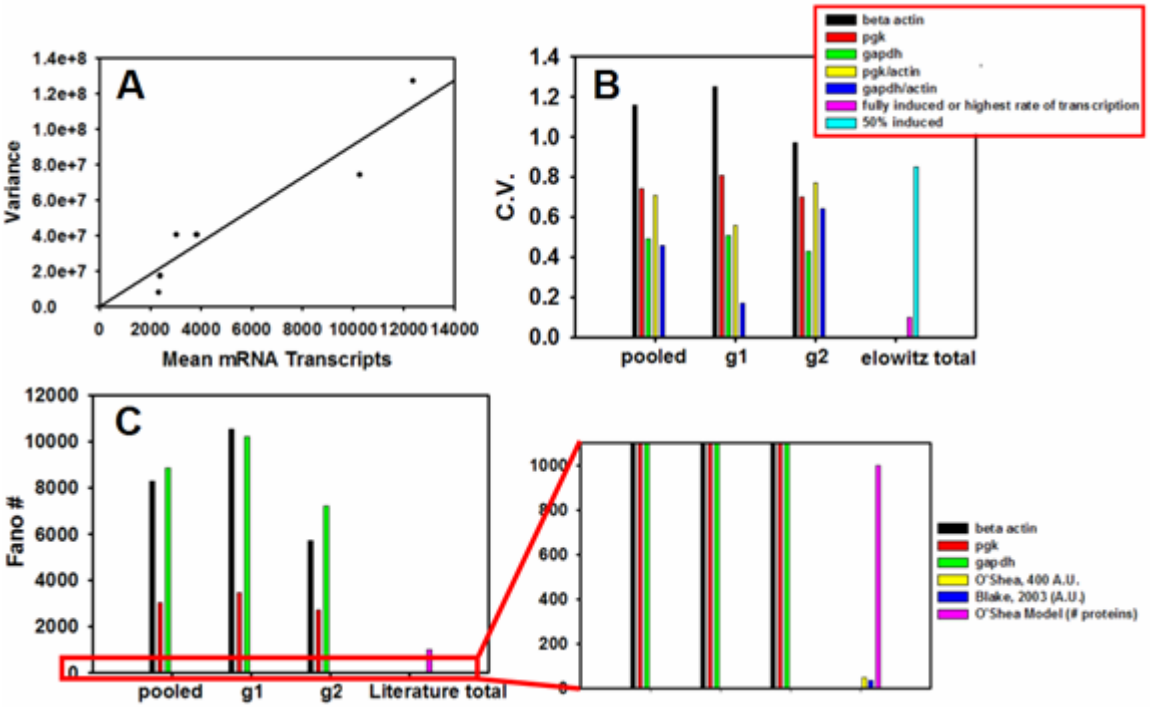


Figure 5.5. Raw noise measurements on NIH/3T3 cells. Extrinsic noise due to cell cycle stage and basic intracellular variability (β-actin normalization) are quantified. Note measurement error (~0.3) has not been subtracted.

5.3 Methods

Although the devices in chapters 2 and 3 are able to process single cells to cDNA in a parallel manner, running only four samples in parallel poses problems for the researcher in terms of obtaining enough samples for meaningful analysis. Single cells are ideally analyzed from the same cell culture or tissue section in order to minimize extrinsic noise in the samples analyzed, brought on by day to day variation (e.g. temperature fluctuations, age of sample). However, there are no widespread methods to process single cells to cDNA for gene expression studies, in a parallel manner. In this section, the author describes a solution to this problem and applies the novel technology to the study of gene expression noise in single NIH/3T3 cells.

5.3.1 Device component optimization

As is the case frequently when developing a nascent microfluidic device into a mature platform capable of carrying out a specific process, the initial device design varied considerably from the optimized design.

5.3.1.1 Column construction

The author first attempted to stack 20 to 80 bead columns in parallel. To do this, beads were resuspended to two times their normal concentration and flowed into a serpentine channel. Once beads were present throughout the serpentine, valves partitioned off each reactor, and lysis buffer pushed the beads against the sieve valves through an input upstream of the bead input. The author found the amount of beads trapped in the serpentine was far from sufficient to build a bead column of appropriate

length (~300 μm). Iterating between the lysis buffer push step and filling the serpentine channel produced better results; however, a full column could still not be stacked after 10 iterations. Concentrating the beads beyond 3X eventually clogged the bead input, so a better solution was needed.

In order to solve the bead stacking problem, the author implemented a fluidic multiplexer to individually address reactors on-chip (Fig. 5.6, Ch. 7). The multiplexer allows for unlimited bead columns to be stacked in a serial fashion. Figure 5.1 is an AutoCAD drawing of the bead stacking portion of a 20X cells-to-cDNA device along with an optical micrograph of 20 stacked columns. The columns are stacked in a uniform fashion regardless of whether 1, 2, 4, or 8 columns are constructed in unison (column pitch of 500 μm, experimental observations). To address multiple flow lines with the plexer, pneumatic pressure was disconnected from individual control lines until the least significant plexer line pair corresponded to the desired number of columns to be stacked. For example, when the author stacked four columns in parallel, the control lines boxed in blue in Fig. 5.6 (two plexer line pairs, four total control lines) were disconnected from their respective pressure sources.

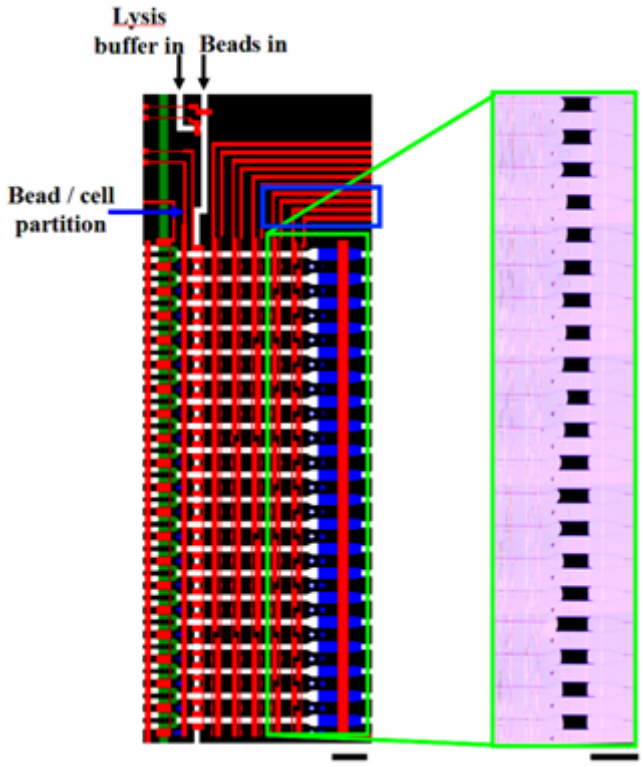


Figure 5.6. Column construction utilizing a fluidic multiplexer. **Left:** AutoCAD drawing of the bead loading module. Red are 24 μm high, rectangular cross-section control channels (SU8-2025); white are flow channels (12 μm high, semicircular cross section molded from Spr 220-7); green are flow channels (AZ-50 photoresist, 40 μm high, semicircular cross section) for cell loading, blue is the column construction area (SU8-2010, 10 μm high, rectangular cross section). The bead/cell partition control channel is closed when beads are stacked. Scale bars are 600 μm. **Right:** Optical micrograph of 20 stacked affinity columns.

5.3.1.2 Cell trapping and lysis

The author's scheme to trap and lyse cells also evolved from the first iteration of the highly parallel cells-to-cDNA device. The first implementation employed sieve

valves present on a 10 μm high flow structure (Fig. 5.7). In this scheme, cells were flown down a serpentine channel, open to atmosphere downstream of the sieve valve. The idea here was that cells would be trapped in the opening(s) of the sieve valves in a serial manner, with reactors closest to the "cell inlet" addressed first (see ch. 7 for microfluidic channel resistance discussion). The only valves downstream of the cell trapping valve were sieve valves for column construction, so flow was always open to atmosphere. This implementation proved ineffective for trapping single cells. Cells were seen escaping through the opening of the sieve valves, most likely due to a remodeling of their respective actin cytoskeletons. Although escaped cells would eventually be caught on the corresponding stacked bead column, and eventually lysed, lysis would no longer be uniform throughout the device. Accordingly, the author changed the cell trapping scheme to no longer include sieve valves. Instead, a 3 μm portion of SU8 molded flow structure was placed where the 10 μm structure was in the previous iteration. The remaining cell loading components remained the same. Again, cells were seen remodeling their cytoskeletons to escape through the small openings.

Finally, the author decided to close the cell trapping module off from the atmosphere (bead/cell partition control channel, Figs. 5.6, 5.7) and to partition simply a cell suspension with valves. Although not all reactors will be addressed with single cells in this scheme, a sufficient number (1 out of 3) should contain single cells, if the suspension is diluted correctly. This scheme worked well in the author's hands, and was implemented in the most mature device discussed in this chapter. At first, lysis by diffusion (Fig. 5.7B) was implemented because single cells were visualized to disappear in a matter of seconds, when exposed to lysis buffer. However, this scheme was changed

to an active mixing protocol (Fig. 5.7C) after beads were visualized to stick in channels when released to an outport. In order to achieve a working device with mixers, the author needed to determine the optimal valve geometry for pumping valves as well as the reaction and module partition valves. Although 200 μm X 200 μm is typically a sufficient valve geometry for a 40 μm high flow channel, the surrounding valve geometry precluded valves this size from actuating at workable pressures. To solve the valve dimension problem, the author redrew the mixing ring portion of the cells-to-cDNA device in AutoCAD, varying dimensions of the "cells in" valve (Fig. 5.8). The full cells-to-cDNA process was simulated by filling the region of the test chip to the left of the cell loading module (lysis buffer in) with food dye. Next, the author filled the "cells in" portion of the chip with food dye, and closed the "cells in" valve array. Consideration was paid to whether mixing occurred between modules and, if not, the minimum actuation pressure of the valve array. After testing six geometries, it was determined 225 μm X 150 μm was the optimal valve dimension. Readers are directed to Ch. 2 for a further discussion on lysis by diffusion and mixing, as well as ch. 7 for a solution to the problem of addressing 100% of the reactors on-chip with single cells.

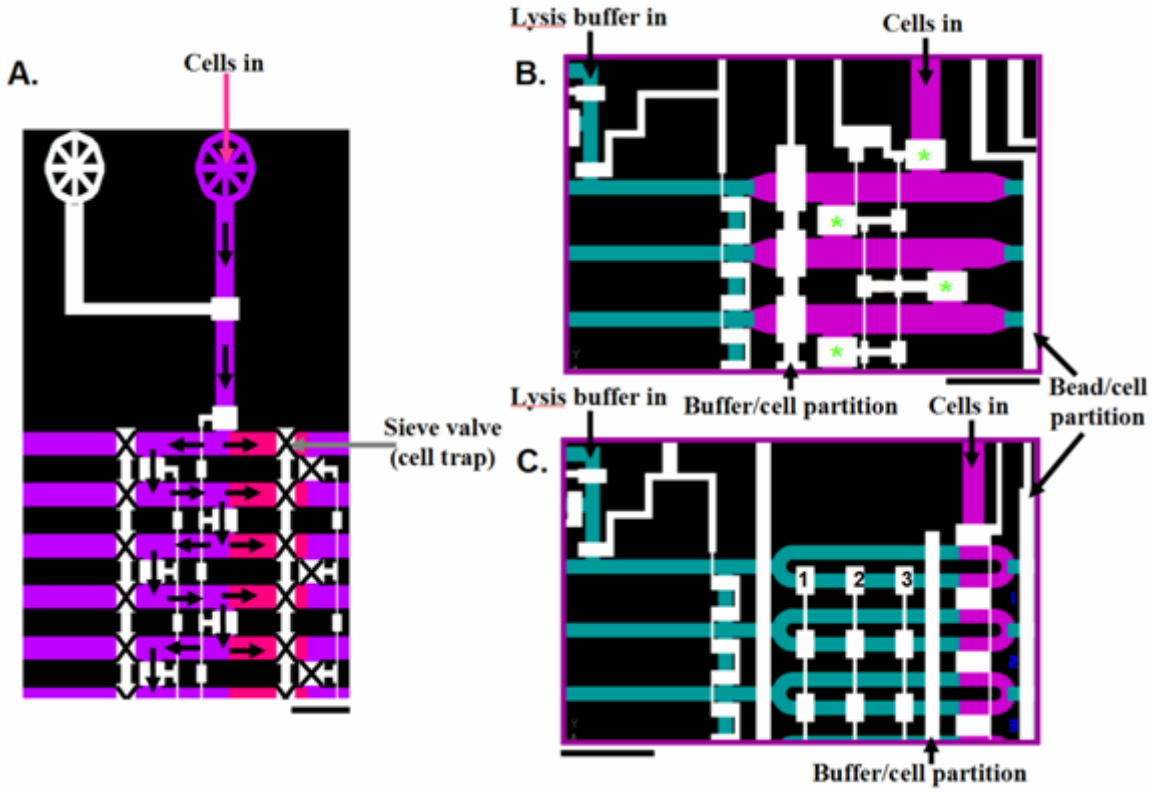


Figure 5.7. Cell trapping schemes. Purple channels are replica molded from 40 µm high AZ-50, cyan channels signify 12 µm high channels molded from AZ-50, and the pink channels represent 10 µm high channels cast from SU8 2010 photoresist. Scale bars are 600 µm. **A.** Cells enter a serpentine channel, where the right portions of the channels open to atmosphere. **B.** Cells enter a serpentine channel, opened to atmosphere at the end of the channel. Once the channel is filled with cell suspension, reactors are portioned off by closing the valves marked with asterisks. Cells are then lysed by opening the "buffer/cell partion" valve. **C.** Cells are partitioned the same way as **(B)**, but are actively lysed in a ring structure by executing a peristaltic pump sequence with the valves marked 1, 2, and 3.

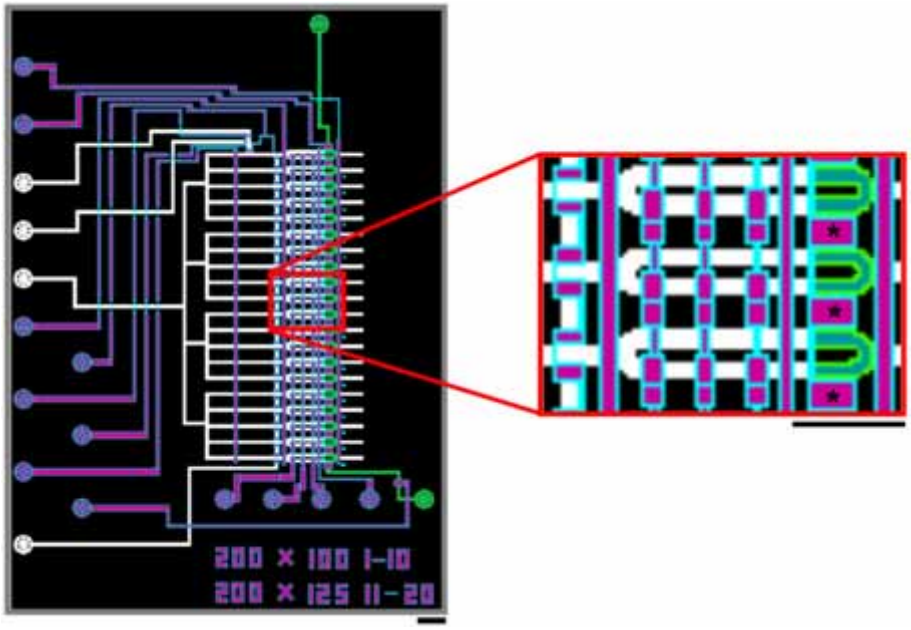


Figure 5.8. Testing for optimal valve geometry in deep, parallel mixing rings. AutoCAD drawing of the chip. Scale bar is 600 µm. Six devices of this size fit on a 3'' silicon wafer. Purple lines represent 25 µm high control channels (SU8-2025), white lines represent 12 µm high channels molded from Spr220-7 resist, and green line depicts 40 µm high flow channels cast from AZ-50. A total of six geometries were tested for the valve array present on the AZ-50 channel (control width X flow width): 1) 200 µm X 100 µm, 2) 200 µm X 125 µm, 3) 200 µm X 150 µm, 4) 225 µm X 125 µm, 5) 225 µm X 150 µm, 6) 225 µm X 175 µm. **Inset:** Geometries of the valves marked with asterisks and the green flow channel were varied by the dimensions listed above.

5.3.1.3 Waste and collection outputs

The early iterations of the densely integrated cells-to-cDNA chip employed the same output for both beads and waste. It was reasoned that further chip complexity was not needed because when the beads were retrieved from the chip, a wash step could

simply be implemented. At first, the author tried to flow beads into a 650 µm output connected to tygon tubing via steel pin interfaces (Fig. 5.9A). This proved unsuccessful because gravity kept the beads in the collection port. Next, the author cut the output lines so product could be flown directly onto the glass slide the chip was mounted on (Fig. 5.9A, blue line). This implementation was far from ideal, as the output clogged with cell debris and lysis buffer on most occasions. When beads were able to be flowed onto the glass slide, the author was unable to collect them all because there was no way to confine the flow. This is problematic, as it could cause inter-reaction contamination. The author then revisited the implementation discussed in chapters 2 and 3. Briefly, flow was directed to waste ports and collection ports, depending on the stage of the process (Fig. 5.9B). In retrospect, the author should not have tried new exotic implementations without trying what had worked well previously.

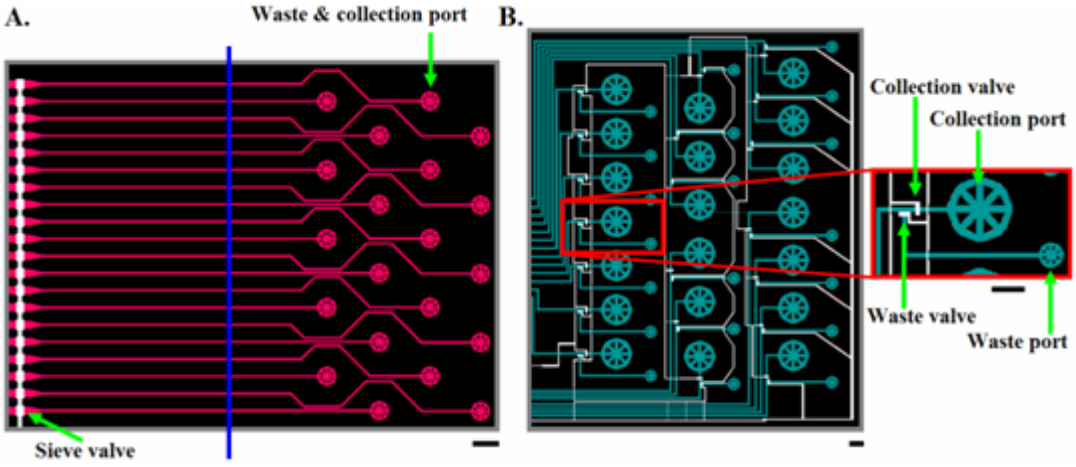


Figure 5.9. Outports for waste and products. All are AutoCAD drawings. Pink: 10 µm high flow channels molded from SU8. White: Control channels cast from 24 µm high SU8. Cyan: 12 µm high flow channels molded from reflowed Spr220-7 resist. Scale bars are 600 µm. **A.** The author's original scheme for diverting waste and products.

Products and waste were delivered to outports either connected to tygon tubing though 23 gauge steel pins or onto a glass slide by cutting the channels where the blue line is in the figure. **B.** The mature waste and product delivery scheme. Waste is diverted to a separate outport by opening the waste valve and closing the collection valve. When products are ready to be collected, the “waste valve” is closed and “collection valve" is opened. Beads are then retrieved via pipetting with gel loading tips.

5.3.2 Negative master mold fabrication for mature devices

All photomasks defining device features are designed with AutoCAD software (AutoDesk, Sausalito, CA), and printed at a resolution of 20,000 dots per inch on transparency films (CAD/Art services). In all optical lithography processes, 3’’ silicon wafers are utilized as substrates; and mold exposures are under UV light (365-405 nm) on an MJB mask aligner (7 mW/cm^2) for the indicated time.

Twenty-four-μm-high features present on control molds are fabricated with a single lithographic step. SU8-2025 (Microchem) is spun on a wafer (3,000 rpm, 45 s.), baked before exposure to evaporate excess solvent (2 min./5 min. at 65°C /95°C), exposed under a negative mask for 75 s., baked after exposure (2 min./5 min. at 65°C/95°C) to facilitate additional resist polymerization, and developed in NanoSU8 developer (Microchem). Once features are developed, the mold is baked again for 40 s. at 95°C to remove remaining solvent.

Three lithographic steps are undertaken for the flow molds. The first step is to define column construction flow channels (10 μm high). For these channels, SU8-2010 (Microchem) is spun onto a wafer (3,000 rpm, 45 s.), baked before exposure to evaporate

excess solvent (1 min./3 min. at 65°C/95°C), exposed under a negative mask for 45 s., baked after exposure (1 min./3 min. at 65°C/95°C), and developed in NanoSU8 developer (Microchem). Upon visualization of developed features, the mold is subjected to a 90 min. hard baked at 150°C. When the mold cooled to room temperature, the second step defining 12-μm-high features (output, bead and buffer delivery channels) is carried out. First, to promote resist adhesion, the mold is exposed to HMDS vapor for 2 min. Twelve-μm-high features are defined with Spr220-7 (Shipley). The resist is spun onto the mold (1,500 rpm, 60 s.), soft baked (90 s. at 105°C), aligned to the first layer (column construction flow lines), exposed under a positive mask for 3 min., and developed in 100% MF-319 developer (Microchem). The features are then annealed/hard baked for 3 hr. 200°C to facilitate channel rounding, necessary for valve closure, and also to chemically modify the resist. If the resist does not exhibit a chemical change (color turns from red to black), features will wash away during the third developing step. For the third step of the flow mold fabrication process, AZ-50 (Clariant) is spun onto the mold (1,600 rpm, 60 s.), after a two minute HMDS exposure. The mold is then soft baked (2 min./5 min. at 65°C/115°C), aligned to the second layer, exposed under a positive mask for 4 min., and developed in 25% 2401 developer (Microchem, diluted with 18 mΩ dI H_2O). Once features were developed, the mold was annealed/hard baked for 3 hr. 200°C for the aforementioned reasons. Additionally, if the AZ-50 resist does not undergo a chemical change (color turns from red to black), the features will crack when subjected to subsequent soft lithography baking steps.

5.3.3 Device fabrication process

All devices are fabricated by multilayer soft lithography (MSL)[19] with the silicone elastomer polydimethylsiloxane (PDMS, General Electric). Each device employs push-up valve geometry[68] and is a three layer elastomeric structure bonded to an RCA-cleaned microscope slide. Negative master molds are first exposed to chlortrimethylsilane (TMCS, Aldrich) vapor for 2 min. to promote elastomer release from molds after baking steps. Thirty g of liquid PDMS (5 parts A:1 part B) is poured onto the flow master, degassed under vacuum, and baked for 45 min. at 80°C. Liquid PDMS (20 parts A:1 part B) is spun onto the control master mold (1,800 rpm for 60 s.) and allowed to settle for 30 min. in order to obtain a uniform elastomer membrane on top of the control features. The mold is then baked for 30 min. at 80°C. Upon completion of the baking steps, the partially cured flow layer is peeled from its mold and 650 μm diameter flow channel access holes punched (Technical Innovations, part# CR0350255N20R4). The layer is then aligned to the partially cured control layer, still on the control master. The two layer structure is then baked for 45 min. The third layer, a featureless elastomeric membrane, is fabricated by spinning liquid PDMS (20 parts A:1 part B) onto a clean silicon wafer (1,600 rpm for 60 s.) followed by baking for 30 min. at 80°C. Once baking is complete, the two layer structure is peeled from the control master, control channel access holes punched (Technical Innovations, part# CR0350255N20R4) and mounted onto the third partially cured PDMS layer. The three layer structure is then baked for 3 hr. at 80°C. The assembled three layer structure is then peeled from the clean silicon wafer, output holes punched (Technical Innovations, part# CR0830655N14R4), cut to size and bonded to an RCA-cleaned microscope slide and baked overnight at 80°C.

5.3.4 General device operation

The on-off valves within each device are controlled by individual pressure sources (Fluidigm) and are interfaced via 23 gauge pins (New England Small Tube) and tygon tubing (VWR). We utilized an NI-DAQ card through a Labview interface (National Instruments) to actuate the pressure sources.

5.3.5 20X and 50X cells-to-cDNA device process flow

An AutoCAD drawing of a 20X device is given in Fig. 5.10. The 50X device is a scaled up version of the 20X device, with the same control and flow channels. However, two additional multiplexer control channels are necessary for the 50X chip because of the additional reactors (Fig. 5.10, 5.11). Lysis buffer is loaded into the white flow channels through inlet 2 (Fig. 5.10) until it reaches the waste outlets, so as to leave no air bubbles. Beads are then loaded into inlet 3 and columns are built serially by addressing flow lines individually with the multiplexer control channels, while keeping the sieve valve (valve present on blue flow channels) actuated. Once columns are built, excess beads still present in the flow channels are pushed with lysis buffer (inlet 2) to the constructed columns. Cell suspension is then loaded into inlet 1 (Fig. 5.10), and once the suspension flows to its outlet, the reactors are closed off by actuating the control channel marked with an asterisk. Cells are contained in the green portion of the mixing ring by keeping the vertical control channels to the east and west of the cell capture module closed. Lysis buffer is then loaded into the remaining portion of the device through inlet 4. Cells are then lysed chemically by mixing cells with lysis buffer in the ~10 nL ring. Mixing

occurs by executing a peristaltic pump sequence[21, 23] with control channels a, b, and c (Fig. 5.10). Cell lysates are then pushed via pneumatic pressure over the affinity columns, followed by washing of the columns with first strand synthesis buffer, dNTPs, and reverse transcriptase. Once the reverse transcriptase reaction fills the flow channels, first strand synthesis is then carried out by heating the device to 40°C on a thermal microscope stage with the beads utilized both as primers (oligo$(dT)_{25}$ sequences) and a solid phase support. The reaction mixture is flowed over the columns for 45 min. at a flow rate of ~20 µm/s. Upon completion of the reaction, the waste valves are closed, collection valves opened, and beads sent to output by opening the sieve valves and flowing columns off the device in a serial manner in PCR buffer, by using the fluid multiplexer to address reactors individually. Beads are collected by pipetting and stored at -20°C in qPCR buffer for future analysis.

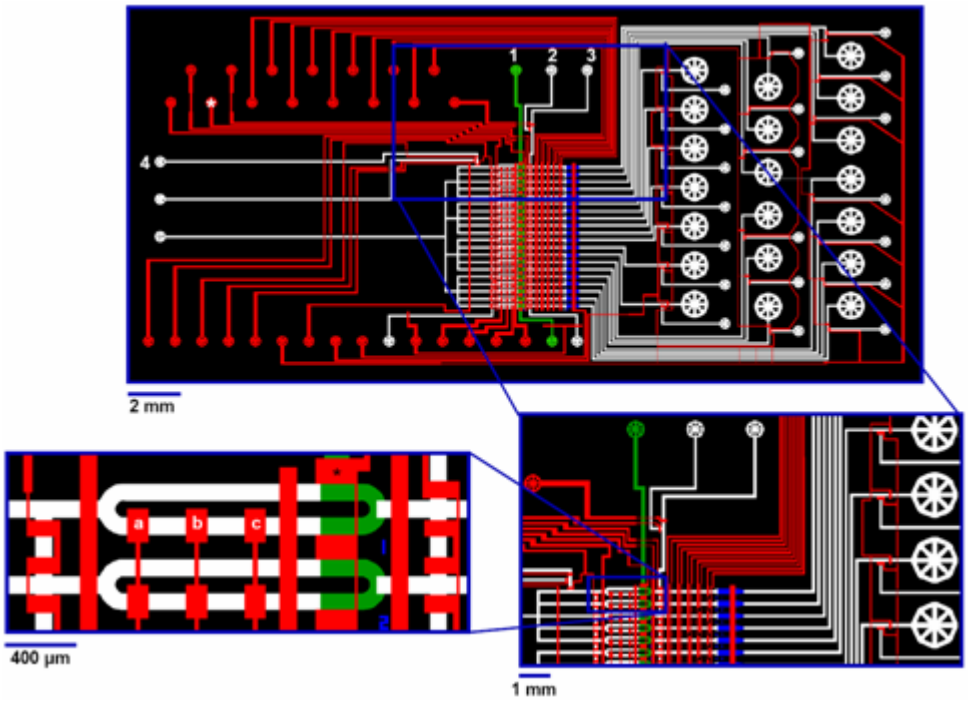


Figure 5.10. 20X cells-to-cDNA device. Control channels are in red, 10 μm high flow channel with rectangular cross section utilized for column construction is shown in blue, 12 μm high flow channels are shown in white, and the 40 μm high flow channel is shown in green. A chip capable of performing 50 reactions in parallel is fabricated with the same flow and control channel layouts.

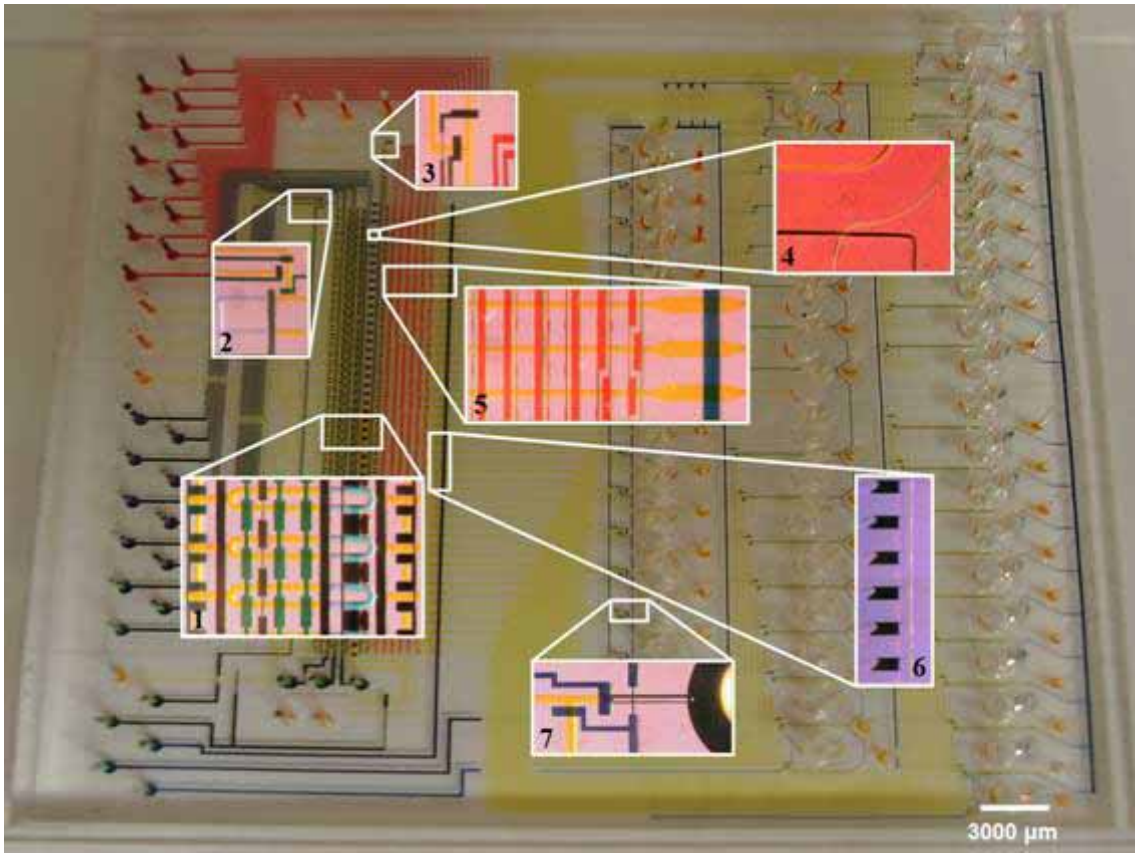


Figure 5.11. 50X cells-to-cDNA device filled with food dye. All flow channels are filled with yellow food dye, multiplexer control channels are filled with red dye, waste and collection control channels are in blue. **Inset 1**: Cell lysis module. Cells are portioned in the flow channels filled with blue dye. The pump valves are in green. **Inset 2**: Buffer inputs. The vertical valve in the picture is kept closed while loading buffer, to

compensate for any negative pressure in the channels caused by valves displacing volume. The back pressure compensation is needed to prevent contamination between reactors. **Inset 3**: Bead and lysis buffer inputs. Beads enter through one of the inlets and lysis buffer pushes residual beads present in channels to the sieve valves through the other inlet. **Inset 4**: Optical micrograph of one NIH/3T3 cell captured (corresponds to blue dye in inset 1). **Inset 5**: Multiplexer control channels and sieve valve channel. **Inset 6:** Six stacked bead columns. **Inset 7**: Output (1) and waste (2) valves, as well as a portion of the collection port.

5.3.6 Device calibration and efficiency measurements

In order to achieve absolute copy number determination from samples processed microfluidically, the author needed to calibrate and quantify the efficiency of each step along the microfluidic cells-to-cDNA process. The first step was to determine benchtop reverse transcriptase (RT) efficiency (results in ch. 4). Briefly, the author generated standard curves by benchtop qPCR or RT-pPCR, utilizing either Dap mRNA or Dap cDNA[69] as the template. The qPCR standard curve generated with the cDNA standards was then utilized to extrapolate measured copy number values for the mRNA samples. This was accomplished by plotting the Ct values obtained from the RT-qPCR standard curves onto the cDNA standard curve. Upon doing so, an average RT efficiency of 77% was achieved.

The author then quantified the microfluidic mRNA isolation step of the cells-to-cDNA process with a synthetic mRNA standard (Affymetrix). Although four synthetic mRNAs are present in the standard, only one (Dap) was probed for. The synthetic

mRNA was partitioned on-chip, mixed with lysis buffer, and run over affinity columns, either 50 in parallel (Fig. 5.12A red) or 20 (Fig. 5.12A black). Collection of the mRNA standards took place by releasing bead columns serially to output ports. It took an average of 3 min. per bead column for collection. Times varied because in some instances beads would stick in channels or disperse in the outport to a greater extent than other columns. Bead sticking is fixed by increasing the pressure applied to the column upon release. When plotting Ct values from the isolated microfluidic templates onto a Dap cDNA standard curve, a 40% efficiency is measured (Fig. 5.12B). Although the standard curve in Fig. 5.12A has an R value of -0.99, there was high variability (c.v. values of 3.58 and 3.37, Table 5.3) in efficiency values for the microfluidically isolated templates. The high disparity could be caused by 1) ambient RNases present during collection and the aforementioned variability in collection times; or 2) different interaction times of mRNAs with their respective affinity column. The author does not think it is the latter causing variability because columns are built with excess capacitance (200,000 oligo$(dT)_{25}$ sequences per bead); and lysates are pushed over the columns at a sufficiently slow velocity to allow for full interaction with the column. Similarly, there was no attempt to determine if time for collection was correlated to the variability observed. Instead, the author subjected identical concentrations of Dap mRNA to the additional microfluidic reverse transcription step, followed by the construction of another highly reliable standard curve (R = -0.99, Fig. 5.12C). The inclusion of the additional step decreased the C.V. values over six fold (Table 5.3). However, the variation observed is still greater than what was seen for samples processed in earlier studies (ch. 3). Next generations of devices will surely implement a binary tree architecture for lysis

buffer loading to assure equal path lengths (ch. 7). This implementation should solve the unequal sample interaction times with the affinity columns.

Microfluidic Process Efficiency Grid

mean, coefficient of variation	Dap cDNA	Dap mRNA → RT-qPCR	Dap → µF mRNA isolation → RT-qPCR	Dap → µF mRNA isolation → µF RT→ qPCR	Dap + 1 pg NIH/3T3 mRNA → µF mRNA isolation → µF RT→ qPCR
Dap cDNA	1,0	0.77	0.4, 3.58	0.18, 0.55	0.12, 0.91
Dap mRNA → RT-qPCR	X	1,0	0.54, 3.37	0.25, 0.55	0.15, 0.48
Dap → µF mRNA isolation → RT-qPCR	X	X	1,0	0.66, 0.57	0.37, 0.47
Dap → µF mRNA isolation → µF RT→ qPCR	X	X	X	1,0	0.68, 0.49
Dap + 1 pg NIH/3T3 mRNA → µF mRNA isolation → µF RT→ qPCR	X	X	X	X	1,0

Table 5.3. Microfluidic process efficiency grid. Ct values subjected to the processes listed in the columns two-five were plotted onto the standard curves listed in the rows of the first column. First number lists the mean efficiency and second number is the corresponding coefficient of variation for all templates tested.

When employing the additional on-chip step, an absolute process efficiency of 18% is obtained. The microfluidic reverse transcription step is 25% as efficient as benchtop RT (Table 5.3). Competing the Dap samples on-chip with 1 pg NIH/3T3 mRNA serves to decrease the absolute process efficiency to 12% from 18% (Fig. 5.11D, blue, Table 5.3, 5.4). On-chip mRNA isolation and first strand synthesis competition effects can be measured by plotting the Ct values obtained from Dap mRNAs loaded on-chip together with NIH/3T3 mRNA, onto the standard curve generated with uncompeted

Dap mRNA (Table 5.4). Upon doing so, a 32% loss in efficiency is calculated. An average c.v. of 0.91 was obtained for all mRNA samples (Fig. 5.12, Dap + NIH/3T3 mRNA) isolated and synthesized on-chip to cDNA (Table 5.3). When narrowing the range of templates to the expected range seen for the genes of interest (ch. 4) in single cells, a c.v. of 0.22 is obtained (Table 5.5). These numbers will serve as a baseline for determining if gene expression noise is exhibited in single NIH/3T3 cells.

Assessing Template Competition: Comparison of the two (Dap, Dap + NIH/3T3) microfluidic cDNA synthesis std. curves

	10 copies	100 copies	1k copies	10k copies	100, 1k	1k, 10k	100, 1k, 10k	all
avg eff	0.99	0.45	0.74	0.63	0.61	0.61	0.61	0.67
variance	0.36	0.03	0.017	0.018	0.043	0.035	0.035	0.11
std dev	0.6	0.17	0.13	0.14	0.21	0.19	0.19	0.33
coeff var	0.6	0.38	0.18	0.22	0.34	0.31	0.31	0.49

Table 5.4. Process efficiencies for individual Dap templates spiked with 1 pg NIH/3T3 mRNA. Efficiency is based on the standard curve generated with Dap cDNA templates.

Dap + 1 pg NIH/3t3 mRNA compared to qPCR std. curve

	10 copies	100 copies	1k copies	10k copies	100, 1k	1k, 10k	100, 1k, 10k	all
avg eff	0.14	0.067	0.15	0.1	0.11	0.13	0.11	0.12
variance	0.0079	0.0007	0.027	0.0005	0.017	0.015	0.012	0.011
std dev	0.089	0.026	0.17	0.023	0.12	0.12	0.11	0.11
coeff var	0.62	0.39	1.08	0.22	1.14	0.95	0.98	0.91

Table 5.5. Reverse transcription competition effects. The Ct values obtained from Dap + NIH/3T3 standards were plotted on the standard curve generated with Dap templates subjected to no competition.

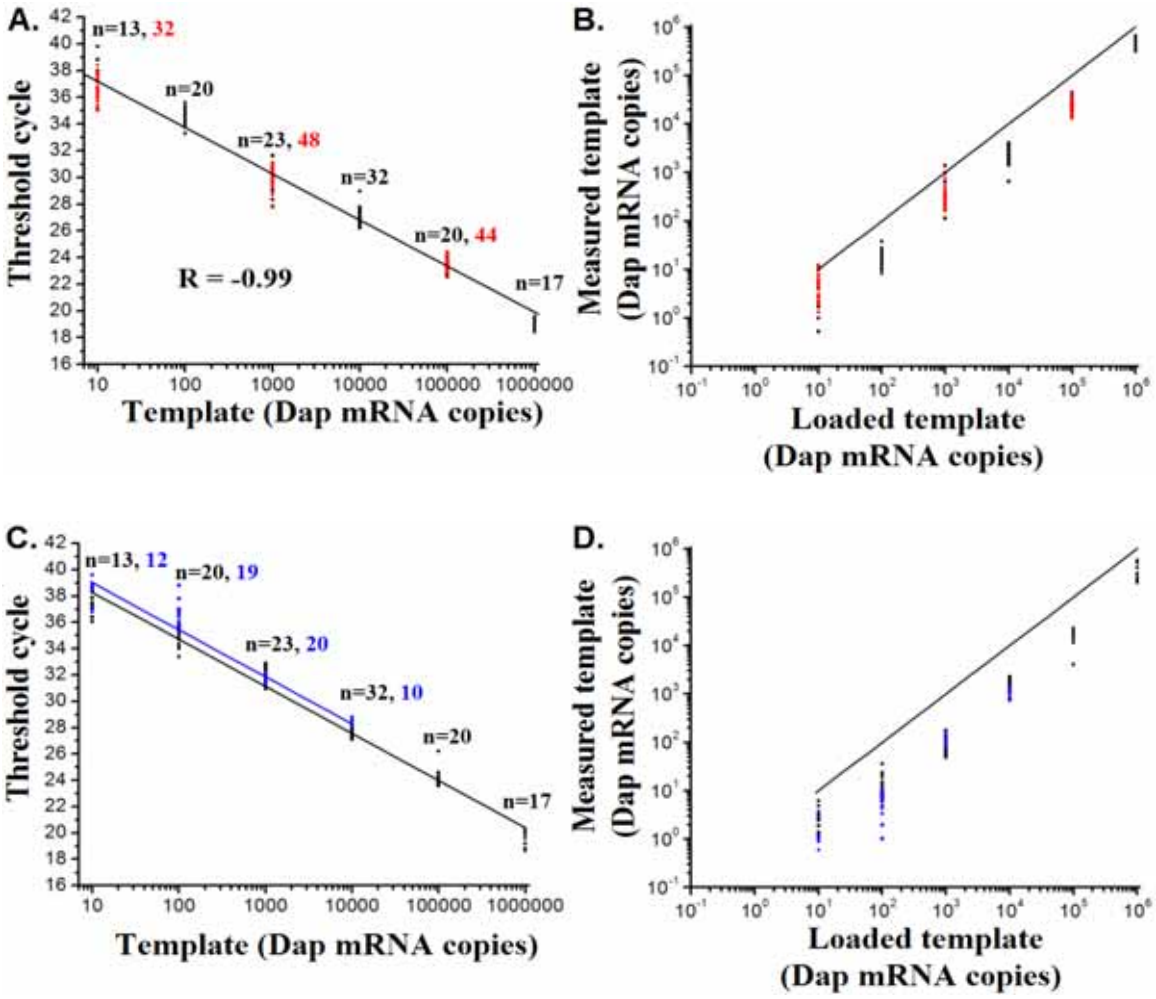


Figure 5.12. 20X and 50X cells-to-cDNA device and process calibrations. **A.** RT-qPCR standard curve (Y = 3.46(X)+40.64, R = -0.99) generated with Dap mRNA standards subjected to the microfluidic mRNA isolation → recovery process, either with 50X (red) or 20X (black) devices. Curve is fitted to the data from both devices. **B.** Microfluidic mRNA isolation → recovery efficiency measurements based on a qPCR standard curve (Y = -3.50(X) + 38.82, R = -0.99, Fig. 4.4). Average efficiency was 40%. **C.** qPCR standard curves generated with Dap mRNA standards subjected to the microfluidic mRNA isolation → first strand synthesis → recovery process. All standards were run

through 20X devices. The black curve (Y =-3.57(X) +41.84, R = -0.99) is fitted to Dap standards while the blue curve (Y = -3.58(X) +42.58, R = -0.99)

is fitted to Dap + 1 pg NIH/3T3 mRNA standards.

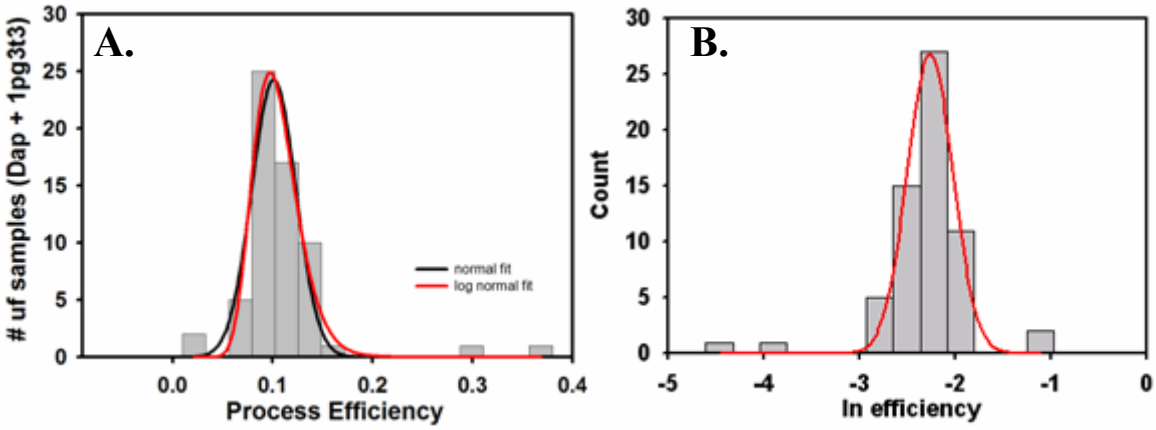


Figure 5.13. Efficiency distribution for the dap + 1pg NIH/3T3 mRNA templates. **A.** Black is normal fit and red is lognormal fit of the data. **B.** Data from (**A**) was ln transformed and fitted to a normal distribution.

In order to quantify cell lysis, specifically RNase activity during lysis, the author spiked cell suspensions off-chip with defined concentrations of the bacterial mRNA standard, loaded the suspension on-chip and subjected it to the 20X cells-to-cDNA process. Results to this study were inconclusive. Regardless of starting concentration samples were amplified the same extent, and at the same Cts as samples with no cells (Fig. 5.14, PBS + Dap mRNA in microfluidic reactor). These results can be explained by the RNase activity prevalent in PBS when not DEPC treated. Because the bacterial mRNAs are mixed directly with PBS, the mRNAs are not protected from buffer based RNases, unlike mRNAs inside a cell. The next set of experiments can either utilize DEPC treated PBS or design a chip-based scheme to carry out mixing of cell suspension,

Dap mRNAs and lysis buffer. On-chip mixing will protect the Dap mRNA from RNases because RNase inhibitor and denaturant is present in the lysis buffer.

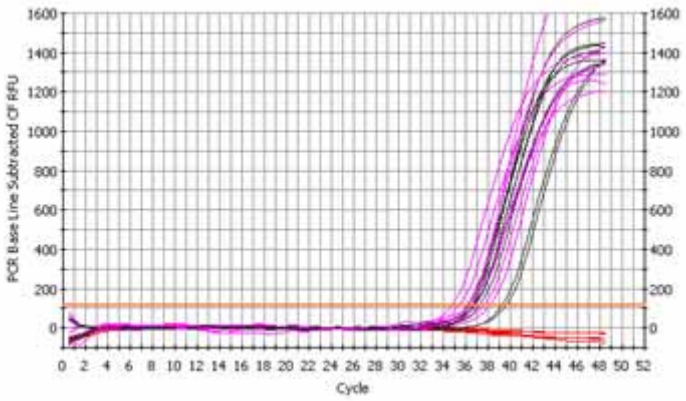


Figure 5.14. Amplifcation plots from Dap mRNA/single cell spike experiments. Red: Off-chip negative controls. Black: PBS + 10,000 copies Dap/microfluidic reactor. Magenta: Single cell + 10,000 Dap/microfluidic reactor.

Negative control experiments

The author verified qPCR signals obtained from on-chip calibrations and single cell measurements were real, and not a product of contamination, by performing negative control experiments. In order to rule out the possibility of the cell suspensions or chips being contaminated with nucleic acids, PBS from the suspensions was subjected to the microfluidic cells-to-cDNA process in parallel with reactors containing NIH/3T3 cells. qPCR was then run on positive and negative templates harvested from chips, and Ct values compared (Fig. 5.14A-C). As the figure illustrates, Ct values for no-template reactions are consistently higher than for single cell samples. Furthermore, the Ct values obtained for the no-template samples are consistent with values obtained for qPCRs with H_2O as template (Fig. 4.7). The plots shown in the figure are indicative of typical results

obtained from single cell and no-template samples harvested from 50X chips, after subjection to the microfluidic cells-to-cDNA process. The author then verified the calibrations presented earlier in the chapter by confirming Dap signal was produced from samples loaded on-chip, and not by some artifact or contamination. The cells-to-cDNA process without the reverse transcriptase enzyme was carried out on highly concentrated Dap mRNA templates (100,000 copies/reactor). Samples were then harvested and subjected to qPCR (Fig. 5.14D). No amplification was observed for any of the microfluidic samples, confirming the earlier calibration and efficiency values. Similar experiments were carried out on cells and NIH/3T3 mRNA, with the consistent results obtained (Fig. 2.8).

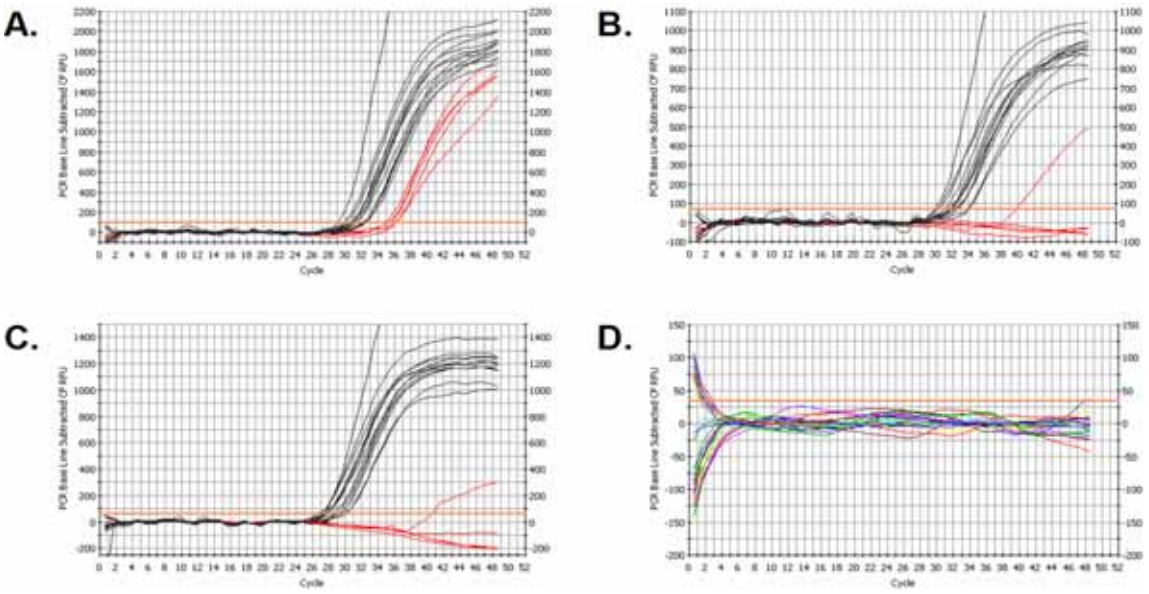


Figure 5.15. Amplification plots from negative control experiments. **A.-C.** Triplex qPCR was carried out on either single NIH/3T3 cell samples (black) or PBS from cell suspensions (red). **A.** β-actin (Hex) **B.** PGK (Cy5) **C.** GAPDH (Fam) **D.** Amplification plots for Dap samples subjected to the microfluidic cells-to-cDNA process sans reverse transcriptase.

5.3.7 Reagent preparation

NIH/3T3 preparation

NIH/3T3 cells were grown to ~70% confluency in Dulbecco's modified Eagle medium (DMEM)/10% calf serum (CS, Hyclone) in 12.5 cm^2 vacuum-gas plasma treated polystyrene culture flasks (BD Falcon). Cells were then washed in 1X PBS, trypsinized, and further diluted in culture medium to stop the trypsinization reaction. After counting the cells, the cell suspension was washed in 1X PBS and pelleted 1 X 10^4 cells per 15 mL conical tube(s) (1800 rpm, 5 min). Each pellet was resuspended in one mL DMEM/10% CS followed by the addition of two μL DyeCycle green (Invitrogen) to each suspension. The suspension(s) was then incubated for 30 min. in 37°C/10% CO_2 incubator.

qPCRs

All qPCRs were carried out on a BioRAD icycler instrument utilizing either Taqman probe or sybr green chemistry. The latter was employed to verify the absence of primer-dimers and to generate products for standard curve construction. The cycling protocols below were modified to include an additional one min. extension step at 72°C, where detection was also enabled, for sybr green qPCRs.

An 86 bp portion from the synthetic bacterial mRNA Dap (Affymetrix, accession # L38424) was amplified with the following protocol and oligonucleotides: 95°C for 15 min. and 50 cycles of 94°C for 15 s., 57°C for 1 min. Detection was enabled during the 57°C step. Reverse transcription was carried out first, either on-chip for 1 hr. at 37°C or 30 min. at 50°C as the initial PCR step. The following oligonucleotides were utilized:

Probe: 5'-**Fam**-TGT GAT GTG TAT TCC ATT CCG CTC GCC A **-BHQ_2**-3'
Sense primer: 5'-ACC GGA TGT CTC GGC ATT AAT C-3'
Antisense primer: 5'-GCA CAA GAA TTT CCG CAG TAC C-3'

PCRs on the mouse genes β-actin (accession # NM_007393, 99 bp), phosphoglycerate kinase (PGK, accession # M15668, 139 bp) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH, accession # NM_001001303, 124 bp) were carried out with the following protocol: 95°C for 15 min. and 50 cycles of 94°C for 15 s., 60°C for 1 min. Detection was enabled during the 60°C step. A modified protocol was implemented when detecting the three genes in the same reaction (above is for one gene detection): 95°C for 15 min. and 50 cycles of 94°C for 45 s., 60°C for 90 s. Detection was enabled during the 60°C step. The following oligonucleotides were employed:

β-actin:

Probe: 5'-**Hex-** CCA TCC TGC GTC TGG ACC TGG CTG GC -**BHQ_1**-3'
Sense primer: 5'- CCA TCT ACG AGG GCT ATG CTC TCC -3'
Antisense primer: 5'- CAC GCT CGG TCA GGA TCT TCA TG -3'

PGK:

Probe: 5'-**Cy5**- ACC ACA GTC CAA GCC CAT CCA GCC AGC A -**BHQ_2**-3'

Sense primer: 5'- ACC TTG CCT GTT GAC TTT GTC ACT G -3'

Antisense primer: 5'- CCT CGG CAT ATT TCT TGC TGC TCT C -3'

GAPDH:

Probe: 5'-**Fam**- CGT GCC GCC TGG AGA AAC CTG CCA AGT **–BHQ_2**-3'

Sense primer: 5'- CCA ATG TGT CCG TCG TGG ATC TG -3'

Antisense primer: 5'- TCC TCA GTG TAG CCC AAG ATG CC -3'

Products from a sybr green PCR were purified, and utilized to generate standard curves for all three genes (Fig. 5.15). These curves are utilized to determine copy numbers of the respective genes in single NIH/3T3 cells. In addition, melting curve analysis was carried out in order to determine how many products are generated with the above primers (Fig. 5.15C). This is necessary because quantification by spectroscopy is only effective if one product is present in the sample. The author employed triplex qPCR (combined three purified samples in one tube) with the PCR purified standards, so as to simulate the single qPCRs discussed in this chapter.

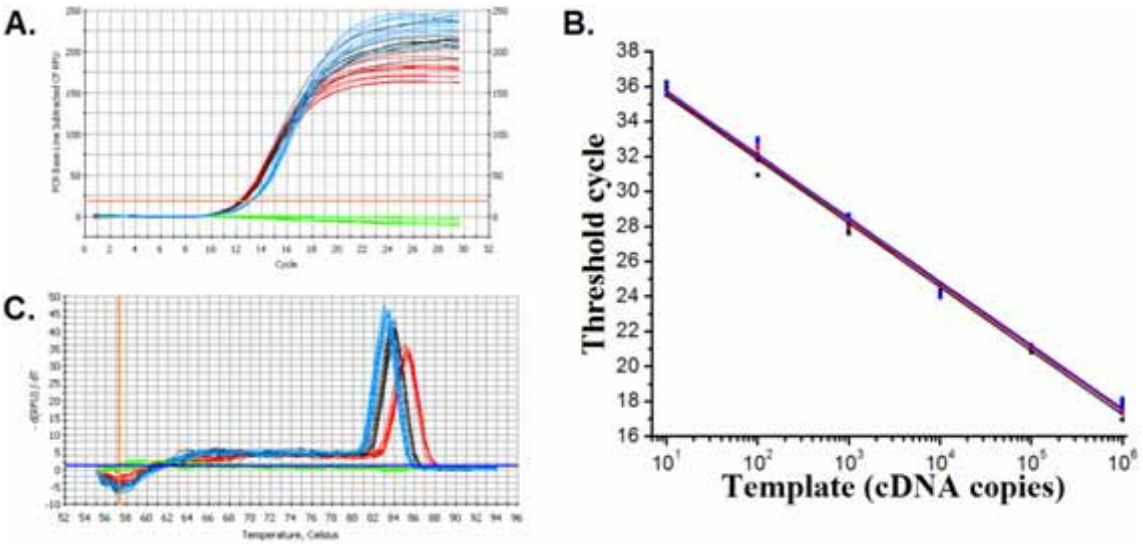


Figure 5.16. Triplex qPCR standard curve construction. Black: GAPDH, red: β-actin, blue: PGK **A.** Amplification plots from simplex sybr green qPCRs. **B.** Products from (**A**) were purified, quantified by spectrophotometry, and utilized as templates for triplex qPCR. GAPDH: Y = -3.65(X) + 39.13, R = -0.99; β-actin: Y = -3.63(X) + 39.23, R = -0.99; PGK: Y = -3.63(X) + 39.39, R = -0.99. **C.** Melting curve analysis of products generated in (**A**). Distinct primer sets yielded distinct products.

Chapter 6:

Parallel Picoliter RT-PCR Assays Using Microfluidics

Anal. Chem. 2006, *78*, 956–958

Parallel Picoliter RT-PCR Assays Using Microfluidics

Joshua S. Marcus,[†,‡] W. French Anderson,[‡,§] and Stephen R. Quake*[,‡,⊥]

Option in Biochemistry and Molecular Biophysics, Department of Applied Physics, California Institute of Technology, MS 128-95, Pasadena, California 91125, and Gene Therapy Laboratories, Keck School of Medicine, University of Southern California, Los Angeles, California 90033

The development of microfluidic tools for high-throughput nucleic acid analysis has become a burgeoning area of research in the post-genome era. Here, we have developed a microfluidic chip to perform 72 parallel 450-pL RT-PCRs. We took advantage of Taqman hydrolysis probe chemistry to detect RNA templates as low as 34 copies. The device and method presented here may enable highly parallel single cell gene expression analysis.

Although there are numerous approaches available for detecting the presence of nucleic acids, the methods most commonly utilized are PCR and reverse transcriptase-PCR (RT-PCR) for DNA and mRNA, respectively.[1] Microfluidic technology offers added advantages for PCR assays, including, but not limited to, decreased cycling times due to high surface-to-volume ratios and, therefore, rapid thermal energy transfer, smaller reaction volumes, portability, and lack of contamination issues.[2] Our group has previously demonstrated microfluidic devices with integrated micromechanical valves as a viable solution to perform parallel PCR in volumes as small as 3 nL.[3,4] Although microfluidic PCR has been demonstrated by numerous groups,[2] there has been little previous work on integrating the reverse transcription enzymatic step into microfluidics.[5−7] In previous studies, material was removed from the chip for detection, starting templates were on the order of 10 million gene-specific RNA molecules, and RT reaction volumes were 5 μL.

Here, we have fabricated microfluidic devices by multilayer soft lithography (MSL)[8] to demonstrate sensitive, highly parallel RT-PCR. We have capitalized on the advantages of small reaction volumes available with microfluidic assays and performed 72 parallel RT-PCR reactions in 450-pL reactors using a design suggested by Unger et al.[9] Using endpoint detection, we were able to detect less than 50 β-actin transcripts from a total RNA template and, thus, achieved the detection sensitivity[10] required for single cell analysis.

EXPERIMENTAL SECTION

Device Fabrication. All devices were fabricated using the process of MSL.[8] Devices were composed of three layers of the silicone elastomer poly(dimethylsiloxane) (PDMS) bonded to a glass coverslip and were fabricated as previously described,[11] but with push-up valve geometry.[12] Negative master molds were fabricated out of photoresist by standard optical lithography and patterned with 20 000 dpi transparency masks (CAD/Art Services). SU8-2025 (Microchem) (24 microns high) was utilized for the control molds. The flow master molds were fabricated out of 10-μm features that were defined in 5740 photoresist (Shipley).

General Device Operation. The on−off valves within each device are controlled by individual pressure sources (Fluidigm). The pressure sources are actuated by a NI-DAQ card (National Instruments) and Labview graphical interface.

RT-PCR Device Operation. Each chip contains eight flow channels which are filled in a west to east fashion with a master RT-PCR reaction. There are two control channels, which form valves when crossing their respective flow channels. When initially filling a specific flow channel, the vertical control channel (Figure 1, green) is opened to allow for purging of the respective flow channel with RT-PCR cocktail. The vertical channel is then closed, and the horizontal reaction partition control channel (Figure 1, blue) is opened, allowing for dead end filling of the nine 450-pL reactions. The reaction partition control channel is closed once the reactors are filled. The 50-μm-wide guard channels[9] (Figure 1, channels filled with yellow food dye) are then pressurized to prevent flow channel evaporation. All control and guard channels are dead-end filled with water.

RNA Amplification and Detection. We utilized human male total RNA as template and probed for a 240-bp portion of the

* To whom correspondence should be addressed. E-mail: quake@stanford.edu.
† Option in Biochemistry and Molecular Biophysics, California Institute of Technology.
‡ Department of Applied Physics, California Institute of Technology.
§ University of Southern California.
⊥ Current address: Bioengineering Department, Clark Center, Stanford University, E300, Stanford, CA 94305.

(1) Bustin, S. A. *J. Mol. Endocrinol.* 2000, *25*, 169−193.
(2) Auroux, P. A.; Koc, Y.; deMello, A.; Manz, A.; Day, P. J. *Lab Chip* 2004, *4*, 534−546.
(3) Liu, J.; Enzelberger, M.; Quake, S. *Electrophoresis* 2002, *23*, 1531−1536.
(4) Liu, J.; Hansen, C.; Quake, S. R. *Anal.Chem.* 2003, *75*, 4718−4723.
(5) Anderson, R. C.; Su, X.; Bogdan, G. J.; Fenton, J. *Nucleic Acids Res.* 2000, *28*, E60.
(6) Obeid, P. J.; Christopoulos, T. K. *Anal. Chim. Acta* 2003, *494*, 1−9.
(7) Obeid, P. J.; Christopoulos, T. K.; Crabtree, H. J.; Backhouse, C. J. *Anal. Chem.* 2003, *75*, 288−295.
(8) Unger, M. A.; Chou, H. P.; Thorsen, T.; Scherer, A.; Quake, S. R. *Science* 2000, *288*, 113−116.
(9) Unger, M. A.; Manger, I.; Lucero, M.; Yong, Y.; Miyashita-Lin, E.; Wienecke, A.; Facer, G.: European Patent Vol. EP1463796, 2004.
(10) Bishop, J. O.; Morton, J. G.; Rosbash, M.; Richardson, M. *Nature* 1974, *250*, 199−204.
(11) Hansen, C. L.; Sommer, M. O.; Quake, S. R. *Proc. Natl. Acad. Sci. U.S.A.* 2004, *101*, 14431−14436.
(12) Studer, V. H., G.; Pandolfi, A.; Ortiz, M.; Anderson, W. F.; Quake, S. R. *J. Appl. Phys.* 2004, *95*, 393−398.

10.1021/ac0513865 CCC: $33.50
Published on Web 01/04/2006

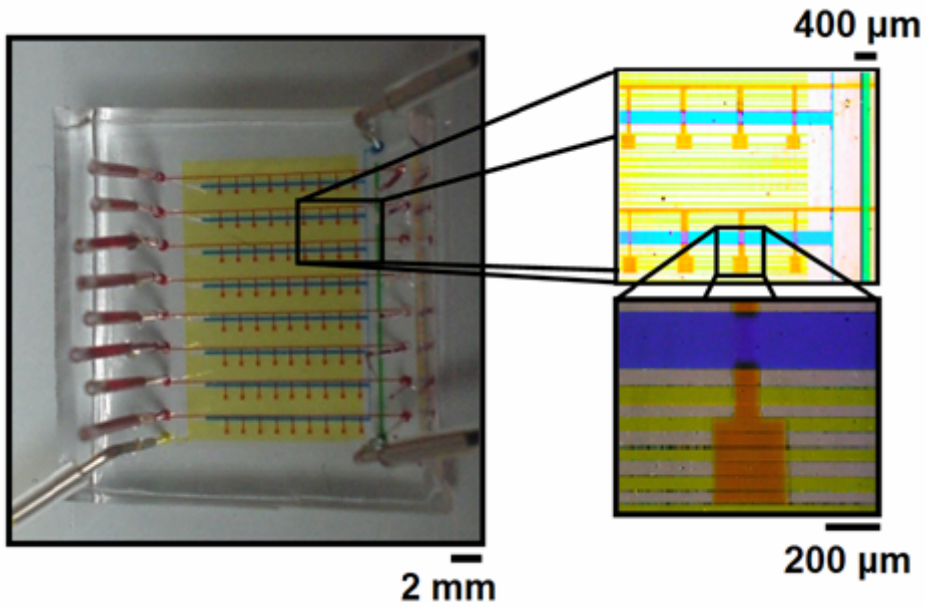


Figure 1. RT-PCR device. **Left**: Photograph of the device loaded with food dye. Guard channels, which supply water vapor to the porous PDMS, are filled with yellow dye, flow channels are filled with orange dye, and the two control channels are filled with blue (horizontal control channel) and green dye (vertical reaction partition control channel). **Right**: Optical micrographs of eight reaction chambers (upper) and one reaction chamber (lower).

β-actin gene using primers and Taqman probe supplied in an Applied Biosystems kit (Part no. 401846). Each probe contains a reporter dye (Alexa 488) covalently attached to the 5′ end and a quencher dye covalently attached to the 3′ end of the probe. Fluorescence resonance energy transfer (FRET) between the two dyes enables the shorter wavelength reporter to be quenched by the longer wavelength dye when the probe is intact. The 5′-to-3′ exonuclease activity of *Taq* polymerase cleaves the probe and thereby increases the distance between the two dyes, and in so doing augments the fluorescence emission of the reporter. Once the chip was loaded, thermal cycling was performed on a flat plate thermal cycler according to the following protocol: 50 °C for 30 min, 95 °C for 10 min, and 30 cycles of 94 and 58 °C for 20 and 50 s, respectively. The reaction chambers on the chip were then probed for fluorescence with a modified flat-bed ArrayWorx scanner (Applied Precision).

RESULTS AND DISCUSSION

We fabricated microfluidic devices with the purpose of performing RT-PCRs in parallel picoliter reactions. A prototype device, with its control and flow lines filled with food dye, is shown in Figure 1. Each chip has eight flow lines, so multiple templates and master mixes can be varied up to 8-fold on a single device. This multiplexing ability allows for standard curve generation, utiliziation of multiple primer sets, and microfluidic reaction optimization on a single device. The RT-PCR cocktail is mixed off-chip, loaded onto the device in a west to east manner, and divided into nine 450-pL chambers, for a total of 72 chambers per chip. Because the chip's chambers have a geometrically defined volume, error due to pipetting variability is not an issue.

Initial experiments with these devices demonstrated proof of principle. We verified the PCR chemistry and cycling parameters by performing PCR on a β-actin cDNA template and detected DNA templates ranging from 6 to 18 000 molecules (data not shown).

We then utilized three devices to demonstrate RT proof of principle and to determine RNA detection sensitivity. Results from these experiments are given in Figure 2. Figure 2a shows a fluorescence scan of a chip utilized to verify RT feasibility. We kept the RNA template constant at 50 ng/µL (Figure 2a, middle four lines) and employed the first two and last two lines as no-template controls. We then used a hydrolysis probe to detect a segment of the β-actin gene, which is present in high abundance[11] inside most cells, with copy numbers ranging from 2000 to 3000 per individual cell.[13] When considering the amount of mRNA in one cell is ~1 pg and 1% of total RNA,[14] we estimate the positive template reactors to contain 680 β-actin copies each. The scan shows a 2.4-fold difference in fluorescence between the positive and negative template reactors.

Next, we determined the sensitivity of the device by varying the total RNA template concentration. RNA was serially diluted to establish a range of 0−4 pg of total RNA per reaction chamber (Figure 2b). Using the same reasoning as above, we estimate the β-actin copy numbers per reactor to be 0, 34, 68, and 136 copies, respectively. Figure 2b shows a graph of mean fluorescence values for the respective chambers of a chip after RT-PCR. This detection sensitivity is comparable to conventional real time RT-PCR in our hands (50 copies/reaction, Figure S-1 of the Supporting Information). There was an increase in fluorescence up to ~50 copies/chamber, followed by a plateau region. We interpret the fluorescence plateau at higher template concentrations to mean our reaction conditions were not optimal or we were detecting in the plateau phase of the reaction. If conditions were not optimized, it is reasonable to speculate reagents depleted before the cycling ended. Accumulation of product may have inhibited further synthesis of the β-actin gene segment. Nevertheless, the mean fluorescence values for the different positive templates were significantly different from the no-template control, as measured

(13) Adams, M. D.; Kerlavage, A. R.; Fleischmann, R. D.; Fuldner, R. A.; Bult, C. J.; Lee, N. H.; Kirkness, E. F.; Weinstock, K. G.; Gocayne, J. D.; White, O.; et al. *Nature* **1995**, *377* (Suppl. 6547), 3−174.
(14) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 4th ed.; Garland Publishing: New York, 1994.

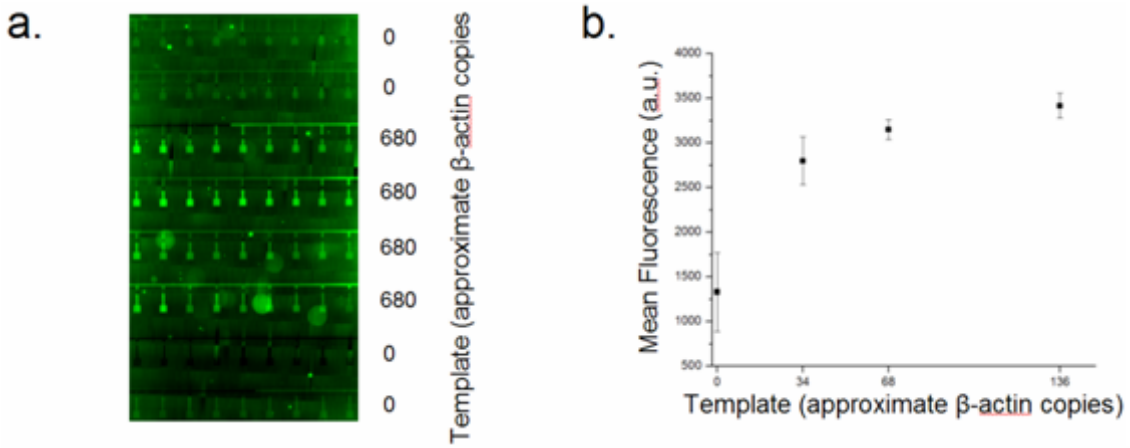


Figure 2. RT-PCR results. A. Scan of a chip after RT-PCR. We are detecting fluorescence at 530 nm, the emission wavelength of the reporter dye (FAM) attached to the Taqman probe. B. Mean fluorescence data from the low-end serial dilution experiment. Points are mean fluorescence of chambers ± standard error after RT-PCR.

by paired *t*-tests. To rule out the possibility of false positive results due to genomic DNA (gDNA) contamination, we ran off-chip real time RT-PCRs with no RT enzyme along with serially diluted human male total RNA templates. Amplification was limited to reactions with 500 pg or 50-pg templates (Figure S-1). We calculated the percent of gDNA contamination by plotting the Ct values of the amplified no-RT reactions onto a standard curve generated with the RT enzyme (Figure S-1). We then divided the calculated template by the initial template present in the respective reaction and arrived at a value of 0.3% for gDNA contamination. Therefore, 333 β-actin molecules would need to be present in a reaction chamber to achieve amplification of 1 gDNA molecule.

It has become popular to use real time PCR, in which detection of products is performed after each thermal cycle, to obtain quantitative information about the amount of starting material. In the work described above, detection was performed at the end of thermal cycling, but it would be straightforward to modify the scanner to include a thermocycling capability such that real time data could be acquired from our device.

CONCLUSIONS

These results demonstrate that RT-PCR assays can be implemented with success on a microfluidic platform, in a regime useful for single cell analysis. We were able to detect a signal for low-abundance mRNA templates in a parallel fashion. The ability to detect low and medium copy transcripts, as presented here, may enable future studies on gene expression in individual cells.

ACKNOWLEDGMENT

The authors thank Kathy Burke, Sebastian Maerkl, and Carl Hansen for helpful discussions. This work was supported in part by a National Research Service Award (T32GM07616) from the National Institute of General Medical Sciences (J.S.M.) and National Institutes of Health (NIH) Grant NIH 1RO1 HG002644-01A1.

SUPPORTING INFORMATION AVAILABLE

Additional information as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review August 3, 2005. Accepted November 24, 2005.

AC0513865

Supporting Information for
Parallel picoliter RT-PCR assays using microfluidics

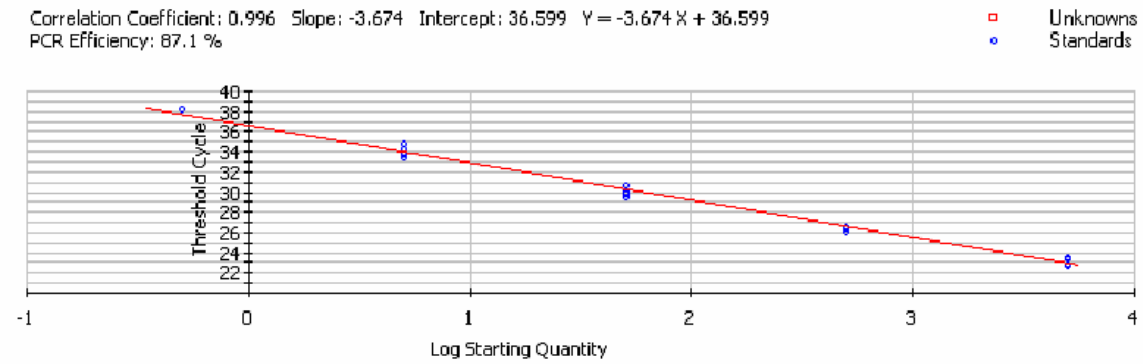
Joshua S. Marcus[1,2], W. French Anderson[2,3], and Stephen R. Quake[2,4,5]

[1]Option in Biochemistry and Molecular Biophysics, [2]Department of Applied Physics, California Institute of Technology, MS 128-95, Pasadena, CA 91125. [3]Gene Therapy Laboratories, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033
4: To whom correspondence should be addressed. Email: quake@stanford.edu
5: Current address: Bioengineering Dept., Stanford University, Clark Center, E300, Stanford, CA 94305

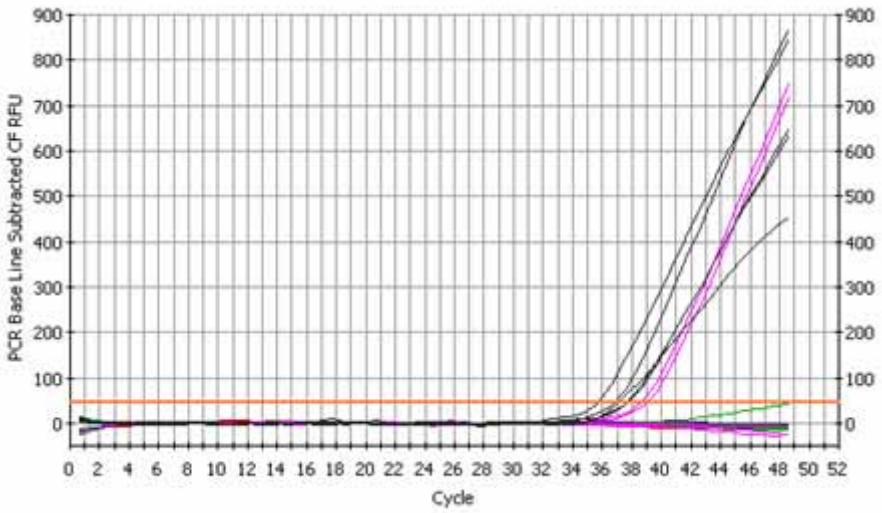
Table of Contents

Figure S-1

a.



b.



Supporting Figure 1: **a.** Benchtop RT-qPCR standard curve utilizing serially diluted human male total RNA as template. We were able to detect the 0.5 pg template in only one of five reactions. **b.** Fluorescence traces of RT-PCR control reactions. Red curves are no-template reactions. All other curves are no RT reactions with serially diluted human male total RNA templates. Templates range from 0.05 pg total RNA (gray) to 500 pg total RNA (black). Only reactions with 500 pg total RNA and 50 pg total RNA (magenta) templates amplified. All other templates (5 pg total RNA, green; 0.5 pg total RNA, blue) did not amplify. We can estimate the percent of gDNA contamination by plotting the Ct values of the no RT reactions onto the off-chip standard curve shown in **a**. When this is done, gDNA contamination is estimated to be 0.3% of total RNA template.

Chapter 7: Advances in Microfluidic chip fabrication and operation

The development of multilayer soft lithography (MSL) by Quake and coworkers[19] has laid the foundation for complex fluid manipulation on the nanoliter scale. This chapter describes the technique of MSL in detail and further discusses the author's contributions to MSL fabrication and device operation.

7.1 Multilayer soft lithography background

MSL,[19] which builds upon the replica molding technique of soft lithography,[70] enables the construction of true sealing valves on-chip. Silicone rubbers have become the material of choice for MSL fabrication because of their optical properties, biocompatibility, and low Young's modulus. In general, fabrication with silicone rubbers, for example General Electric RTV 615, proceeds by mixing two parts of liquid elastomer, which then will cure into a flexible solid in the presence of heat and a platinum catalyst. Part A consists of polydimethylsiloxane polymers functionalized with vinyl groups, and part B is rich in silicon hydride groups. When mixed at a stoichiometric ratio (10A:1B), the addition reaction of the vinyl groups to the silicone hydride groups forms covalent bonds with no excess functional groups. However, if the two parts are mixed off-ratio, excess groups in separate layers will be available for covalent bonds between layers, creating a multilayer monolithic device (Fig 7.1).

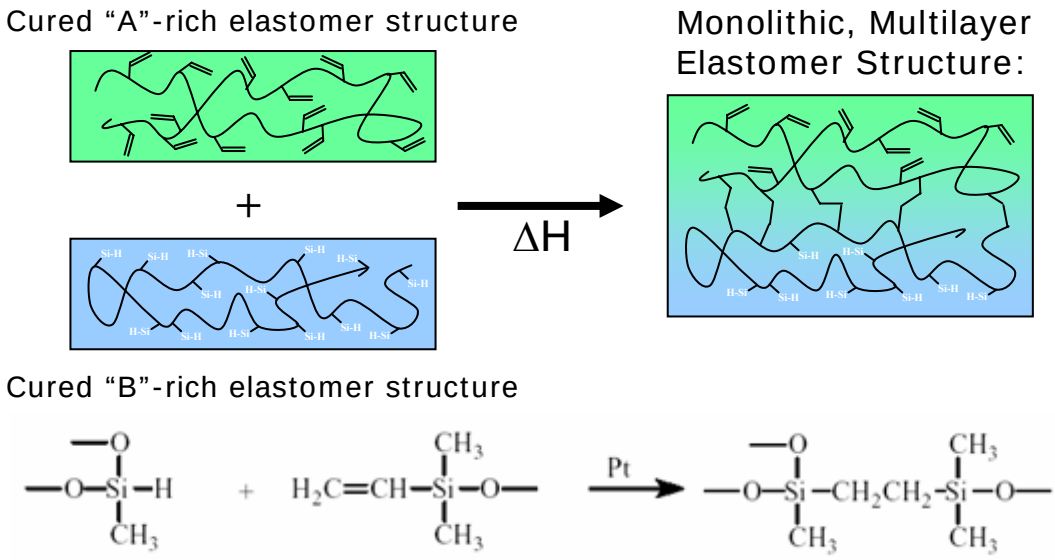


Figure 7.1. MSL bonding process. Silicone hydride groups are covalently cross-linked to vinyl groups by addition reaction. Excess of one functional group enables multi-layer bonding and monolithic device fabrication.

Figure 7.2 depicts the process for creating a microfluidic chip by MSL. Photomasks are drawn with AutoCAD (Autodesk) and printed on transparency film at a resolution of 20,000 dpi (CAD/Art Services). The two-dimensional masks for the control and flow layers are then used to define three-dimensional features by standard optical lithography. The negative molds are subsequently used as casts for the liquid elastomer, polydimethylsiloxane (PDMS). The top layer is cast thick (flow), rich in vinyl groups and platinum catalyst while the bottom layer (control), rich in silicon hydride groups, is cast thin by spincoating. The elastomer is then baked and partially cured on the molds, after mold treatment with trimethylchlorosilane, to facilitate elastomer release. The thick layer is peeled off its mold, and flow channel access holes punched and aligned to the thin layer. Interlayer covalent bonding is then facilitated by baking the aligned layers at

80°C. The resulting two layer monolithic device is peeled from the control mold followed by punching of control ports, to enable control line access. The two layer device is then bonded to a third layer of partially cured 20A:1B PDMS, peeled off, and bonded to an RCA-cleaned glass microscope slide.

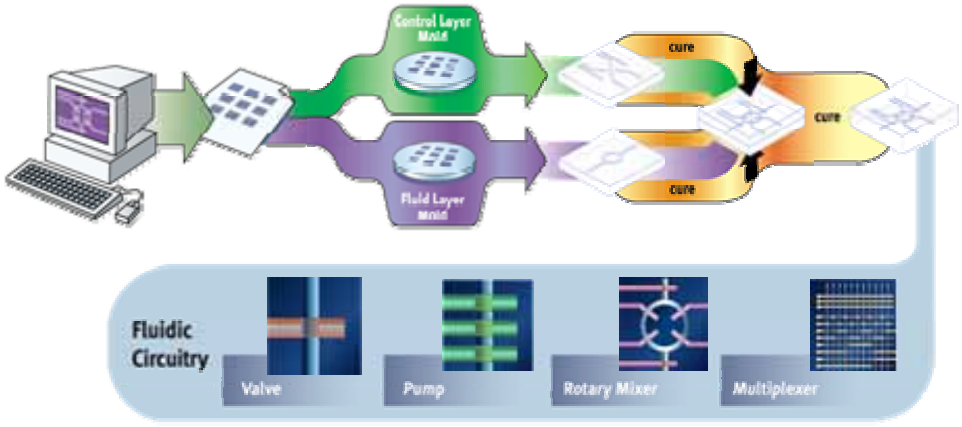


Figure 7.2 MSL process flow (cartoon courtesy of C. Hansen).

7.2 Push-up and Push-down valves

As shown in Fig. 7.2, the valve is the basic building block for a variety of fluidic components. There are two geometries conducive for creating a valve out of PDMS: push-up and push-down (Fig. 7.3). In both cases, pneumatic pressure is applied to the control (actuation) channel, which causes deflection of the PDMS membrane downward or upward to seal the flow channel. Because of the non-uniform PDMS membrane present in the push-down valve architecture (depicted in green in Fig. 7.3), actuation pressure does not scale well with a decrease in flow channel aspect ratio.[68] The push-down valve is limited to flow channel heights of ~13 μm and flow channel aspect ratio of 10:1 (width:height).

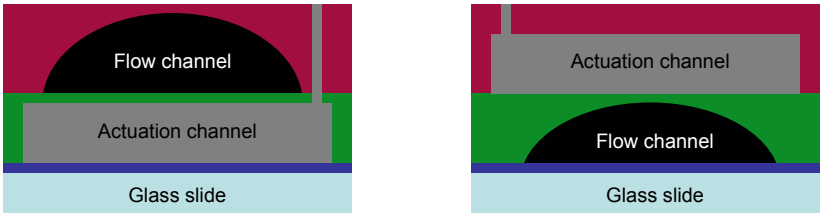


Figure 7.3. Cross sectional cartoon view of push-up (**left**) and push-down (**right**) valve geometries. The blue layer above glass slide is a thin layer of PDMS. Figure is not drawn to scale.

In order to manipulate eukaryotic cells on-chip, which have diameters ranging from 10-35 μm, higher flow structures, and as a result lower aspect ratios for flow channels, are needed. Accordingly, Vincent Studer along with others[68] developed the push-up valve. The uniformly thin membrane allows for the actuation pressure to scale well with aspect ratio (Fig. 7.4). Furthermore, actuation pressure of a push-up valve scales with membrane thickness (Fig. 7.4), and as expected, devices with thinner membranes actuate at lower pressures.

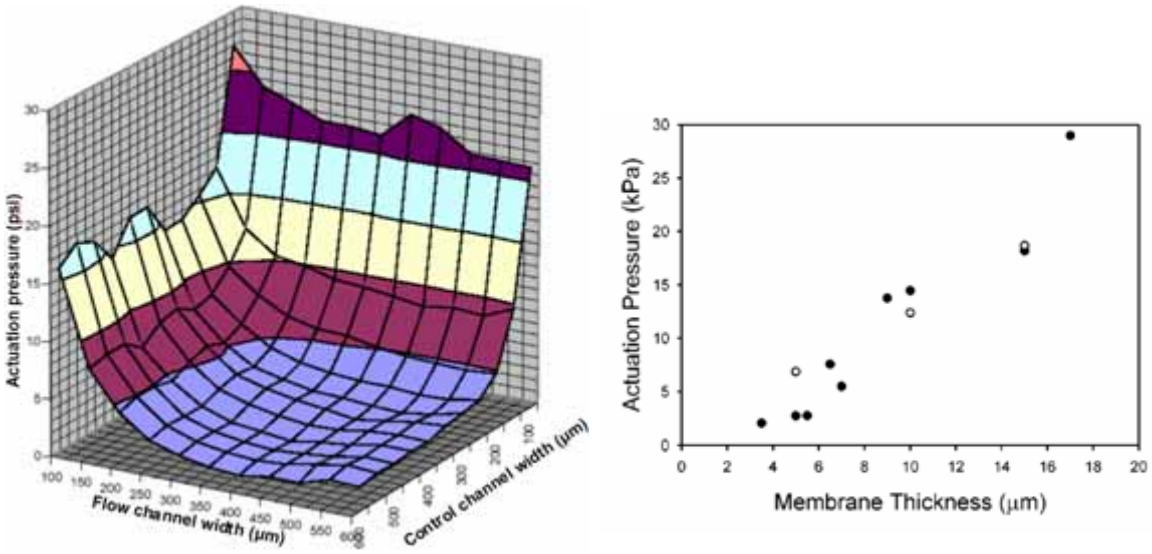


Figure 7.4. Scaling properties of a push-up valve. **Left**: Actuation pressure is symmetrical with channel width. PDMS membrane is ~5μm. **Right**: Actuation pressure

as a function of membrane thickness for a 300 µm X 300 µm valve. The flow channel is 56 µm high. Closed circles are measured values and open circles are simulations.[68]

A comparison of the resistance across a flow channel provided by the two valve geometries demonstrates the push-up valve closes at an order of magnitude lower pressure than its push-down counterpart (Fig. 7.5). For these measurements, identical valve parameters of 300 µm X 200 µm, 46 µm high flow channel, 10 µm membrane were utilized. The measurements were taken by first measuring the interface of two fluid stream positions as a function of input pressure of one of the streams. The interface was then tuned to the middle of the channel and pressure was applied to a valve located close to one of the inputs. The change in the position of the interface was then plotted against the pressure in the valve. Finally, normalized resistance (percentage of input pressure) was extrapolated and plotted as a function of actuation channel pressure.[68] The results are shown in Fig. 7.5.

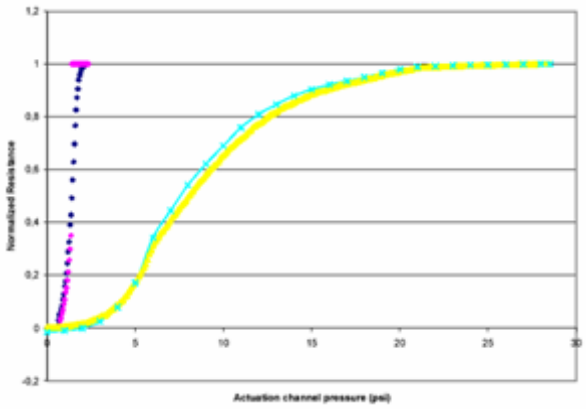


Figure 7.5. Plot of the normalized resistance provided by a push-up (purple, pink) and push-down (blue, yellow) valve as a function of actuation channel pressure. The first color listed is for increasing actuation channel pressures and the second color listed is for decreasing pressures. The push-up valve shows a hysteresis due to membrane sticking to the top of the flow channel.[68]

7.3 Optical lithography considerations

A key consideration in microfluidic chip design is the minimization of dead volume. Consequently, valve dimensions should be as small as possible in areas of the device that are not crucial for process functionality. For example, when building microfluidic affinity columns for the mRNA isolation devices presented in chapter 5, one would ideally want to utilize 100 μm X 100 μm valves (or smaller) for the bead multiplexer valves. If the multiplexer implements 200 μm X 200 μm valves, as would be the case if channels were cast out of AZ-50, cell lysates would have to be pushed pneumatically an additional 1.5 mm for the 20X device and 2 mm for the 50X device. The additional length would increase the possibility of lysate sticking to channels and therefore loss of material before affinity purification. Furthermore, in the case of the 50X device, the larger valves would result in a larger device, making the possibility of fitting it on a 3" silicon wafer unlikely. Because manipulating cells on-chip requires channels with heights on the order of 40 μm, a three layer lithography process (Su8 2010, Spr220-7, AZ-50) had to be implemented to achieve an optimal device.

Two level molds that incorporate SU8 as the shorter feature are straightforward to implement and were utilized for the devices presented in chapters 2 and 3. Because of the considerations presented in the above paragraph, three layer molds had to be fabricated for the denser (20X and 50X) mRNA isolation/cDNA synthesis devices. The incorporation of a third layer on top of a positive resist, in this case Spr220-7, turns out to be a non-trivial problem. 2401 developer (AZ-50 developer) dissolves Spr220-7 features after a basic reflow protocol (20 min. @ 125°C; Fig. 7.6), so the reflow protocol had to

be changed to include a hard bake step at a higher temperature. Others in the lab had experienced similar problems when processing tall SU8 features on a mold with processed positive resist (personal communication with Carl Hansen). Hansen found hard baking temperatures above 180°C for 1 hour resulted in a permanent chemical change to the 5740 photoresist. Investigation of hard baking protocols for the Spr220-7 resist yielded a similar result. A combined reflow/hard bake step of 200°C for 2 to 3 hours resulted in the Spr resist changing color from red to black (Fig. 7.6). The time varies from mold to mold because the hot plates do not have 100% level surface, so amount of contact varies depending on where molds are placed on the respective hot plate. After the chemical change, the Spr film is resistant to solvents such as acetone and 2401 developer (AZ-50 developer).

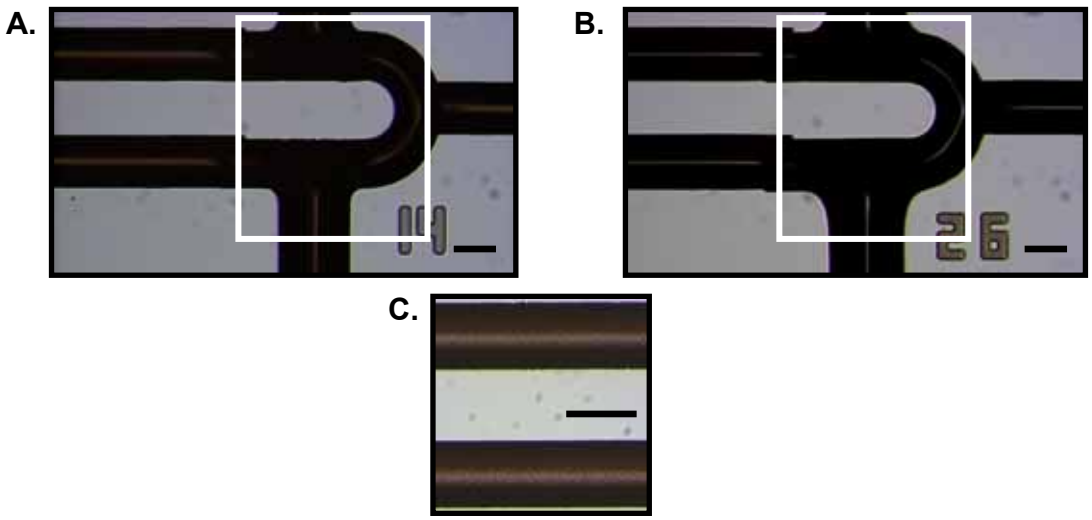


Figure 7.6. Comparison of Spr220-7/AZ-50 films that have been subjected to different hard bake protocols. AZ-50 features are boxed in white. The remaining features are Spr220-7. Spr220-7 features were processed first, followed by a hard bake and subsequent AZ-50 processing and final AZ-50 hard bake. Scale bars are 100 μm. **A.**

Spr220-7 and AZ-50 were not hard baked to the point of a chemical change (features are red). **B.** Spr220-7 and AZ-50 were hard baked until films changed color. **C.** Spr220-7 channels after reflow process in (**A**) and exposure to AZ-50 developer.

7.4 Sieve valves for robust separation assays

The sieve valve, or digital filter, discussed primarily in chapter 3, provides an elegant solution for on-chip separation assays. The sieve valve is aligned to a flow structure with rectangular cross section, allowing for fluid flow through the valve, but not for micron size particles to pass. Because full closing valves need to be aligned to rounded flow structures, multilayer molds are needed to implement microfluidic separation assays. Separation assays can be run successfully when a range of pressures is applied to the sieve valve, because of its digital nature. This is in direct contrast to a prior solution,[24] where the valve used to construct an affinity column was analog and pressure needed to be finely tuned. SU8 films, because of their mechanical properties, inability to be re-flown at high temperatures, and wide range of heights (3 μm-200 μm) are the resist of choice when implementing digital filters. Sieve valve height and width is tuned according to the particle size that needs to be captured. For example, 3 μm beads are trapped sufficiently (~20 psi applied to filter) when utilizing a 200 μm X 200 μm valve present on a rectangular 10 μm high channel. Applications of the digital filter include chromatography columns to cell capture.

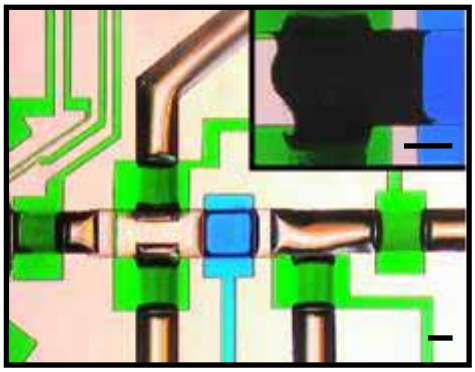


Figure 7.7. Optical micrograph of the chromatography portion of the cells-to-cDNA device presented in chapter 2. Sieve valve is shown in blue. **Inset**: 3 μm beads derivatized with oligo$(dT)_{25}$ sequences are stacked against the sieve valve for mRNA affinity capture. Scale bars are 100 μm.

7.5 Rapid parallel mixing

As discussed in chapter 2, lysing a cell by free interface diffusion[25] poses problems for bead recovery after mRNA affinity capture/first strand cDNA synthesis (Fig 2.1). Although all mixing in microfluidic devices of the length scales presented in this thesis is ultimately accomplished by diffusion, because flow is laminar; accelerated mixing can be achieved by actively decreasing diffusion lengths. Pumping cells and lysis buffer in a ring-shaped channel allows for the repeated folding of fluids onto themselves, thereby allowing for full homogenization of the ring contents and therefore complete lysis. A detailed discussion of the scaling laws that govern microfluidic mixing are presented elsewhere.[71] However, in order to achieve rapid parallel mixing of 20 to 50 reactors on-chip, the basic rotary mixer design needed to be changed from the implementation presented in chapters 2 and 3 to an architecture that could be implemented in more of a linear reaction scheme. Therefore, the ring structures were

stretched out from a circle to an oval (Fig. 7.7). The oval implementation allows for a high density reactor design, with 50 parallel reactors easily fitting onto a 2'' X 2'' chip.

Mixing in the lysis rings is governed by the flow rate in each of the three pump lines. Flow rate in a channel with rectangular cross section can be given as

$$Q = \frac{wh^3}{12\mu L}\Delta P \quad (1)$$

where Q is the volume flow rate, h is the channel height, w is the channel width, L is the channel length, ΔP is the driving pressure, and μ is the fluid viscosity.[22] Consequently, to obtain complete mixing in each ring, the flow rate in each pump line needed to be sufficiently fast to get valve actuation in each reactor. Because flow rate scales favorably with channel dimensions, the author chose to implement pump lines with 125 μm valves along with 60 μm crossovers. Typically, control channels that address 12 μm flow lines would have 100 μm valves with 15 μm crossovers. The linearity of the cells-to-cDNA process allows for pump inputs to be placed at either end of the chip and not in the middle, as in a previous implementation (chapter 3). The bulky rotary mixers in early cells-to-cDNA chips made it necessary to include 15 μm crossovers as well as generally thin control lines (50 μm) for the pump lines, making complete mixing impossible unless inputs were in the middle of the chip. Figure 7.8 shows complete mixing in 50 parallel reactors is feasible.

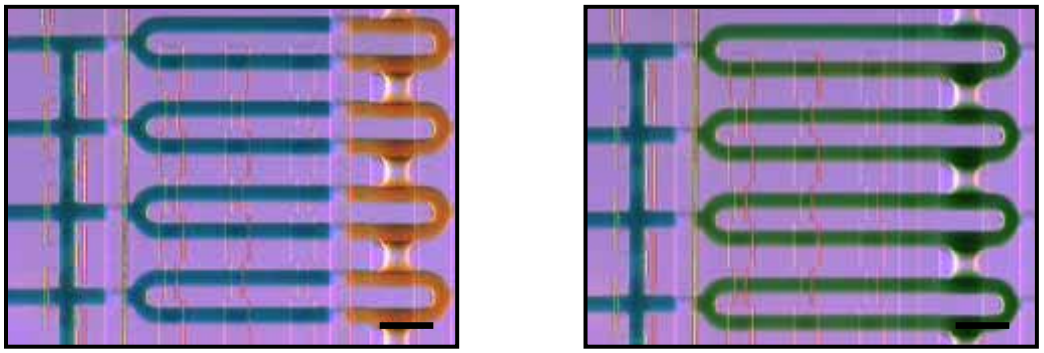


Figure 7.8. Optical micrographs of reactors one-four of the 50X cells-to-cDNA device before (**left**) and after (**right**) mixing. The reactors shown are the farthest from the pump inputs. Scale bars are 250 μm.

7.6 Column construction with one flow input

In order to build affinity columns on a microfluidic device for highly parallel sample analysis, various technical issues need to be addressed. The main consideration has to be given to fluidic resistance, which in a rounded channel is expressed as

$$R = \frac{8\mu L}{\pi r^4} \tag{2}$$

where R is the fluidic resistance, L is the channel length, μ is the fluid viscosity, and r is the channel radius. If every column were built simultaneously in parallel, column lengths would vary tremendously because most of the beads would stack against the sieve valves closest to the bead input (path of least resistance). Serial bead stacking with a fluidic multiplexer overcomes the parallel stacking problem. The multiplexer uses an array of complementary valve pairs organized in a binary-tree architecture to select 1 of N flow lines using $2\text{Log}_2\text{N}$ control lines (Fig. 7.8).[71] The fluidic multiplexer has been implemented previously for large scale integration devices,[20] as well as protein

crystallization and combinatorial mixing chips.[22, 25] Implementing the fluidic multiplexer to stack bead columns allows for unlimited columns to be built on-chip, with the size of the chip being the limiting factor for number of columns implemented. In this scheme, columns can be built either one, two, four, or eight at a time in a uniform fashion.

As previously discussed in section 7.3, dead volume issues also need to be addressed in order for optimal device operation. Three layer flow molds minimize dead volume and were implemented for all devices containing a multiplexer element.

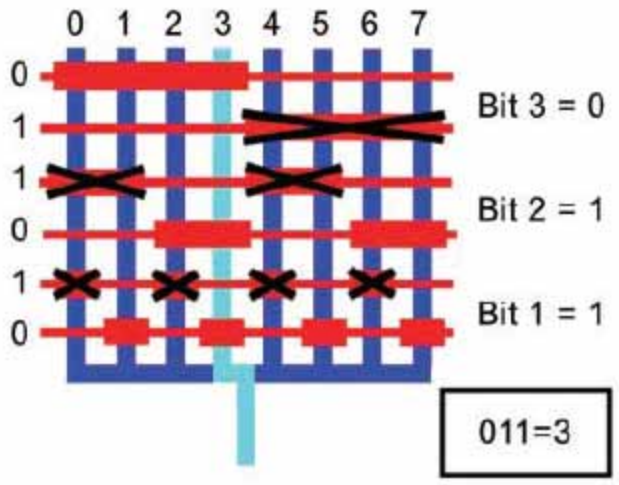


Figure 7.9. Parallel multiplexing structure.[20] Each pair of complementary valves composes a bit. Figure not drawn to scale.

7.7 Cell input with 100% reactor addressability

The final technological innovation by the author was to implement a binary tree multiplexer to allow for 100% reactor addressability with single cells. In earlier devices, reactors were loaded one at a time (chapter 3) or by flowing cell suspension onto a chip and massively partitioning it (chapter 5). The first implementation was not optimal

because of the need for separate valves for each cell input and output. The second implementation was not optimal because Poisson statistics govern that only one third of the reactors will be filled by single cells, if the suspension is correctly diluted. One way to keep the device's complexity the same while addressing each reactor with individual cells is to implement a binary tree multiplexer (Fig. 7.10, 7.11) that shares control line inputs with the bead multiplexer. N consecutive bifurcations originating at a single channel allow for the 2^N inlet channels to be connected through equivalent fluidic paths.[71]

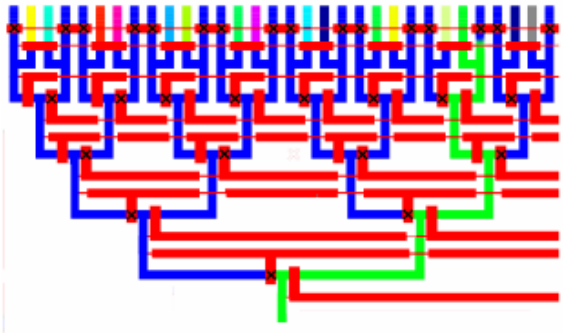


Figure 7.10. Binary tree multiplexing structure. Low dead-volume junctions and interleaved flush lines (blue) allow for zero reagent cross-contamination and addressability of each reactor. Cell suspension or buffer (green) is flushed through unique path of binary tree. Figure adapted from Carl Hansen's thesis.[71]

An AutoCAD drawing of a 16x device, which utilizes multiplexers for both cell and bead loading, is given in Fig 7.11. The same design has also been implemented to fabricate a device with 32 reactors. The additional multiplexer has no affect on device complexity (e.g. inputs from outside world) because every level of the multiplexers is

addressed with the same input (Fig. 7.11, white box). Lysis buffer is first introduced into the device as described in chapter 5, with slight modifications. Instead of keeping the lysis buffer confined to the bead stacking and output modules of the device, it is also introduced into the mixing rings, by opening the control channel marked with an asterisk (Fig. 7.11, bottom right). Once lysis buffer is in the chip, affinity columns are built as described in chapter 5. Cell suspension is then loaded onto the chip through inlet 1, and multiplexer valves are actuated to address each reactor with cells. At the final multiplexer level, flow can be directed to a waste output by opening valve **b** and closing valve **a,** or into the ring by opening valve **a** and closing valve **b** (Fig. 7.11, bottom right). If flow is directed into the ring, valve **c** can be toggled on and off until the desired number of cells are captured in the mixing ring. Once a cell is captured in the mixing ring, the next reactor is addressed, until all reactors are filled. Excess cells are then washed away to the waste outputs in PBS. After the two washes, cells are chemically lysed by executing a peristaltic pump sequence described above. The process of capturing mRNA, synthesizing first strand cDNA and recovery are identical to the 20X chip process described in chapter 5.

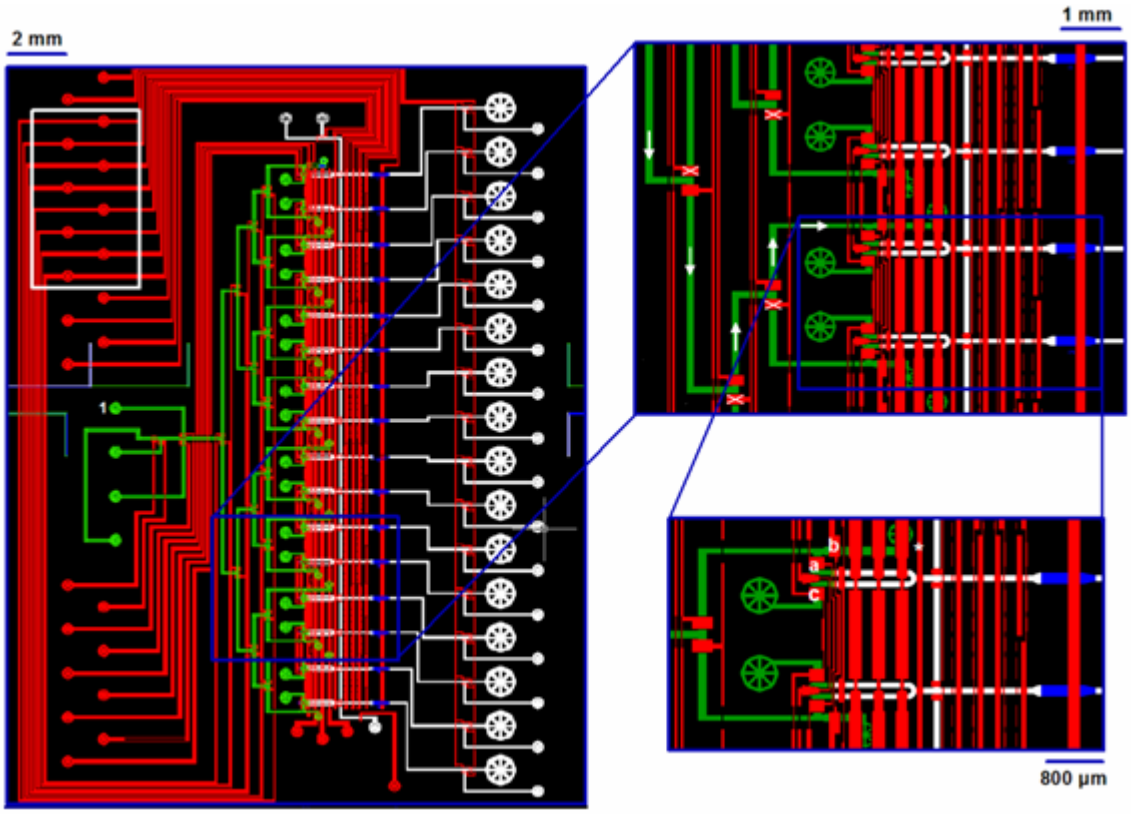


Figure 7.11. AutoCAD drawing of 16X cells-to-cDNA device with 100% reactor addressability. Control channels are in red, 10 µm high flow channels with rectangular cross section utilized for column construction are shown in blue, 12 µm high rounded flow channels are shown in white, and the 40 µm high rounded flow channel is shown in green. The same design is utilized for a device capable of handling 32 reactors. White arrows in top inset show where flow would be directed when certain multiplexer lines are closed (shown with white x's).

Appendix A: Protocols

A.1 Photolithography processes

A.1.1 Su8-2010 10 μm/Spr220-7 15 μm/AZ-50 40 μm flow molds

1. Spin Su8-2025 at 3000 rpm for 45 s. with an acceleration of 10.
2. Soft bake mold for 1 min./3 min. at 65°C /95°C.
3. Expose mold 50 s. real-time on MJB mask aligner (7 mW/cm^2).
4. Bake mold post-exposure for 1 min./3 min. at 65°C /95°C.
5. Develop in Su8 nano developer. Rinse in fresh Su8 nano developer and determine if developed by looking at mold under microscope.
6. Once developed, hard bake mold at 150°C for 2 hr.
7. Expose mold to HMDS vapor for 90 s.
8. Spin Spr220-7 (cold, straight from refrigerator) at 1500 rpm for 1 min. with an acceleration of 15.
9. Soft bake mold for 90 s. at 105°C.
10. Expose mold under a 20,000 dpi positive transparency mask (CAD/Art Services) for 3.2 min. real-time on MJB mask aligner (7 mW/cm^2).
11. Develop mold in MF-319 developer and rinse under a stream of H_2O. Determine if developed by looking at mold under microscope. Spr develops rather quickly, except for areas around the Su8 layer. Therefore, some areas may get overdeveloped when trying to remove residual resist around Su8 portions.
12. Hard bake 2 hr. at 200°C.
13. Expose mold to HMDS vapor for 90 s.
14. Spin AZ-50 (cold, straight from refrigerator) at 1600 rpm for 1 min. with an acceleration of 15.
15. Soft bake mold for 1 min./5 min./1 min. at 65°C/115°C/65°C.
16. Expose mold under a 20,000 dpi positive transparency mask (CAD/Art Services) for 4 min. real-time on MJB mask aligner (7 mW/cm^2).
17. Develop mold in 3:1 H_2O:2401 developer. Rinse mold under a stream of H_2O.
18. Once developed (determine by visualization under microscope), hard bake 3 hr. at 200°C.

Note: Temperatures for hard bakes are ramped up and down from room temperature by either turning on or off the hot plate. This will prevent resist cracking.

A.1.2 Su8-2025 23 μm control molds

1. Spin Su8-2025 @ 3000 rpm for 45 s. with an acceleration of 10.
2. Soft bake mold for 2 min./5 min. at 65°C/95°C.
3. Expose mold under a 20,000 dpi negative transparency mask (CAD/Art Services) 1.2 min. real-time on MJB mask aligner (7 mW/cm^2).
4. Bake mold post-exposure for 2 min./5 min. at 65°C /95°C.

5. Develop in Su8 nano developer. Rinse in fresh Su8 nano developer and determine if developed by looking at mold under microscope.
6. Once developed, bake mold at 95°C for 45 s to evaporate excess solvent.

A.2 Three-layer RTV device fabrication (push-up valves)

1. Prepare 5:1 GE RTV A:RTV B (mix 1 min., de-foam 5 min.).
2. Expose flow mold to TMCS vapor for 2 min.
3. Pour 30 g 5:1 GE RTV A:RTV B on respective flow mold.
4. De-gas flow mold under vacuum.
5. Bake flow mold 45 min. at 80°C.
6. While flow mold is de-gassing, prepare 20:1 GE RTV A:RTV B (mix 1 min., de-foam 5 min.).
7. Expose control mold to TMCS vapor for 2 min.
8. Spin 20:1 RTV mix at 2000 rpm for 60 s. with a 15 s. ramp.
9. Let RTV settle on control mold for ~30 min. before baking 30 min. at 80°C.
10. Bake control mold 30 min. at 80°C.
11. Cut devices out of flow mold and punch holes with 650 μm diameter punch tool (Technical Innovations #CR0350255N20R4).
12. Clean flow device with transparent tape and align to control mold.
13. Bake 2-layer device for 45 min. at 80°C.
14. While 2-layer device is baking, prepare 20:1 GE RTV A:RTV B (mix 1 min., de-foam 5 min.) to spin on blank silicon wafer.
15. Expose blank(s) to TMCS vapor for 2 min.
16. Spin 20:1 RTV mix on blank wafer(s) at 1600 rpm for 60 s. with a 15 s. ramp.
17. Bake blank wafer for 30 min. at 80°C.
18. Cut out 2-layer device(s) from control mold(s), clean with tape and mount on blank wafer(s). Check for debris and collapsed valves. Collapsed valves can be fixed by applying pneumatic pressure with a syringe to the respective control channel(s). This should overcome valves sticking to channels. Once pressure is applied and released, peel device back from blank wafer and re-mount.
19. Bake 3-layer RTV device(s) for 6-18 hr. Less is best (devices can still handle ~30 psi without delaminating).
20. If output holes need to be punched, do so with technical innovation titanium nitride coated punch (#CR0830655N14R4).
21. Cut 3-layer device(s) out, clean with tape and mount on RCA-cleaned glass slide(s). Check for collapse as in (18).
22. Bake finished devices overnight at 80°C.

A.3 RCA Cleaning

1. Clean microscope slides in Micro-90 microcleaning soap (Cole Parmer) in heated ultrasonic tank for 30 min.
2. Rinse slides 3x in DI H_2O to get rid of residual soap.
3. Make RCA cleaning solution, 6:4:1 H_2O:NH_4OH(37%):H_2O_2(30%).
4. Heat solution on hot plate (~60°C, between 3 and 4 on hot plate dial), with samples immersed in it for 80 min. Use magnetic stir bar to stir solution.
5. Rinse samples 5x in DI H_2O and store long term in DI H_2O.
6. Blow samples dry with N2 stream before using.

Notes: RCA cleaing procedure cleans out most organic and metal contamination because of the strong oxidation activity. Hydroxyl groups are left on the finished surface and thus render the surface hydrophilic. Treatment for more than an hour makes the glass surfaces a bit rough. This roughness creates more contact area for a PDMS device, which promotes adhesion to the glass slide.

A.4 Real-time PCR master mix recipes for microfluidic samples.

For conventional samples, volume of water is reduced by 5 µL for every 100 µL master rxn. This allows for 1 µL template per 20 µL rxn.

A.4.1 Sybr green PCR using the Roche Lightcycler 1.5 or BioRAD iCycler.

2X Qiagen sybr green master mix	50 µL
3' Primer (20 µM stock, 300 nM final)	1.5 µL
5' Primer (20 µM stock, 300 nM final)	1.5 µL
Nuclease free H_2O	47 µL
	---
	100 µL

-20 µL to each bead pellet.

A.4.2 Sybr green RT-PCR using the Roche Lightcycler 1.5 or BioRAD iCycler.

2X Qiagen sybr green master mix	50 µL
3' Primer (20 µM stock, 300 nM final)	1.5 µL
5' Primer (20 µM stock, 300 nM final)	1.5 µL
Sensiscript reverse transcriptase	1.0 µL
40 U/µL RNase Inhibitor	1.0 µL
Nuclease free H_2O	45 µL
	---
	100 µL

-20 µL to each bead pellet.

A.4.3 Taqman PCR using the BioRAD iCycler.

2X Qiagen quantitect taqman master mix	50 µL
3' Primer (20 µM stock, 400 nM final)	2.0 µL
5' Primer (20 µM stock, 400 nM final)	2.0 µL
Taqman probe (10 µM stock, 200 nM final)	2.0 µL
Nuclease free H_2O	44 µL
	---
	100 µL

-20 µL to each bead pellet.

A.4.4 Taqman RT-PCR using the BioRAD iCycler.

2X Qiagen quantitect taqman master mix	50 µL
3' Primer (20 µM stock, 400 nM final)	2.0 µL
5' Primer (20 µM stock, 400 nM final)	2.0 µL
Taqman probe (10 µM stock, 200 nM final)	2.0 µL
Sensiscript reverse transcriptase	1.0 µL
40 U/ µL RNase inhibitor	1.0 µL
Nuclease free H_2O	42 µL
	---
	100 µL

-20 µL to each bead pellet.

A.4.5 Taqman multiplex PCR using the BioRAD iCycler.

2X Qiagen quantitect taqman mux master mix	50 µL
3' Primer gene 1 (20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 1 (20 µM stock, 300 nM final)	1.5 µL
3' Primer gene 2 (20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 2 (20 µM stock, 300 nM final)	1.5 µL
3' Primer gene 3 (20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 3 (20 µM stock, 300 nM final)	1.5 µL
Taqman probe 1 (10 µM stock, 200 nM final)	2.0 µL
Taqman probe 2 (10 µM stock, 200 nM final)	2.0 µL
Taqman probe 3 (10 µM stock, 200 nM final)	2.0 µL
Nuclease free H_2O	35 µL
	---
	100 µL

-20 µL to each bead pellet.

A.4.6 Taqman RT-PCR using the BioRAD iCycler.

2X Qiagen quantitect taqman mux master mix	50 µL
3' Primer gene 1(20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 1 (20 µM stock, 300 nM final)	1.5 µL
3' Primer gene 2 (20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 2 (20 µM stock, 300 nM final)	1.5 µL
3' Primer gene 3 (20 µM stock, 300 nM final)	1.5 µL
5' Primer gene 3 (20 µM stock, 300 nM final)	1.5 µL
Taqman probe 1 (10 µM stock, 200 nM final)	2.0 µL
Taqman probe 2 (10 µM stock, 200 nM final)	2.0 µL
Taqman probe 3 (10 µM stock, 200 nM final)	2.0 µL
Sensiscript reverse transcriptase	1.0 µL
40 U/ µL RNase inhibitor	1.0 µL
Nuclease Free H_2O	33 µL
	---
	100 µL

-20 µL to each bead pellet.

A.5 Thermal cycling protocols

A.5.1 Conventional benchtop thermal cycling (50 µL reactions)

A.5.2 Sybr green PCR using the Roche Lightcycler 1.5 or BioRAD iCycler.

Cycle 1:	(1X)		
	Step 1:	95.0ºC	for 15:00
Cycle 2:	(50X)		
	Step 1:	94.0ºC	for 00:15
	Step 2:	55.0ºC	for 00:30
	Step 3:	72.0ºC	for 00:30
	Step 4*:	79.0ºC	for 00:10

*Data collection and real-time analysis enabled.

A.5.3 Sybr green RT-PCR using the Roche Lightcycler 1.5 or BioRAD iCycler.

Cycle 1:	(1X)		
	Step 1:	50.0ºC	for 30:00
Cycle 2:	(1X)		
	Step 1:	95.0ºC	for 15:00
Cycle 3:	(50X)		
	Step 1:	94.0ºC	for 00:15
	Step 2:	55.0ºC	for 00:30
	Step 3:	72.0ºC	for 00:30
	Step 4*:	79.0ºC	for 00:10

*Data collection and real-time analysis enabled.

A.5.4 Taqman PCR using the BioRAD iCycler

Cycle 1:	(1X)		
	Step 1:	95.0ºC	for 15:00
Cycle 2:	(50X)		
	Step 1:	94.0ºC	for 00:15
	Step 2*:	60.0ºC^	for 01:00

*Data collection and real-time analysis enabled.
^ 60.0ºC is used for mammalian genes and 57.0ºC is used for bacterial genes.

A.5.5 Taqman RT-PCR using the BioRAD iCycler

Cycle 1:	(1X)		
	Step 1:	50.0ºC	for 30:00
Cycle 2:	(1X)		
	Step 1:	95.0ºC	for 15:00
Cycle 3:	(50X)		
	Step 1:	94.0ºC	for 00:15
	Step 2*:	60.0ºC^	for 01:00

*Data collection and real-time analysis enabled.
^ 60.0ºC is used for mammalian genes and 57.0ºC is used for bacterial genes.

A.5.6 Taqman multiplex PCR using the BioRAD iCycler

Cycle 1:	(1X)		
	Step 1:	95.0ºC	for 15:00
Cycle 2:	(50X)		
	Step 1:	94.0ºC	for 00:15
	Step 2:	60.0ºC	for 00:45
	Step 3*:	60.0ºC	for 00:45

*Data collection and real-time analysis enabled.

A.6 Labeling cells with DyeCycle green dye (Invitrogen# V35004)

1. Trypsinize cells (0.5 mL) and when cells detach from plate, stop reaction with 4.5 mL DMEM/10% CS.
2. Count cells in culture and spin down $1X10^4$ cells per tube (5000 rpm, 5 min.)
3. Resuspend each pellet in 1 mL DMEM/10% CS.
4. Add 2 µL dye to each mL of cell suspension (15 mL concical tube).
5. Incubate suspension for 30 min. in 37°C/10% CO_2 incubator.

Appendix B: Intracellular Single Cell Studies Using Microfluidics

Note: Appendix B is the author's contribution to a chapter entitled "Progress in nanofluidics for cell biology" in the second edition of the *CRC handbook of nanoscience, engineering, and technology*, to be published in October 2006.

Introduction

It has become increasingly clear that studying cell populations, regardless of what is being assayed (e.g. mRNA, protein, or small molecule levels) is only sufficient for obtaining average values over the particular population. Vast cellular heterogeneity is a common theme in all biological organisms, and the ability to interrogate cells in an individual manner could elicit the presence of subpopulations and provide insight into specific processes, such as stem cell differentiation events, which are masked at the population level. Although questions pertaining to what genes and proteins are expressed in single cells, to how individual cells communicate and respond to different stimuli are all active research areas, tools to answer these questions are generally lacking. The main reason for the lack of technologies to study the single cell can be contributed to the poor yields encountered when subjecting these precious samples to multi-step processes.

The typical channel dimensions found in microfluidic devices (10 μm-100 μm in x, y, and z) and the ability to manipulate nanoliters of reagents on-chip have made the devices encouraging platforms for the analysis of single cells.[18] Furthermore, the economy-of-scale benefits along with the ability to parallelize and automate processes are significant advantages not found with conventional biological assays. The focus of this section will be on micro/nanofluidic assays developed to interrogate the intracellular

content of single cells, although cell population assays that are amenable to individual cells will also be considered. We will begin by discussing assays in which single cells can be biochemically modified, followed by assays to determine expression of genes, proteins, and other signaling molecules. It is not the section's intention to provide a comprehensive list of intracellular assays applied in microfluidic devices, but rather to highlight techniques that have the most promise for single cell analysis.

Single cell electroporation

Electroporation, or the application of an electrical field to induce changes in membrane permeability, has been harnessed in batch assays to introduce polar and charged agents such as nucleic acids, proteins, dyes, and drugs into cells. Single cell electroporation allows for intracellular biochemical manipulation as well as investigation of cell-to-cell variation in response to various substances. Recently, electroporation assays have been developed to perform the technique on single cells in microchannels.[72-75] These efforts utilized cell loading with a syringe in order to control flow rate and to apply back pressure to the microchannel for effective trapping. The earlier devices[72, 75] were fabricated with standard silicon microfabrication techniques and trapped an individual cell in a micropore present between two electrodes. The more recent work took advantage of soft lithography[70] to trap cells in tapered microchannels. Cross sections of the devices that employ the two electroporation schemes are given in Fig. B.1. Each group effectively trapped a single cell between two electrodes and was able to show electroporation events by either measuring molecular uptake or release, by fluorescence. Current jumps due to membrane poration could also be studied by treating the cell as a

resistor in the existing electrical circuit. Although cells were successfully trapped by either strategy, a significant difference between the two strategies was the effective voltage needed to electroporate cells. In the PDMS chip[73], ≤ 1 V was needed to electroporate a single HeLa cell, while in the silicon devices,[72, 75] 15-20 V were needed to reversibly electroporate various individual mammalian cells. In each case, however, molecular transfer rate was close to 100%, a direct contrast to conventional schemes.[74] A main avenue of research in this area will be integration of downstream on-chip steps after electroporation/molecular transfer events, as well as development of more efficient trapping strategies to increase sample throughput.

Cellular nucleic acid isolation

The isolation of gDNA or mRNA from bacterial and/or mammalian single cells is a crucial step for many biological and medical applications. Construction of cDNA or gDNA libraries at the single cell level could allow for gene discovery in rare cells and the elucidation of molecules important for various processes. The following section will focus on the isolation of nucleic acids on-chip from single and/or a small number of cells. It is known that pipetting DNA or RNA results in its shearing,[76] and therefore focus will be given to reports that take advantage of the low Reynolds number flow (shear-free flow) inside microfluidic channels to lyse cells.

Early work on cell lysis/gDNA isolation[77] focused on lysing single GFP expressing *Escherichia coli* (*E. coli*) bacterial cells (treated off-chip with lysozyme to break down the peptidoglycan layer) by diffusional mixing with dI water at the junction of a T channel, which was molded in PDMS and bonded to a glass coverslip by oxygen

plasma treatment. Efficient lysis was demonstrated by the release of GFP into the microchannel. The lysate was then transported via dielectrophoresis (DE) to a trapping region of the device, where the channel constricted and contained a series of 10 μm gaps both to maximize the electric field and trap the *E. coli* chromosomes. DNA was visualized with a nucleic acid stain, with cells being stained before loading on-chip. It was shown that manipulating the applied DC and AC currents allows for chromosome shuttling to downstream DE traps.

Anderson *et al*.[78] developed a polycarbonate machined device capable of automated microarray sample preparation and subsequent hybridization. A solid phase of cellulose is packed into a chamber, and homogenous lysate generated off-chip is drawn by vacuum through the extraction chamber. The nucleic acid is then eluted for downstream PCR. The nucleic acid isolation procedure yielded a sensitivity of 300 copies.

The first reported integrated mammalian cell capture, lysis, and nucleic acid isolation chip-based assays[24] utilized the technique of multilayer soft lithography (MSL)[19, 79] for device fabrication. In this report, both *E. coli* and murine NIH/3T3 cells were isolated in nanofluidic channels by active mechanical valves. Cells were lysed on-chip by diffusional mixing, followed by nucleic acid isolation by affinity capture. Columns of microbeads were built against a partially closed microvalve, with valve pressure and lysate flow rate having to be tuned precisely to carry out successful assays. Hong *et al*. demonstrated single mammalian cell lysis followed by mRNA capture/purification as well as bacterial cell lysis/gDNA isolation and purification from as little as 30 bacterial cells. Recent advances in the Quake group have made column

construction more robust and digital in nature, as opposed to the analog predecessor (Fig. B.2). Marcus *et al*. demonstrated single mammalian cell resolution with the new method.[23] In addition, first strand cDNA synthesis was integrated with the aforementioned processes by utilizing the oligo$(dT)_{25}$ derivatized beads, which comprised the affinity column, as both primers and a solid-phase support. The additional step did not have adverse effects on sensitivity. By utilizing off-chip quantitative RT-PCR (RT-qPCR) on the bead:mRNA, or the bead:cDNA, complexes the authors demonstrated the capability to detect mRNA copy numbers spanning 6 orders of magnitude (10-10^6 copies).

Nucleic acid amplification/detection

PCR has proven to be the most sensitive method for amplifying small amounts of nucleic acids for the purpose of global and specific gene identification, with single cell studies now commonplace in the literature.[3, 4, 10, 13, 14, 40-42, 44] Nanofluidic channels offer a robust platform for PCR[28, 29, 78, 80-84] for various reasons. Microfluidic devices provide efficient thermal transfer due to high surface to volume ratios and offer diagnostic and forensic applications based on the ability for faster thermal cycling and on-site detection. The capacity to partition reagents into thousands of chambers[20, 29] enables concentrated reactions, digital detection,[85] and therefore absolute copy number determination in single cells or otherwise.[84] Despite the overwhelming promise for applying microfluidic PCR to single cells, little has been reported[23, 78, 80, 86] on devices that integrate cell lysis and RNA or DNA purification with an amplification reaction. Integration of cell lysis is critical because of the aforementioned shearing issue and because it establishes the

temporal resolution of the particular downstream measurement(s).[87] Furthermore, to realize true economy of scale and to maintain concentrated samples throughout the full purification and amplification processes, microfluidic process integration is essential. On-chip PCR has been reviewed extensively elsewhere,[76, 88] and therefore only one application of microfluidic PCR will be discussed here. Other gene expression microanalytical techniques will also be reviewed.

Unger *et al*.[84] harnessed MSL technology to design a PCR device capable of partitioning PCR reactions into tens of thousands of picoliter compartments (Fig B.3). The proof-of-principle work isolated varying concentrations of β-actin cDNA molecules down to 0.1 copies per chamber and employed fluorescent hydrolysis probes specific to β-actin for readout of the method. The observed number of positive chambers matched almost identically with what is calculated by Poisson statistics (Fig B.3). More recently, this technology has been applied by Quake's group to study multiplexed gene expression in single mammalian cells, as well as to determine phenotypes of unculturable bacteria found in the termite gut.[24] In the case of mammalian cells, single cells are seeded in 96 well plates or microcentrifuge tubes followed by chemical lysis off-chip in a modified RT-PCR buffer. The lysate is then loaded onto the device and partitioned via actuation of the valve array. The chip is then thermal cycled and fluorescence detected with a flatbed scanner. Besides fluorescence detection with oligonucleotide probes or DNA binding dyes, the presence of on-chip PCR products has been elucidated by electrochemical[80, 89-91] and electrophoretic methods.

Nucleic acid sequence based amplification (NASBA) has been applied in nanoliter volumes inside microfluidic devices, although not to single cells.[92-94] The

technique is gaining popularity in nanosystems because the process is isothermal (41°C) and therefore simplifies chip design, and obviates the problem of evaporation at hot start and denaturation temperatures (~95°C). Jayaraman's group[60, 61] developed PDMS-based microfluidic devices to study TNF-α activated NF-κB gene expression dynamics with EGFP reporters, inside living mammalian cells. The authors presented two similar devices, both of which allowed for cells to be seeded in chambers along a microfluidic network, and subsequently grown to the desired confluency. Medium exchange every 12 hours was accomplished with syringe pumps to replenish metabolites and to remove toxic wastes and dead cells. The devices differed in fluidic architecture so that eight[60] or four[61] different concentrations of activator (applied via syringe pumps) could be applied to cells in parallel, enabling dose response curves to be generated for populations of cells as well as single cells (Fig B.4). Future microfluidic work based on these studies will surely incorporate the monitoring of multiplexed gene expression dynamics using multiple reporters, as well as dynamics of heterogeneous cells, grown together in culture, by the utilization of distinct reporters in each cell type. Furthermore, intrinsic and extrinsic noise[53] in single cell gene expression will be able to be resolved by these methodologies.

Microfluidic electrophoresis

An extensive body of work has been done to study intracellular contents of single cells by electrophoresis inside nanofluidic channels.[87, 95-101] The majority of applications has focused on the detection of small molecules and metabolites,[95, 97-99] as well as DNA content.[96] Electrophoretic detection of specific protein expression has been accomplished in microfluidic devices in immunoassay-based formats[102] as well as by probing for

aromatic amino acids; but has yet to be integrated on-chip with single cell samples.[103] However, a device integrating cell lysis with detection of a GFP chimera has been disclosed.[104]

Fang and coworkers[97, 99] published two reports in the past year disclosing a method that utilized microfluidic electrophoresis to detect reactive oxygen species (ROS) and/or glutathione (GSH) inside single human erythrocytes. These molecules are known to mediate numerous pathological processes, including cancer and brain trauma. The researchers utilized photolithographic and wet chemical etching to define a 12 μm deep and 48 μm wide microfluidic T-channel in soda-lime glass (Fig B.5). Single cells were transported into the lyis/separation portion of the channel by electroosmotic flow (EOF) by applying electrical potentials to the reagent reservoirs. Lysis of the single cell was accomplished by either subjecting the cell to a stronger electric field,[99] or by subjecting it to the same electric field after a buffer exchange.[97] For detection, the authors took advantage of native nonfluorescent dyes rhodamine 123 (DHR 123) and 2,3-naphthalene-dicarboxaldehyde (NDA) that fluoresce when oxidized by either ROS or GSH, respectively. Labeling of the cells was accomplished either before sample loading or directly after cell lysis, by diffusive mixing. The authors utilized laser induced fluorescence (LIF) for detection, and were able to resolve ROS and GSH at sensitivities two orders of magnitude lower than conventional detection limits.

Munce *et al*.[98] employed a scheme to electrophoretically separate the intracellular contents of four individual cells in parallel, followed by multiplexed detection of the cell permeable molecule calcein AM (which fluoresces after cleavage by esterases inside the cell) and the DNA minor groove binder, Hoechst 3342, by LIF. Acute myeloid leukemia

(AML) cells were transported via optical tweezing to a tapered portion of the microchannel where cells underwent lysis by electromechanically shearing. The technique offers selective lysis of cells that contact the channel opening (Fig B.6), which allows for the simultaneous injection of the cell's cytoplasmic contents into predetermined channels.[98] The channels were defined in PMMA by laser ablation, and electrodes were formed by inserting stainless steel wires into the cathode and anode reservoirs in the microfluidic device. The majority of future studies in this field will most likely harness soft materials, such as PDMS,[19, 70] in order to achieve greater parallelization and process integration and separation via mechanical valves.

Proteomics applications

In general, reported single cell protein expression studies, whether specific or on the scale of the proteome, are far fewer in number than single cell gene expression studies. Numerous factors contribute to this fact, including the inability to amplify proteins from a single cell and the variability of gene products expressed due to truncations, splice variants, and post-translational modifications.[105] Tools that operate on the length scale of proteins, such as microfluidic devices harnessing nanofluidic channels, could have far-reaching implications for single cell protein expression experiments.

Benchtop-based immunoassays, such as ELISA, are the workhorse for determining intracellular protein expression in bulk samples. ELISA has been transferred to microfluidic formats,[103] but has not been utilized to detect analytes from cell lysates. The key consideration for these assays is how to immobilize the first layer of the immunostack inside a nanochannel. Two methods that immediately come to mind are

surface derivatization and solid phase extraction with antibody-coated microbeads. Because of the high surface to volume ratio found in nanofluidic channels, surface derivatization will employ a greater capacity for capture than beads or conventional microtiter plate-based assays. However, bead assays with the scheme employed by Marcus *et al*. (Fig B.2) may provide a robust way to interact cellular lysate with antibodies for proteomic applications. Kartalov *et al*. derivatized epoxide floors of PDMS nanochannels with antibodies to cancer markers.[106] Because the device utilized active valving, specific channels could be individually addressed with antibodies of choice, enabling multiplexed detection from one sample. Once antibodies are bound to the surface, the antigen-containing sample can be pressure-driven over each section of the device, enabling all antibody parameter space to be sampled. Protein digests have also been detected in microfluidic systems by two-dimensional (2-D) electrophoresis, although not from cellular samples. Ramsey's group[107, 108] has separated tryptic digests of model proteins by micellar electrokinetic chromatography (MEKC) followed by capillary electrophoresis. The second report demonstrated higher separation resolution because of the optimized channel geometry. The more recent work employed asymmetric turns and decreased the second dimension's injection plug length, therefore achieving faster sampling and a single order of magnitude improvement in peptide peak resolution.

Other proteomics tools being transferred to the nanoscale are devices that prepare samples for mass spectrometry (MS) and x-ray diffraction analysis. Mass spectrometry is a powerful tool in that it can elucidate global networks of proteins; and x-ray diffraction of macromolecular crystals is the most sensitive method for determining molecular

structure. Microfluidic MS applications have been reviewed extensively elsewhere[103] and therefore will not be discussed further. Because structure implies function, elucidating protein structure is an extremely active pursuit in both biochemistry and cell biology laboratories. Hansen *et al*. took advantage of the unique fluid physics present on the microscale[25, 109] to screen protein crystallization conditions by barrier interface metering (BIM). The PDMS device fabricated by MSL is capable of screening 144 protein crystallization conditions with only 10 nL of protein for each condition, which is 2 orders of magnitude less than conventional methods such as hanging drop.

Quantifying intracellular protein and small molecule levels

Detecting proteins and small molecules in single living cells has been accomplished via fluorescence assays inside microfluidic devices. In the case of small molecules, most reported work[18, 110-112] has been carried out probing intracellular Ca^{2+} concentration, $[Ca^{2+}]_i$, in mammalian cells, mainly due to the established commercial dyes available and the importance of Ca^{2+} as a universal second messenger. Protein expression has been examined by quantifying GFP variants transfected inside cells.[20, 61, 113, 114] Thorsen *et al*. utilized microfluidic LSI[20] to probe for the expression of a particular enzyme in single *E. coli*, in a massively parallel fashion. The researchers isolated cytochrome C peroxidase (CCP) expressing cells in subnanoliter compartments by actuating a valve array. The cells were then allowed to interact with substrate, followed by fluorescent product detection. In the future, high throughput screening should become the norm with this technology.

In the $[Ca^{2+}]_i$ reports referred to above, cells were either stained with respective calcium indicators off-chip,[18, 110, 111] or staining was implemented via diffusive mixing on-chip.[112] Mixing fluorescent dye with cells on-chip was inefficient, taking nearly an hour to maximize signal to noise, with respect to the cell and its surrounding channel. Although dye loading on-chip minimizes cell damage, more robust and efficient methods are needed to shorten the diffusion length scale, possibly by implementing valves[19] to enclose the reaction. Manipulation of cells on-chip also varied from researcher to researcher. Two reports[18, 112] disclosed microfabricated structures inside fluidic channels to retain single cells, while another group took advantage of a hydrodynamic pressure difference between two parallel flow channels to dock cells in a hard-wired dam between the channels.[111] The final group[110] flowed cells through microchannels and numerous detection zones, in order to gain real-time monitoring of $[Ca^{2+}]_i$. The $[Ca^{2+}]_i$ assays implemented were similar in nature, employing diffusive mixing of calcium agonists with cells followed by detection of a fluorescent dye product. The next generation of calcium assays will certainly integrate the culture of mammalian cells along with real-time calcium monitoring. The ability to partition cell subpopulations in nanoliter compartments will allow for high throughput screening applications as well as the study of intrapopulation effects.

Conclusions

Microfluidic assays for cell interrogation and manipulation show great promise for robust single cell analysis. Harnessing the length scales present within these devices will provide highly concentrated assays and the ability to analyze single cells in a

massively parallel fashion. The fluid physics present in the nanoliter regime, as well as the ability to manipulate single cells in channels on the order of their diameter, may provide answers to single cell biological questions not obtainable through conventional analysis.

Acknoweledgements

The authors thank W. French Anderson for critical reading of the chapter. JSM is a graduate student in Stephen R. Quake's group at the California Institute of Technology and is supported by National Institutes of Health (NIH) grant 1R01 HG00264401A1.

Figure B.1

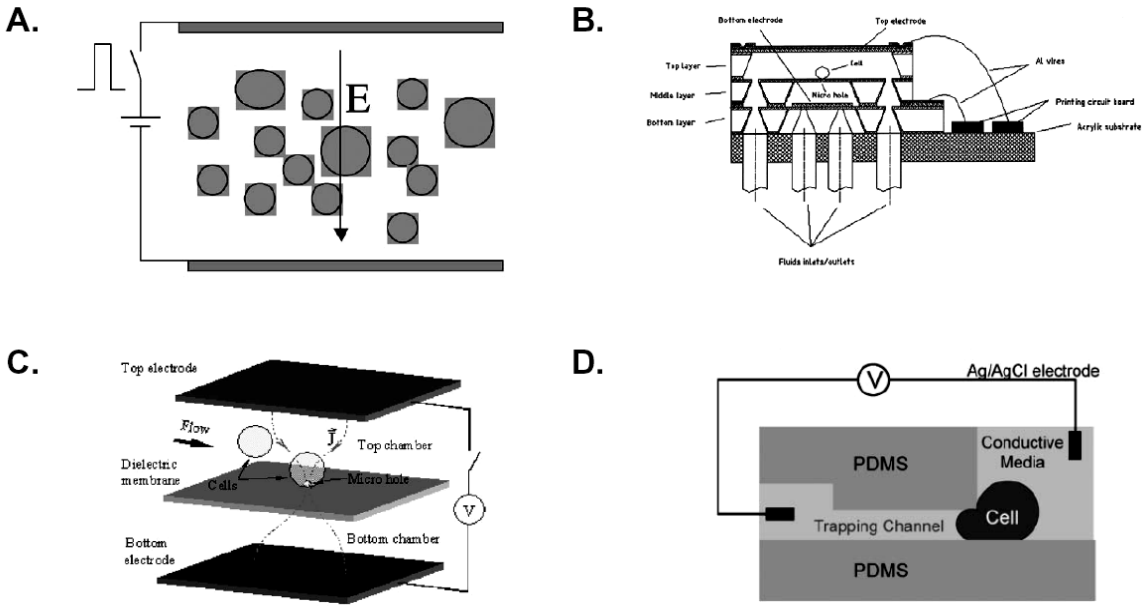


Figure B.1. Electroporation schemes. **A.** Bulk electroporation. Only a portion of the cells are electroporated because the electric field is not uniform with respect to the cells exposed to it. **B., C.** Cross sections of silicon single cell electroporation devices. Cells are transported in a microchannel and captured in a pore by applying back pressure to the channel. Reprinted from *Sens. Actuators A*, Vol. 89, Huang, Y., and Rubinsky, B., Microfabricated electroporation chip for single cell membrane permeabilization. pp. 242-249, 2001; and *Sens. Actuators A*, Vol. 104, Huang, Y., and Rubinsky, B., Flow-through micro-electroporation chip for high efficiency single-cell genetic manipulation, pp. 205-212, 2003, with permission from Elsevier. **D.** Cross section of a PDMS based single cell electroporation device. Cells are captured between electrodes by hard wiring the device to constrict channels to a height smaller than a cell's diameter. Khine, M., et al. (2005). A single cell electroporation chip. *Lab Chip 5*, pp. 38-43, reproduced by permission of The Royal Society of Chemistry.

Figure B.2

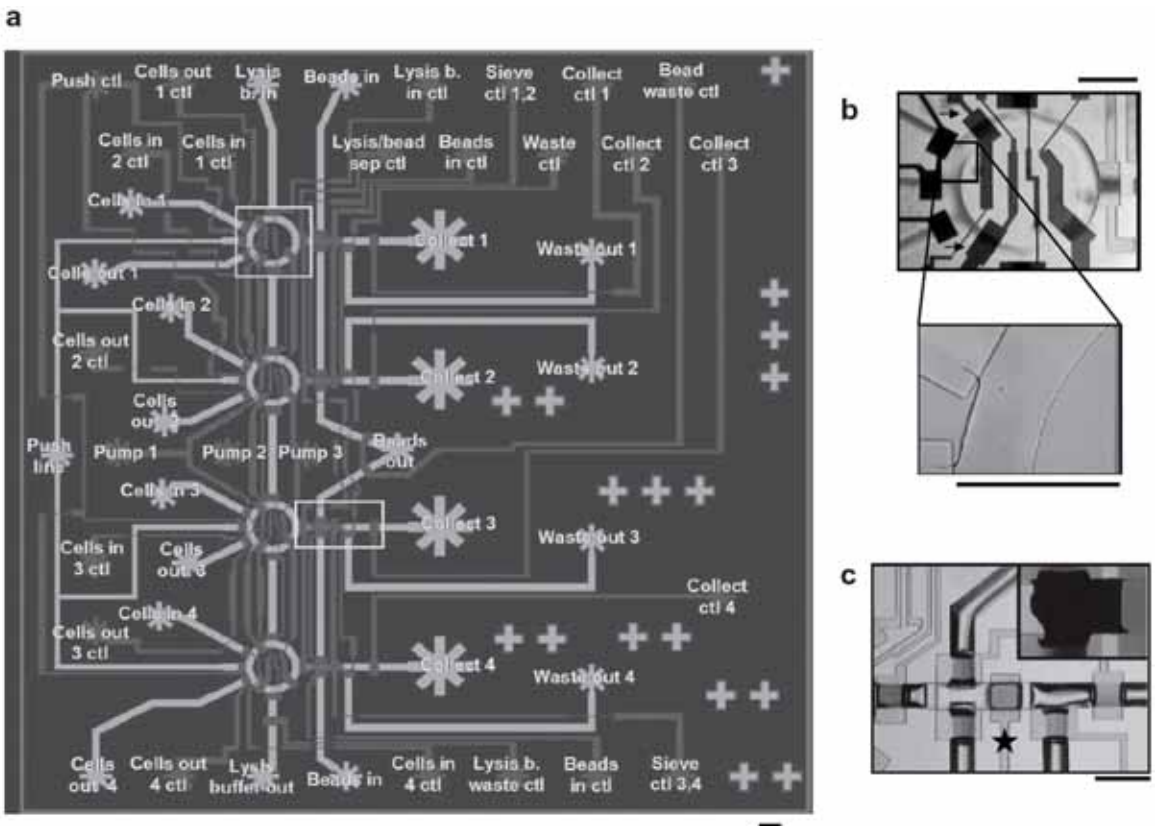


Figure B.2. Fourplex mRNA isolation/first strand synthesis device.[23] **a.** AutoCAD drawing of a device with inputs and outputs labeled according to function. Rounded flow channels are depicted in light gray and control channels are shown in dark gray. Unrounded (rectangular profile) flow channels for affinity column construction are shown in dark gray. Portions of drawing in white boxes are shown in b and c, respectively. **b.** Optical micrographs of the lysis ring and an NIH/3T3 cell captured in the ring. A cell is captured by opening the respective cell in and out valves and keeping pump 1 and pump 2 closed (marked with arrows in figure). Cells are then loaded under pneumatic pressure with a pipet tip. **c.** Optical micrographs of the affinity column construction area and a stacked column against a sieve valve (marked with star in figure).

Scale bars are 400 μm. Marcus, J. S., Anderson, W. F., and Quake, S. R. (2006). Microfluidic single cell mRNA isolation and analysis. *In Press (Anal. Chem.)*, reproduced with permission from the American Chemical Society.

Figure B.3

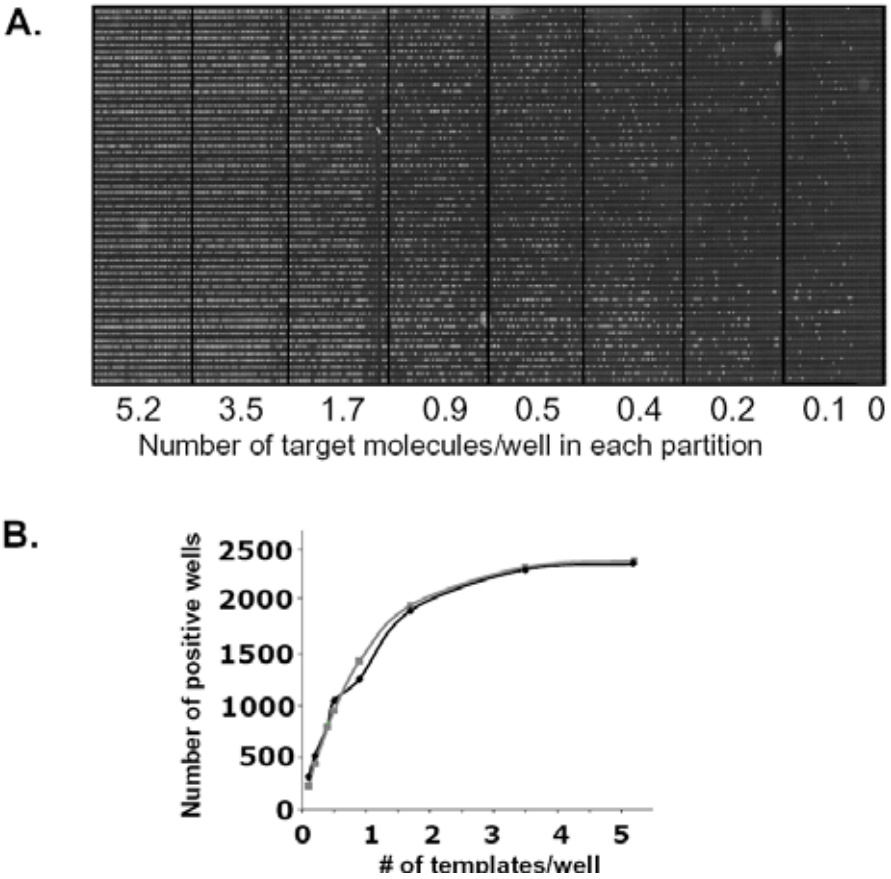


Figure B.3. Digital PCR results from Unger *et al*. **A.** Scan of a chip after 30 cycles (check on this) of PCR. β-actin cDNA molecules were serially diluted and detected with Taqman™ chemistry and a flatbed scanner. **B.** The positive reactors in **(A)** are nearly identical to what is calculated by Poisson statistics. The black line is number of observed positive wells while the gray line is the number calculated by Poisson statistics. Adapted from http://www.fluidigm.com/didIFC.htm

Figure B.4

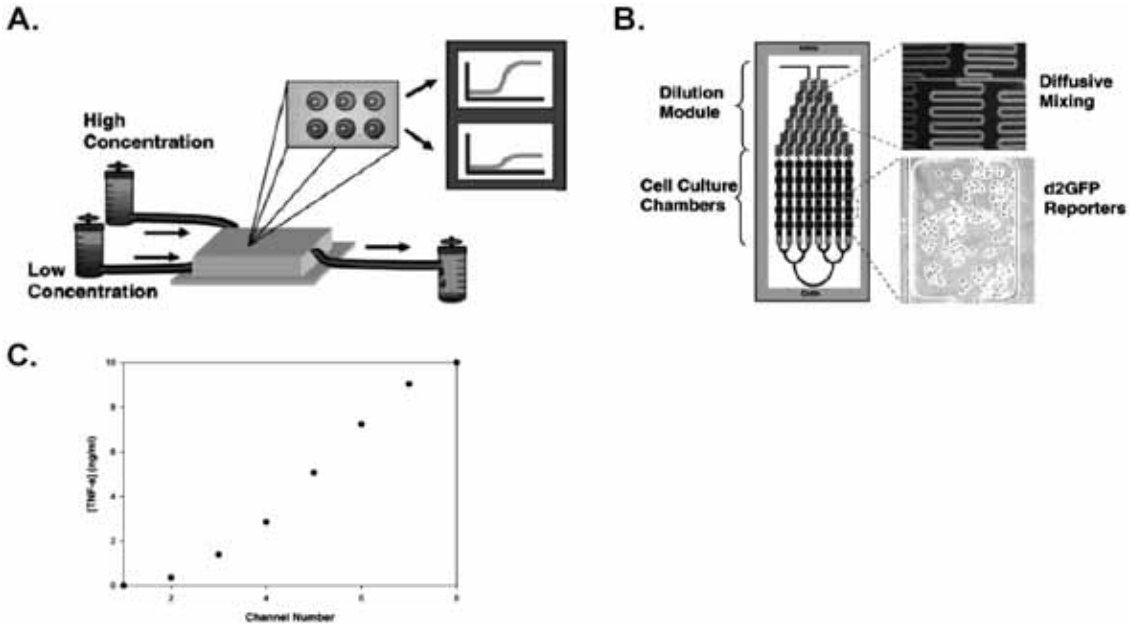


Figure B.4. Living Cell Array **A.** Chip loading scheme. Cells are seeded into the device with a syringe pump through one of the inlets. Chemicals and medium are also loaded with syringe pumps. **B.** Medium and chemicals are diluted via diffusive mixing in the dilution module and are delivered to the cells. TNF-α (spiked with fluorescein) gradient is formed via diffusive mixing. Cells are visualized with d2GFP reporters. **C.** TNF-α dose response curve for chambers in the LCA. Thompson, D. M., et al. (2004). Dynamic gene expression profiling using a microfabricated living cell array. *Anal. Chem.* 76, pp. 4098-4103, reproduced with permission from the American Chemical Society.

Figure B.5

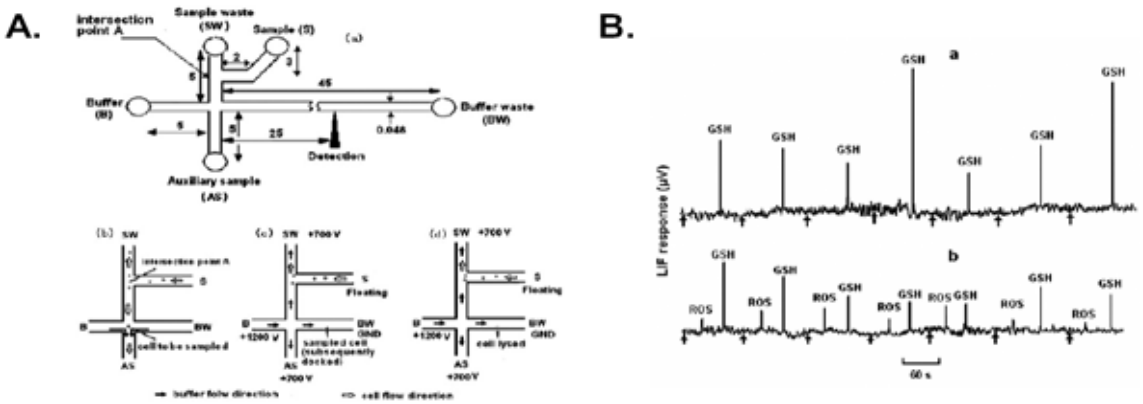


Figure B.5. Determination of erythrocyte contents by capillary electrophoresis. **A. a.** Schematic of the T-microchannel etched in soda-lime glass. Channels are 12 μm high and 48 μm wide. **b.** Single cell is transported from S through intersection point A to AS and SW by hydrostatic pressure. **c.** Set of electrical potentials are applied to inject single cell from sample channel into separation channel. Voltages are turned off once cell is docked in channel. **d.** Docked cell is lysed by applying the same voltage used for cell transport under different buffer conditions. **B.** Recorded electropherograms of seven cells injected consecutively. The separation distance was 25 mm and took place with a voltage of +240 V/cm. Arrows indicate the starting point for separations. **Top –** Fresh cells. **Bottom –** Stimulated with H_2O_2 for 10 min. Ling, Y. Y., Yin, X. F., and Fang, Z. L. (2005). Simultaneous determination of glutathione and reactive oxygen species in individual cells by microchip electrophoresis. *Electrophoresis 26*, pp. 4759-4766, reproduced with permission from John Wiley and Sons.

Figure B.6

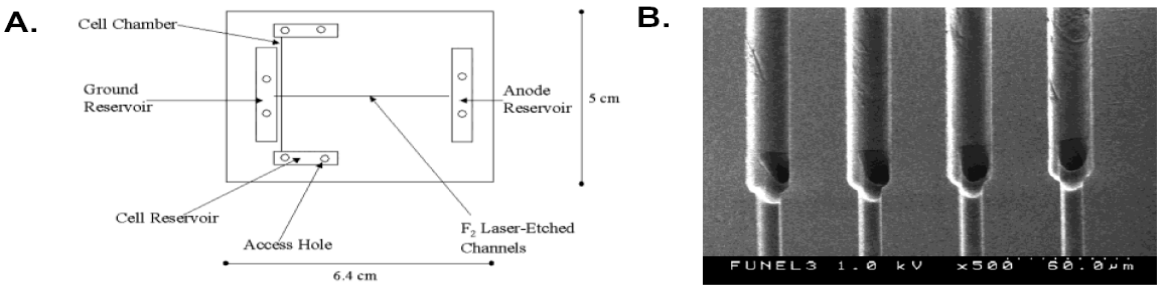


Figure B.6. Parallel intracellular study by microfluidic lysis/electrophoresis. **A.** Layout of device utilized for parallel cellular electromechanical lysis/electrophoresis. One laser etched channel is shown for clarity. **B.** Scanning electron micrograph of the tapered intersection of the injection and separation modules of the channels. Munce, N. R., et al. (2004). Microfabricated system for parallel single-cell capillary electrophoresis. *Anal. Chem. 76*, pp. 4983-4989, reproduced with permission from the American Chemical Society.

Appendix B - Figure References

Hong, J. W., Studer, V., Hang, G., Anderson, W. F., and Quake, S. R. (2004). A nanoliter-scale nucleic acid processor with parallel architecture. Nat Biotechnol *22*, 435-439.

Huang, Y., and Rubinsky, B. (2001). Microfabricated electroporation chip for single cell membrane permeabilization. Sens Actuators, A *89*, 242-249.

Huang, Y., and Rubinsky, B. (2003). Flow-through micro-electroporation chip for high efficiency single-cell genetic manipulation. Sens Actuators A *104*, 205-212.

Khine, M., Lau, A., Ionescu-Zanetti, C., Seo, J., and Lee, L. P. (2005). A single cell electroporation chip. Lab Chip *5*, 38-43.

Ling, Y. Y., Yin, X. F., and Fang, Z. L. (2005). Simultaneous determination of glutathione and reactive oxygen species in individual cells by microchip electrophoresis. Electrophoresis.

Marcus, J. S., Anderson, W. F., and Quake, S. R. (2006). Microfluidic single cell mRNA isolation and analysis. *In Press.*

Munce, N. R., Li, J., Herman, P. R., and Lilge, L. (2004). Microfabricated system for parallel single-cell capillary electrophoresis. Anal Chem *76*, 4983-4989.

Thompson, D. M., King, K. R., Wieder, K. J., Toner, M., Yarmush, M. L., and Jayaraman, A. (2004). Dynamic gene expression profiling using a microfabricated living cell array. Anal Chem *76*, 4098-4103.

Unger, M., Manger, I., Lucero, M., Yong, Y., Miyashita-Lin, E., Wienecke, A., and Facer, G. (2004). European Patent. *Vol. EP1463796.*

Wieder, K. J., King, K. R., Thompson, D. M., Zia, C., Yarmush, M. L., and Jayaraman, A. (2005). Optimization of reporter cells for expression profiling in a microfluidic device. Biomed Microdevices *7*, 213-222.

doi:10.1016/S0022-2836(03)00423-6 J. Mol. Biol. (2003) 329, 611–622

JMB Available online at www.sciencedirect.com SCIENCE DIRECT®

Prudent Modeling of Core Polar Residues in Computational Protein Design

Daniel N. Bolon[1], Joshua S. Marcus[1], Scott A. Ross[2] and Stephen L. Mayo[2]*

[1]*Biochemistry and Molecular Biophysics Option, California Institute of Technology, Mail Code 114-96, Pasadena CA 91125, USA*

[2]*Howard Hughes Medical Institute and Division of Biology MS 147-75, California Institute of Technology, 1200 East California Boulevard, Mail Code 114-96, Pasadena CA 91125, USA*

**Corresponding author*

Hydrogen bond interactions were surveyed in a set of protein structures. Compared to surface positions, polar side-chains at core positions form a greater number of intra-molecular hydrogen bonds. Furthermore, the majority of polar side-chains at core positions form at least one hydrogen bond to main-chain atoms that are not involved in hydrogen bonds to other main-chain atoms. Based on this structural survey, hydrogen bond rules were generated for each polar amino acid for use in protein core design. In the context of protein core design, these prudent polar rules were used to eliminate from consideration polar amino acid rotamers that do not form a minimum number of hydrogen bonds. As an initial test, the core of *Escherichia coli* thioredoxin was selected as a design target. For this target, the prudent polar strategy resulted in a minor increase in computational complexity compared to a strategy that did not allow polar residues. Dead-end elimination was used to identify global minimum energy conformations for the prudent polar and no polar strategies. The prudent polar strategy identified a protein sequence that was thermodynamically stabilized by 2.5 kcal/mol relative to wild-type thioredoxin and 2.2 kcal/mol relative to a thioredoxin variant whose core was designed without polar residues.

Keywords: protein design; hydrogen bond; energy function; thermodynamic stability; protein core

Introduction

For protein design to succeed, the underlying model must accurately describe the physics involved in protein thermodynamics. Under this principle, computational protein design has emerged as a powerful tool for studying the forces involved in stabilizing native protein structures. A design cycle,[1] iterating between computational modeling and experimental analysis of designed sequences, has been used to improve the predictive capability of force fields for protein design.[2] Using a protein design cycle, it was found that including a benefit for hydrophobic burial improved the correlation between predicted and experimental stability.[1] The energetic benefit of hydrophobic atoms interacting with a hydrophobic environment compared to an aqueous environment had previously been demonstrated by solvent transfer experiments.[3] The unfavorable interaction of solvent with hydrophobic molecules, compared to solvent–solvent interactions gained upon hydrophobic burial, is believed to explain this energetic benefit.[4] The burial of hydrophobic residues in the cores of folded proteins is thought to be a dominant force in stabilizing the native state over the denatured state.[5] Given the energetic benefit of burying hydrophobic atoms, protein design methods commonly restrict core positions to hydrophobic amino acids.[1,6–13] In addition to maximizing hydrophobic burial, the restriction of core residues to hydrophobic identity results in a significant reduction in combinatorial complexity compared to the inclusion of polar residues.

Polar groups form favorable electrostatic

Abbreviations used: MCU, main-chain atoms that do not make MC/MC hydrogen bonds; DSC, differential scanning calorimetry; t_m, melting temperature; ΔH, enthalpy of unfolding; ΔC_p, difference in heat capacity between the unfolded and folded state; ΔG, free energy of unfolding; m value, slope of ΔG *versus* denaturant concentration; C_m, midpoint of guanidinium chloride unfolding transition; wt, wild-type.

E-mail address of the corresponding author: steve@mayo.caltech.edu

0022-2836/03/$ - see front matter

interactions with aqueous solvent.[14] The burial of polar groups within protein structures shields polar atoms from solvent and results in an unfavorable electrostatic desolvation energy. The net electrostatic effect of burying a polar group can be favorable and will depend on intra-molecular interactions as well as desolvation.

Analysis of folded protein structures indicates that many polar side-chain atoms are buried.[15,16] We previously noted a strong positive correlation between protein size and the number of polar residues at core positions.[17] This correlation suggests that polar residues may play a structural role in energetically stabilizing the folded state and energetically distinguishing a single well-folded state from alternative conformations.[17] The mutation of polar residues involved in intra-molecular hydrogen bonding to hydrophobic residues frequently results in thermodynamically destabilized proteins.[18–21] In a survey of protein structures, McDonald & Thornton[22] found that the vast majority of buried polar atoms form intra-molecular hydrogen bonds.

These results suggest that protein core design could be improved by including polar residues provided that their hydrogen-bond potential is largely satisfied. With the intent of constructing hydrogen bond rules for protein core design, we analyzed hydrogen bonding in a previously published set of unrelated high-resolution protein structures.[23] Hydrogen bond rules developed based on this survey together with a geometrically softened hydrogen bond potential form a prudent method for modeling polar residues in protein cores. This prudent polar strategy for core optimization was incorporated into the ORBIT (optimization of rotamers by iterative techniques) protein design software.[9] As an initial test, core residues of *Escherichia coli* thioredoxin were designed with prudent polar rules as well as with no polar residues. The resulting sequences from the prudent polar and no polar designs were both thermodynamically stabilized relative to wild-type (wt) thioredoxin. At 25 °C the prudent polar molecule was stabilized by 2.2 kcal/mol relative to the no polar molecule, indicating the benefit of including polar residues in protein core design.

Results

Using a previously published dataset of 100 unrelated, high-quality protein crystal structures at better than 1.7 Å resolution,[23] we classified side-chain positions as core, boundary, or surface using a previously published algorithm.[24] Positions classified as core are highly solvent-inaccessible, while positions classified as surface are largely solvent-accessible (Figure 1). Positions classified as boundary have an intermediate solvent accessibility profile compared to surface and core (data not shown).

In the same data set, we compared hydrogen bonding at surface and core positions for the following group of polar amino acids: Ser, Thr, Asn, Asp, Gln, Glu, Lys, and Arg (Table 1). In this dataset, Richardson and co-workers previously added hydrogen atoms whose positions were based on both covalent and non-covalent considerations.[23] We find that the majority of these polar side-chains at core positions form intra-molecular hydrogen bonds consistent with the previous findings of McDonald & Thornton.[22] In contrast, the majority of the polar side-chains at surface positions do not make intra-molecular hydrogen bonds. Using a geometric hydrogen bond definition, as described later, we find that most polar residues at core positions (62.7%) make hydrogen bonds to main-chain atoms (Table 1). Furthermore, the majority of polar residues at core positions (53.6%) make at least one hydrogen bond to a MCU atom (main-chain atom not involved in hydrogen bonding to other main-chain atoms). Very few polar core residues (7.3%) form hydrogen bonds to crystallographically defined water molecules, while a slightly greater proportion of surface polar residues (17.4%) form these interactions. At boundary positions, where

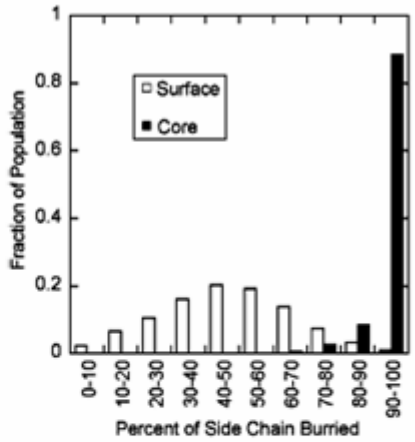


Figure 1. Solvent-accessible surface area (ASA) distribution of side-chains classified as surface and core. The fraction buried is the ratio of ASA in the folded state (surfacing all side-chain atoms while excluding all other atoms) to ASA in the reference state (surfacing all side-chain atoms while excluding only the local backbone N, C, C^{α}, and O atoms).

Table 1. Hydrogen bonding of polar side-chains

	Core	Surface
Number of side-chains	1835	3360
Percent with no intra-molecular HB	15.0	63.5
Percent with HB to MC atom	62.7	16.3
Percent with HB to MCU atom	53.6	12.7
Percent with HB to crystallographic water	7.3	17.4

Includes Ser, Thr, Asn, Asp, Gln, Glu, Lys, and Arg.

the geometric classification criteria are insufficient to determine if a side-chain will interact with solvent or bury itself in the protein core, polar residues make an intermediate number of intramolecular hydrogen bonds compared to surface and core positions (data not shown).

Hydrogen bonding by core side-chains was analyzed for each polar amino acid type (Table 2). The partitioning of each amino acid between core and surface or boundary positions divides polar residues into two groups. In the first group (Ser, Thr, Asn, Asp, Gln, Glu, Lys, and Arg) relatively few are located at core positions (11–32%), while more than 65% of Cys, Tyr, and Trp residues and 45% of His residues are found at core positions. Amino acids in the first group (Ser, Thr, etc.) make more total hydrogen bonds and more hydrogen bonds to MCU atoms than amino acids in the second group (Cys, His, Tyr, Trp). Within the first group, as the number of hydrogen bond donors and acceptors per amino acid type increases, so does the total number of identified hydrogen bonds per residue. From this analysis, hydrogen bonding requirements for polar core residues were constructed (Table 3). These rules require core polar residues to make a minimum number of total hydrogen bonds and a minimum number of hydrogen bonds to desirable partners. Desirable hydrogen bond partners are polar atoms that are likely to be buried from solvent and are defined here as MCU atoms whose hydrogen bond vectors are oriented towards the core as well as atoms of side-chains that hydrogen bond to MCU atoms. For example, in order for Asn to be considered at a core position it must at a minimum form one hydrogen bond to a desirable partner and one additional hydrogen bond (two total H-bonds, one desirable H-bond). The number of total H-bonds required increases as the number of donors and acceptors per amino acid type increases while the number of desirable H-bonds required is never greater than one.

Table 2. Side-chain hydrogen bonding at core positions

Amino acid	Total number	Core number	Percent core	HB[a] per core residue	HB_{MCU}[b] per core residue
Ser	1204	360	30	1.5	0.6
Thr	1131	365	32	1.4	0.5
Asn	915	250	27	2.5	0.9
Asp	1134	248	22	2.2	0.9
Gln	633	131	21	2.3	0.9
Glu	999	156	16	1.9	0.7
Lys	1046	116	11	2.2	0.6
Arg	718	209	29	4.7	1.4
Cys	335	240	72	0.3	0.1
His	376	171	46	0.2	0.3
Tyr	691	452	65	1.0	0.3
Trp	278	192	69	0.9	0.3

[a] Total hydrogen bonds.
[b] Hydrogen bonds to main-chain atoms that are not involved in hydrogen bonds to other main-chain atoms.

Table 3. HB requirements for core positions

Amino acid	Total H-bonds[a]	Desired H-bonds[b]
Ser	1	1
Thr	1	1
Asn	2	1
Asp	2	1
Gln	2	1
Glu	2	1
Lys	2	1
Arg	4	1
Cys[c]	–	–
His[c]	–	–
Tyr	0	0
Trp	0	0

[a] Minimum number of H-bonds required for core residues.
[b] Minimum number of H-bonds to MCU atoms or to the atoms of side-chains that hydrogen bond to MCU atoms.
[c] The H-bonding trends seen in the database survey were deemed insufficient to generate prudent hydrogen bond requirements for these residues.

For the thioredoxin core design target, the amino acid identity of 32 positions was optimized. The wt sequence at these positions contains two Tyr, two Thr, and one Asp, but is otherwise hydrophobic. For this design target, three different design strategies were employed: no polar, prudent polar, and unrestricted polar. For the no polar strategy, polar residues were excluded from the core as in previous design efforts.[17] For the prudent polar strategy, core polar rotamers were eliminated if they did not meet the hydrogen bond requirements (Table 3). For the unrestricted polar strategy, polar rotamers were included without hydrogen bond requirements.

Compared to the no polar strategy, the prudent polar strategy included a total of 14 additional polar amino acid rotamers that satisfied the prudent polar hydrogen bond requirements. Six of these additional rotamers were at positions whose wild-type side-chain is polar: two rotamers at position 66 (wt Thr) and four rotamers at position 77 (wt Thr). The other eight additional rotamers were at positions whose wild-type side-chain is non-polar: one rotamer each at positions 7 (wt Leu), 27 (wt Phe), 53 (wt Leu), and 72 (Ile), and two rotamers at positions 12 (wt Phe), and 75 (wt Ile).

In terms of combinatorial complexity, the no polar strategy resulted in 10^{49} possible rotamer solutions, the prudent polar design strategy resulted in 10^{50} possible rotamer solutions, and the unrestricted polar strategy resulted in 10^{69} possible rotamer solutions. Computational optimization was performed using an algorithm based on the dead-end elimination (DEE) theorem[25–27] and proceeded rapidly for the no polar and prudent polar strategies, but stalled for the unrestricted polar strategy (Figure 2). The no polar global minimum energy conformation (GMEC) was found in 2.4 processor hours, while the prudent polar GMEC was found in 1.6 processor hours. After 180

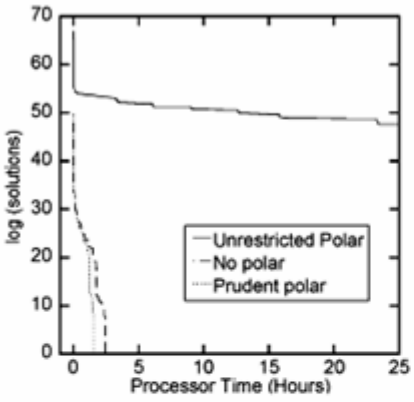


Figure 2. DEE optimization progress as a function of processor time for the different design strategies.

processor hours, the unrestricted polar optimization was stopped with greater than 10^{42} possible solutions remaining. The optimized sequence from the no polar strategy has six mutations (D26I, Y49F, L58I, T66L, Y70F, and T77A) corresponding to 81% identity relative to the wt sequence (Table 4). The optimized sequence from the prudent polar strategy has three mutations (D26I, L58I, Y70F) corresponding to 91% sequence identity relative to the wt sequence (Table 4).

Models of the no polar and prudent polar designed structures indicate similar core packing compared to wt, but different H-bond interactions (Figure 3). Of the six mutations in the no polar designed sequence (D26I, Y49F, L58I, T66L, Y70F, and T77A), five remove polar groups relative to wt. In the wt structure,[28] three of these residues (Y49, T66, and T77) form intra-molecular H-bonds that are recovered in the prudent polar design. The hydroxyl group of Y49 is within hydrogen bond distance (2.9 Å) to a side-chain carboxylate oxygen of D104. D104 is at a surface position and in contact with solvent. T66 is within hydrogen bond distance (2.9 Å) of the side-chain of D9, which makes putative hydrogen bonds with the backbone amides of both T66 and G65. T77 makes a putative hydrogen bond to the carbonyl oxygen of G74 in both the wt structure (2.7 Å) and the prudent polar design (2.8 Å). The carbonyl oxygen of G74 is buried from solvent and is not involved in hydrogen bond interactions with other main-chain atoms.

The no polar and prudent polar designed sequences have three common mutations relative to wt (D26I, L58I, and Y70F). In models of the no polar and prudent polar designs, I26, I58, and F70 all make extensive hydrophobic contacts. In the structure of wt thioredoxin, D26 is hydrogen bonded through a bridging water molecule to the carbonyl oxygen of C35 (the carbonyl oxygen of C35 makes a putative hydrogen bond to the main-chain amide of I38). Hydrophobic atoms surround L58 and the hydroxyl group of Y70 contacts solvent but does not form intra-molecular H-bonds.

The wt, no polar, and prudent polar sequences were all expressed in *E. coli* with high yield (20–50 mg purified protein/l culture). Circular dichroism (CD) of the wt, no polar, and prudent polar proteins indicates that all three proteins have similar secondary structure content (Figure 4). 1D ^{1}H NMR spectra of all three proteins show similar chemical shift dispersion (Figure 5(A)).

The number and pattern of resonances in the 2D ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the prudent polar variant are similar to those in the spectrum of the wild-type thioredoxin (Figure 5(B)), suggesting that the variant primarily populates a single conformation and that this conformation is similar to that of the wild-type protein. Several outlying peaks are nearly isochronous in the spectra of the two proteins. Seven of these backbone amide peaks (T8, G65, G71, G74, G84, A87, G97) were readily assigned by comparison with published thioredoxin chemical shifts.[29,30] The corresponding residues sample different regions of the structure. Similarity of the two proteins' chemical shifts thus indicates that the structural similarity inferred between the two proteins is global.

All three proteins denature cooperatively in the presence of increasing concentrations of guanidinium chloride (Figure 6(A)). Chemical denaturation indicated that the no polar protein and the prudent polar protein were thermodynamically

Table 4. Amino acid sequence comparison of core positions

Position	4	7	12	16	19	22	23	24	25	26	27	38	42	45	46	49	53	55	56	58	66	67	70	72	75	77	78	79	80	81	99	103
Secondary structure:	c	c	h	c	c	b	b	b	b	b	b	c	h	h	h	c	b	b	b	b	c	h	h	c	c	b	b	b	b	b	h	h
Sequences:																																
Wt	I	L	F	V	A	A	I	L	V	D	F	I	I	L	A	Y	L	V	A	L	T	A	Y	I	I	T	L	L	L	F	L	L
No polar	–	–	–	–	–	–	–	–	–	I	–	–	–	–	–	F	–	–	–	I	L	–	F	–	–	A	–	–	–	–	–	–
Prudent polar	–	–	–	–	–	–	–	–	–	I	–	–	–	–	–	–	–	–	–	I	–	–	F	–	–	–	–	–	–	–	–	–

Amino acid identity of positions included in the core optimization of thioredoxin. The first line indicates amino acid position. The second line indicates the secondary structure for each position (h—helix; b—beta strand; c—coil, neither helix nor strand), as determined by MOLMOL.[48] The amino acids of the wt protein, as well as the two designed proteins, are shown in single letter code.

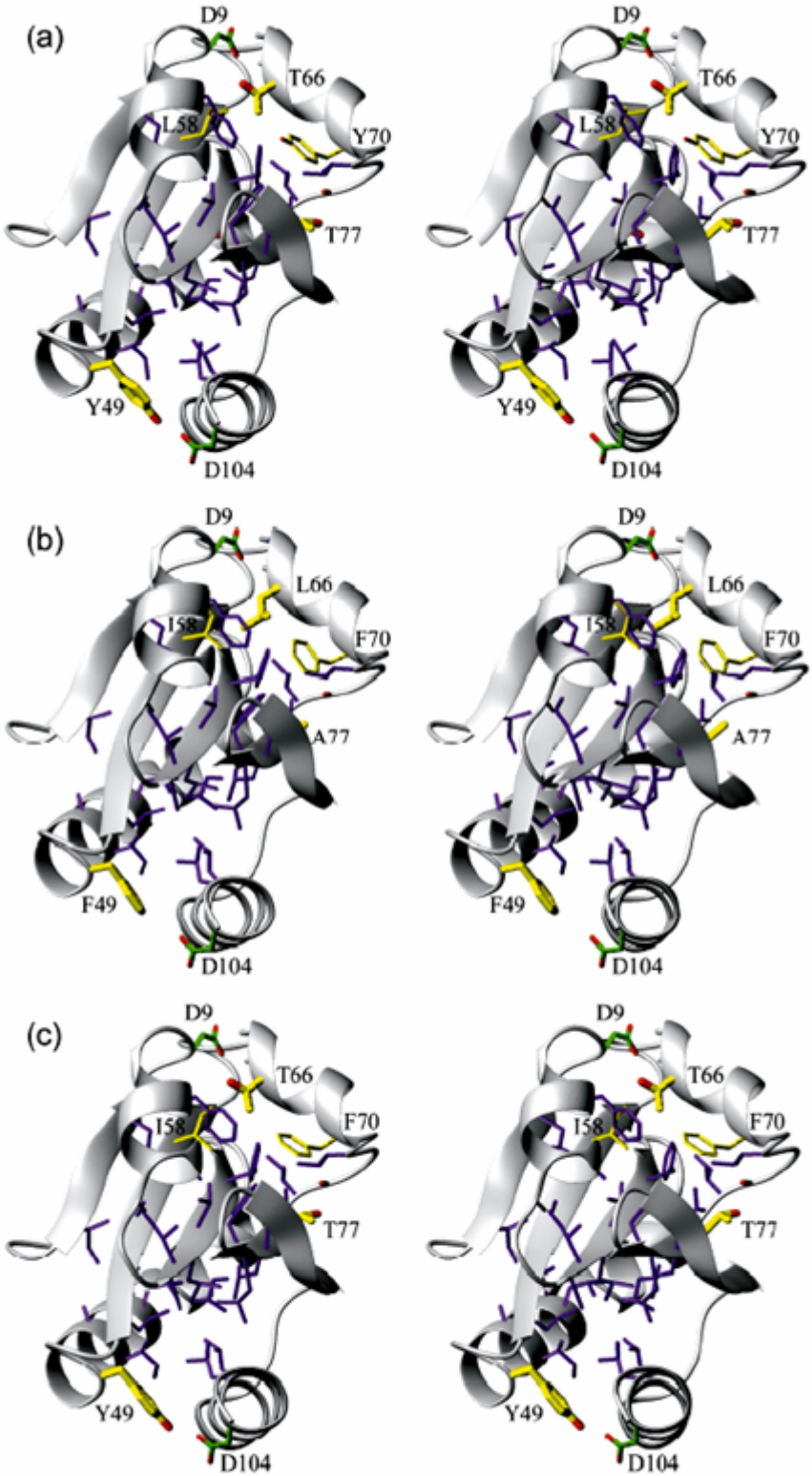


Figure 3. Three-dimensional molecular representations of the various thioredoxin structures. (a) Wt crystal structure.[28] (b) Model of the no polar design. (c) Model of the prudent polar design. Side-chains of the 26 core positions with identical amino acid identity are shown in purple. Side-chains at the six variable positions (26, located behind the β-sheet in these representations, 49, 58, 66, 70, and 77) are shown with carbon atoms colored yellow and oxygen atoms colored red. Side-chains at two fixed positions (9, 104) involved in H-bonds to core side-chains are shown with carbon atoms colored green and oxygen atoms colored red. Images were created with the program MOLMOL.[45]

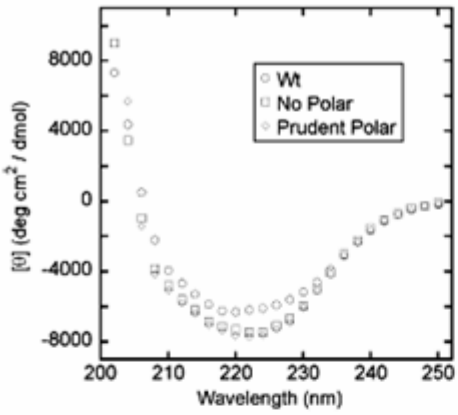


Figure 4. Far-UV circular dichroism spectra of the wt, no polar, and prudent polar proteins.

stabilized relative to the wt protein (Figure 6(A) and Table 5). The no polar and prudent polar proteins are 0.3 kcal/mol and 2.5 kcal/mol more stable than the wt protein, respectively. The *m*-value is similar for all three proteins, though it is slightly reduced for the no polar protein relative to the other two proteins. The results of differential scanning calorimetry (DSC) indicate that t_ms for the no polar protein and the prudent polar protein are both increased by 10 °C relative to the wt protein (Figure 6(B) and Table 5). DSC also shows that the ΔH of unfolding is 38% (41 kcal/mol) greater for the prudent polar protein than for the no polar protein (Table 5).

Discussion

The folding of a protein into a compact globular form requires the burial of both polar and non-polar atoms. Both statistical surveys[22] and mutational experiments[18–21] indicate the energetic benefit of forming intra-molecular hydrogen bonds between polar atoms that are buried from solvent. Buried polar atoms in proteins are shielded from solvent and therefore have unfavorable electrostatic desolvation energies.[14] In order to offset this unfavorable energy, buried polar atoms must make favorable intra-molecular interactions (e.g. hydrogen bonds).

The majority of buried polar atoms in folded globular protein structures are main-chain atoms.[22] The formation of secondary structure creates hydrogen bonds between many of these buried main-chain atoms; however, some polar main-chain atoms that orient towards the core are not involved in main-chain/main-chain hydrogen bonds. We note a strong positive correlation (Figure 7) between protein size and the number of core MCU atoms (main-chain atoms that are not involved in hydrogen bonds to other main-chain atoms). Polar side-chains at core positions that are able to form intra-molecular hydrogen bonds to core MCU atoms may energetically favor the folded state, provided they also largely fulfill their own hydrogen bond potential. In our database survey, we note that the majority of polar side-chains at core positions do form hydrogen bonds to MCU atoms (Table 1). As protein size increases and the number of MCU atoms increases (Figure 7) so should the demand for polar side-chains as hydrogen bond partners. This may partly account for the strong positive correlation that has been noted between protein size and the number of polar side-chains at core positions.[17]

Hydrogen bond rules for polar residues at core positions were constructed based on the database survey. These rules require that polar side-chains at core positions hydrogen bond to desirable partners as well as largely satisfying their own hydrogen bond potential (Table 3). The relatively small number of hydrogen bonds to crystallographically identified water molecules by core polar residues (Table 1) may be due to the energetically unfavorable burial of the partner water molecules from bulk solvent. Due to the limited number of hydrogen bonds to crystallographic water molecules revealed in our survey and the likely energetic penalty of burying a water molecule, water-mediated hydrogen bonds at core positions were not included in the prudent polar core design strategy. The prudent polar hydrogen bond requirements (Table 3) developed for core design will probably not transfer directly to boundary design because side-chains at boundary positions are capable of interacting with solvent. However, the prudent polar strategy could be applied to boundary positions that are likely to be highly buried based on the generic classification scheme of Marshall & Mayo.[13]

Given the variable geometry seen for hydrogen bonds in protein structures,[22] as well as the energetic benefit[18–21] of forming hydrogen bonds to buried polar atoms, a geometrically softened hydrogen bond potential was used in combination with the hydrogen bond requirements (Table 3) for core polar residues. For hydrogen bonds by core side-chains to template atoms (includes main-chain atoms and fixed side-chain atoms) whose geometries meet minimum criteria, an energetically beneficial hydrogen bond potential was applied. Given the clear importance of forming hydrogen bonds to desirable partners, this potential applied a minimum energetic benefit of 6.0 kcal/mol for desirable partners and a minimum energetic benefit of 4.0 kcal/mol for other partners. In the current ORBIT force field, these minimum hydrogen bond benefits are similar in magnitude to the energetic benefit of burying a methyl group. Thus, core polar rotamers that meet the hydrogen bond rules will be able to energetically compete with hydrophobic residues in the design of protein cores.

For the thioredoxin core design target, the unrestricted polar strategy failed to converge due to the large combinatorial complexity of the

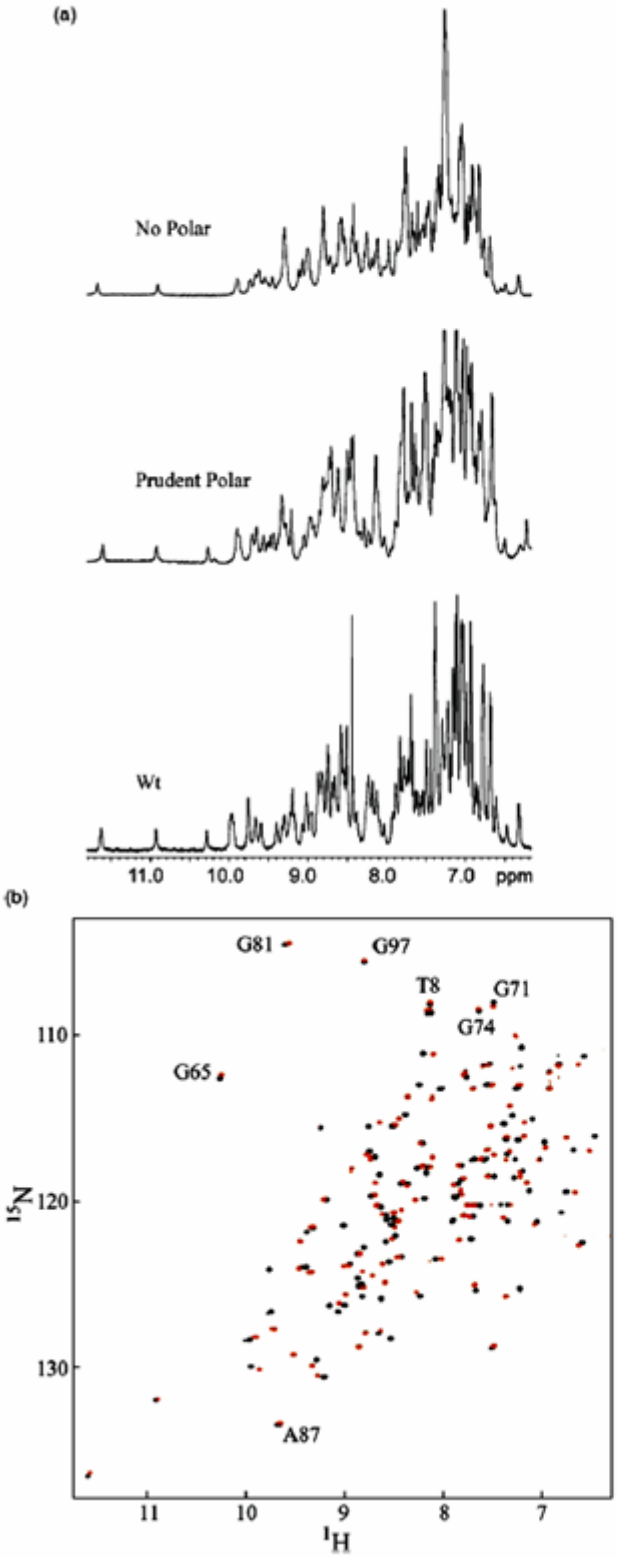


Figure 5. NMR spectra. (A) 1D ^{1}H NMR spectra of the wt, no polar and prudent polar proteins. (B) 2D ^{1}H–^{15}N HSQC NMR spectra of the wt (black) and prudent polar (red) proteins; seven presumed equivalent backbone amide peaks are identified.

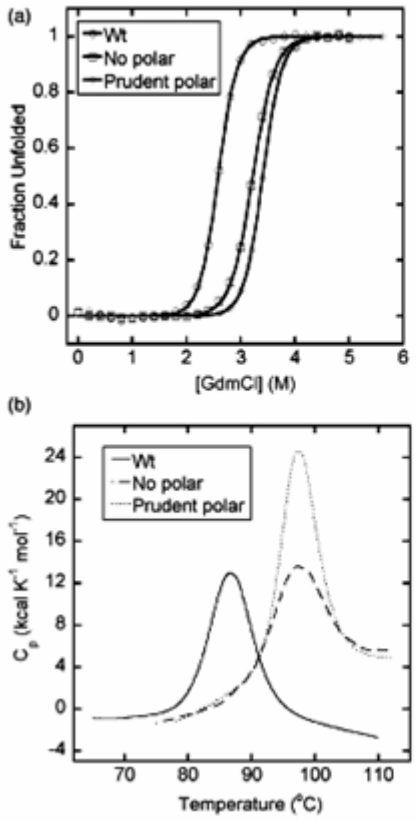


Figure 6. (A) Guanidinium chloride denaturation of the wt, no polar, and prudent polar proteins. (B) Thermal denaturation monitored by DSC.

calculation. The prudent polar strategy significantly reduced the combinatorial complexity relative to the unrestricted polar strategy, indicating that the hydrogen bond requirements effectively eliminated many polar side-chain rotamers from consideration. This decrease in combinatorial complexity results in the faster optimization trajectory that converges on a GMEC solution for the prudent polar strategy compared to the unrestricted polar strategy (Figure 2). The initial combinatorial complexity of the prudent polar strategy was slightly greater than that of the no polar strategy. However, the computational search was moderately faster for the prudent polar strategy than the no polar strategy. The energetic contributions of the prudent polar strategy simplify the search by emphasizing rotamer/template energies relative to rotamer/rotamer energies.[27]

Both the no polar and prudent polar designed sequences are very similar to the wild-type core sequence (81% and 91% identical, respectively). As the force field and rotamers used in the calculation were free from any knowledge of the wt structure, this result suggests that the wt core sequence is highly optimized for packing interactions.[31] Compared to the no polar sequence, the prudent polar sequence recovers three wt residues (Y49, T66, T77) involved in intramolecular hydrogen bonding. These results suggest that the wt thioredoxin core is highly optimized to form hydrogen bonds between buried polar atoms; in the wt structure[23] these amino acids (Y49, T66, T77) are positioned to make hydrogen bonds that fulfill the database-derived hydrogen bond rules. Our database survey as well as the survey of McDonald & Thornton[22] indicates that naturally occurring globular proteins tend to form hydrogen bonds to the majority of buried polar atoms. Given this observation, a conservative design strategy would leave wt core polar residues fixed and optimize the identity of core positions whose wt identity is hydrophobic. This strategy should yield protein sequences that form well-folded structures. However, given the highly coupled nature of protein cores, optimizing the identity of all core side-chains using a prudent strategy for modeling core polar residues can identify a solution with improved thermodynamic stability.

The sequences predicted for the thioredoxin core by the no polar and prudent polar strategies both form folded protein structures similar to wt based on the CD and proton NMR spectra (Figures 4 and 5).

The increase in stability of the no polar and prudent polar designed proteins relative to the wt protein (Table 5) can largely be explained by the D26I mutation. The previously described D26A mutation results in a protein that is thermodynamically stabilized relative to wt.[32] This increase in

Table 5. Experimental thermodynamic data

Protein	ΔG^a (kcal mol^{-1} M^{-1})	C_m^b (M)	m-value^c (kcal mol^{-1})	t_m^d (°C)	ΔH_{vh}^e (kcal mol^{-1})	ΔH_{cal}^f (kcal mol^{-1})	β^g
Wt[h]	9.0	2.6	3.4	87	117	117	1.00
No polar	9.3	3.2	2.9	97	106	106	1.00
Prudent polar	11.5	3.4	3.4	97	148	147	1.01

[a] Free energy of unfolding at 25 °C.
[b] Midpoint of guanidinium chloride unfolding transition.
[c] Slope of ΔG *versus* denaturant concentration.
[d] Calorimetrically determined melting temperature.
[e] van't Hoff enthalpy.
[f] Calorimetric enthalpy.
[g] $\Delta H_{cal}/\Delta H_{vh}$.
[h] From Bolon & Mayo.[17]

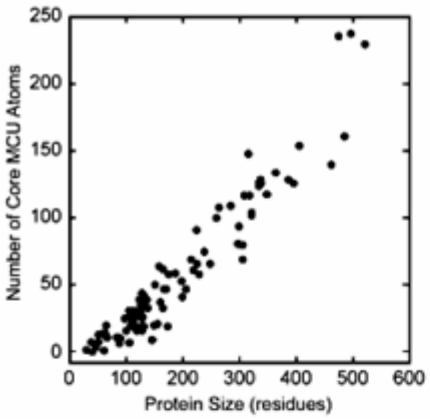


Figure 7. Correlation of the number of core MCU atoms with protein size.

stability is attributed to the removal of a buried polar group, whose hydrogen bond potential is largely unsatisfied, from the protein core.[32] The D26I mutation results in a similar increase in stability.[17] The stability of D26I in an otherwise wild-type thioredoxin (12.1 kcal/mol) is slightly greater than that of the prudent polar thioredoxin (11.5 kcal/mol). This indicates that the L58I and Y70F mutations in the prudent polar thioredoxin are slightly destabilizing relative to D26I thioredoxin. The physical basis for this slight destabilization is not readily apparent from analysis of the native thioredoxin structure. While the prudent polar thioredoxin molecule is slightly less stable than a molecule with the D26I mutation alone, in terms of evaluating protein core design methods, the relevant comparison is between prudent polar thioredoxin and no polar thioredoxin.

The 2.2 kcal/mol increase in stability for the prudent polar thioredoxin design compared to the no polar thioredoxin design indicates the advantage of the prudent polar strategy over the previously used no polar strategy for protein core design. The energetic difference between the prudent polar and no polar molecules can largely be accounted for by the interactions of T66 and T77 in prudent polar thioredoxin (L66 and A77 in no polar thioredoxin). Both T66 and T77 form hydrogen bonds to buried atoms in prudent polar and wt thioredoxin. Individual mutations at these positions to hydrophobic identity in the wt background each result in about a 1.0 kcal/mol decrease in stability.[17]

The results of our database survey suggest that buried polar interactions are largely fulfilled in most protein structures, indicating that buried polar interactions are important for protein stability. The prudent polar calculation on thioredoxin did not predict new polar residues in the core, suggesting that buried polar interactions are highly optimized in wild-type thioredoxin. It will be interesting to examine the results of prudent polar core optimization of other protein targets for new buried polar interactions and to determine their experimental effect on protein stability and conformational specificity.

Conclusions

The prudent polar strategy improves protein core design relative to a no polar strategy by including polar amino acids in the design when they are likely to improve thermodynamic stability. The specified hydrogen bond requirements eliminate most core polar amino acid rotamers from consideration, resulting in a tractable increase in combinatorial complexity. The prudent polar strategy should improve core design for large protein design applications, as larger proteins tend to have a greater number of core MCU atoms. In addition to protein core design, the prudent polar strategy may be useful in the modeling of other design targets that include buried polar groups such as protein–protein and protein–small molecule interfaces including computational enzyme design.[33]

Here, we use geometric hydrogen bond criteria to identify core polar amino acid rotamers that are likely to be energetically favorable. We model core hydrogen bonds using a geometric potential that is adjusted to energetically compete with hydrophobic amino acids given other terms in our force field.[34] The modeling of electrostatic interactions in protein design is an active area of current research.[34,35] As more accurate electrostatic models are developed for protein design they can readily be used in a prudent polar strategy in place of the current energy functions.

Materials and Methods

Database analysis and construction of hydrogen bond rules for protein core design

We analyzed hydrogen bonding in a previously published dataset of 100 high-resolution protein structures.[23] Richardson and co-workers previously modified the protein structures in this dataset by adding experimentally undetermined hydrogen atoms and positioning them to optimize hydrogen bonding and minimize van der Waals clashes.

Side-chains were classified as core, boundary, or surface based on the orientation of the C^{α}–C^{β} vector relative to a solvent-accessible surface computed with the template main-chain heavy atoms (N, C, C^{α}) as previously described.[24] A position was classified as core if the distance from its C^{α}, along its C^{α}–C^{β} vector, to the solvent-accessible surface was greater than 5.0 Å, and if the shortest distance from its C^{β} to the surface was greater than 2.0 Å. A position was classified as surface if the sum of the distance from its C^{α}, along its C^{α}–C^{β} vector, to the solvent-accessible surface and the shortest distance from its C^{β} to the surface was less than 2.7 Å. Positions not classified as core or surface were classified as boundary. The same geometric criteria were used to

classify main-chain carbonyl and amide groups. For classification purposes, the bond distances of the N–H and C–O bonds were adjusted to 1.54 Å to match with the C^{α}–C^{β} bond length for which the classification criteria were originally parameterized.

Side-chain accessible surface areas (ASA) were calculated with the Connolly algorithm,[36] using the Lee & Richards definition of solvent-accessible surface area[37] with atomic radii from the DREIDING force field,[38] an atomic add on radius of 1.4 Å, and a probe radius of 0 A. The fraction buried was calculated as the ratio of ASA in the folded state (surfacing all side-chain atoms while excluding all other atoms) to ASA in the reference state (surfacing all side-chain atoms while excluding only the local backbone N, C, C^{α}, and O atoms).

For our hydrogen bond analysis, we used the geometric hydrogen bond definition of McDonald & Thornton.[22] This hydrogen bond definition requires a donor–hydrogen-acceptor angle of greater than 90° and a hydrogen-acceptor distance of less than 2.5 Å, which corresponds to a heavy atom donor–acceptor distance of approximately 3.5 Å for a linear donor–hydrogen-acceptor angle. Hydrogen bond donors and acceptors were those defined by McDonald & Thornton.[22] All hydrogen bonds meeting this criteria were identified within each protein structure in the dataset. In a subsequent step, side-chain/main-chain hydrogen bonds were searched to identify side-chain hydrogen bonds to MCU atoms.

Computational modeling

Simulations were performed using coordinates from the 1.7 Å X-ray structure of thioredoxin.[28] All side-chains identified as core were optimized in the calculations with the exception of Trp28, Cys32, and Cys35. The side-chain of Trp28 adopts a non-canonical conformation and is not described in our rotamer library, while Cys32 and Cys35 form an intra-molecular disulfide bond. The remainder of the structure (including Trp28, Cys32 and Cys35) was held fixed during the optimizations. All optimized side-chains were modeled using a backbone dependent rotamer library[39] that was expanded about $\chi 1$ and $\chi 2$ dihedral angles for aromatic residues, expanded about $\chi 1$ dihedral angles for aliphatic residues, and unexpanded for polar residues as previously described.[24]

The calculations used an energy function based on the DREIDING force field.[38] A Lennard-Jones 12-6 potential with van der Waals radii scaled by 0.9 was applied to all rotamer-template and rotamer–rotamer interactions as previously described.[8] Electrostatic energies were computed using a distance-dependent dielectric constant of 40r (where r is the inter-atomic distance) with partial atomic charges as previously described.[40] A surface area-based atomic solvation potential was applied to hydrophobic atoms. Surface areas for the solvation potential were calculated using the Connolly algorithm[36] and the Lee & Richards[37] definition of solvent-accessible surface area with unscaled atomic radii and an add-on radius of 1.4 Å. A hydrophobic burial benefit of 48 cal/mol Å^2 and a hydrophobic exposure penalty of 76.8 cal/mol Å^2 were used with the two-body decomposition method of Street & Mayo.[41] The solvation potential was not applied to polar atoms.

For the no polar and unrestricted polar calculations, an explicit hydrogen bond potential was applied to all groups that contained hydrogen bond donors and/or acceptors.[40] This hydrogen bond potential consisted of distance-dependent and angle-dependent terms:

$$E_{\mathrm{HB}} = D_0\left\{5\left(\frac{R_0}{R}\right)^{12} - 6\left(\frac{R_0}{R}\right)^{10}\right\}F(\theta,\phi,\varphi)$$

where R_0 (2.8 Å) is the heavy atom donor–acceptor equilibrium distance, D_0 (8 kcal/mol) is the well-depth, and R is the heavy atom donor–acceptor distance. The angle term is essentially a $\cos^4$ term and is designed to penalize unfavorable geometries based on the hybridization state of the donor and acceptor atoms as previously described.[40] For the no polar design, core positions were restricted to the following seven amino acids: Ala, Val, Leu, Ile, Phe, Tyr, Trp. For the unrestricted polar calculations, core positions were restricted to the following amino acids: Ala, Val, Leu, Ile, Phe, Tyr, Trp, Ser, Thr, Asn, Asp, Gln, Glu, Lys, Arg.

For the prudent polar calculation, core positions were initially restricted to the following amino acids: Ala, Val, Leu, Ile, Phe, Tyr, Trp, Ser, Thr, Asn, Asp, Gln, Glu, Lys, Arg. An explicit hydrogen bond potential was applied to all groups that contained hydrogen bond donors and/or acceptors. For rotamer–rotamer interactions (those between optimized side-chain rotamers) the hydrogen bond potential was as described above. For interactions between core rotamers and template atoms, the potential differed from the default 12-10 potential described above in that it included an angle-independent minimum energetic benefit:

$$E_{\mathrm{HB}} = -D_1 + (D_0 - D_1)\left\{5\left(\frac{R_0}{R}\right)^{12} - 6\left(\frac{R_0}{R}\right)^{10}\right\}F(\theta,\phi,\varphi)$$

where D_1 is the minimum energetic benefit. In order to avoid benefiting grossly improper geometry, the hydrogen bond potential was only applied when $2.5\ \text{Å} < R < 3.2\ \text{Å}$, and the donor–hydrogen-acceptor angle was greater than 90 degrees. This potential was used in part to account for the altered dielectric constant in protein cores as well as to account for backbone flexibility and the use of discrete rotamers. The low dielectric constant in the protein core should increase the electrostatic component of hydrogen bonds in the protein core.[42] The minimum energy benefit effectively reduces the angle dependence of the hydrogen bond potential, and is intended to energetically benefit residues that can form geometrically favorable hydrogen bonds given slightly greater side-chain and backbone flexibility. These calculations used a minimum energetic benefit of 6.0 kcal/mol for hydrogen bonds with desirable partners and 4.0 kcal/mol for all other hydrogen bonding by core side-chains.

Desirable hydrogen bond partners were identified in the template atoms (portions of the molecule held fixed during optimization) using the same analysis program used for the database survey. During the rotamer-template energy calculation, polar residues that did not make the required number of hydrogen bonds (Table 3) were eliminated.

For all design strategies (unrestricted polar, no polar, and prudent polar), an algorithm based on the dead-end elimination theorem[25–27] was used to search for the global minimum energy conformation. Calculations were performed on six 195 MHz R10000 processors (Silicon Graphics, Mountain View, CA) running in parallel.

Mutagenesis and protein purification

The genes for all variants were constructed by inverse PCR[43] using the gene for wt thioredoxin (Invitrogen) cloned into pET-11a (Novagen). Mutations were confirmed by DNA sequencing. BL21(DE3) cells (Invitrogen) were used for protein expression. Cells were grown to mid log phase and induced with 0.5 mM IPTG (ICN) for three hours at 37 °C. Cells were lysed by sonication and centrifuged twice at 20,000*g* for 30 minutes. Acetonitrile (EM Science) was added to the soluble fraction to a final concentration of 60%. After centrifugation, the supernatant was evaporated to half volume and purified by reverse phase high performance liquid chromatography using an acetonitrile/water gradient containing 0.1% (w/v) trifluoroacetic acid (Applied Biosystems). Protein identities were confirmed by mass spectrometry. Protein concentration was determined by UV absorbance in 7 M GdmCl (ICN) based on extinction coefficients at 280 nm of 13,700 $M^{-1}\,cm^{-1}$ for the wt molecule,[44] 12,400 $M^{-1}\,cm^{-1}$ for the prudent polar molecule, and 11,100 $M^{-1}\,cm^{-1}$ for the no polar molecule.

CD analysis

CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit and using a 1 cm path length cell. Protein samples were 5 μM in 50 mM sodium phosphate at pH 7.0. Guanidinium chloride denaturations were monitored at 219 nm at 25 °C for comparison to previously published data.[17,45] ΔG values, m values, and error estimates were obtained by fitting the denaturation data to a two-state transition as described.[46] GdmCl concentration was determined by refractometry.

Differential scanning calorimetry

Protein samples at ~0.7 mg/ml were dialyzed against 50 mM sodium phosphate buffer at pH 7.0. DSC data were collected on a N-DSC II instrument from Calorimetric Sciences Corp.[47] Samples were degassed under vacuum for 15 minutes prior to scanning. Scans were performed at a rate of 1.0 °C/minute. Data were analyzed using the program cpcalc (Applied Thermodynamics).

NMR spectroscopy

For 1D ^1H spectra, unlabeled protein samples at 0.4 mM were prepared in 90:10 H_2O:^2H_2O buffered at pH* 7.0 with 50 mM sodium phosphate. Spectra were acquired on a Varian Inova 600 MHz spectrometer at 25 °C.

For 2D ^1H–^{15}N HSQC spectra, labeled protein was obtained from cells grown in minimal growth medium containing $(^{15}NH_4)_2SO_4$ as the sole nitrogen source. Samples for NMR spectroscopy (500 μl) were prepared with 0.3 mM (prudent polar) or 1.0 mM (wild-type) protein in 95:5 H_2O:^2H_2O buffered at pH 7.0 with 50 mM sodium phosphate. NMR spectroscopy was performed on a Varian Inova 600 MHz spectometer at 25 °C. ^1H–^{15}N HSQC spectra were acquired with 256 t_1 × 1349 t_2 complex points. Eight transients were acquired per t_1 increment, with a 1.5 seconds relaxation delay. The ^1H spectral width was 9000 Hz with the carrier centered on the water frequency; the ^{15}N spectral width was 2500 Hz. Data were doubled in f_1 by linear prediction; zero filling in both dimensions yielded transformed data matrices of 1024 × 2048 complex points.

Acknowledgements

We thank P. S. Shah for aid in protein expression and purification, P. Strop for helpful discussions and M. Ary for critical comments on the manuscript. This research was supported by the Howard Hughes Medical Institute and the Ralph M. Parsons Foundation (S.L.M.), the Helen G. and Arthur McCallum Foundation, the Evelyn Sharp Graduate Fellowship, and grant GM07616 from the National Institutes of Health (D.N.B.).

References

1. Dahiyat, B. I. & Mayo, S. L. (1996). Protein design automation. *Protein Sci.* **5**, 895–903.
2. Gordon, D. B., Marshall, S. A. & Mayo, S. L. (1999). Energy functions for protein design. *Curr. Opin Struct. Biol.* **9**, 509–513.
3. Nozaki, Y. & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **246**, 2211–2217.
4. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63.
5. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
6. Hellinga, H. W. & Richards, F. M. (1994). Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl Acad Sci. USA*, **91**, 5803–5807.
7. Desjarlais, J. R. & Handel, T. M. (1995). *De novo* design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018.
8. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc. Natl Acad Sci. USA*, **94**, 10172–10177.
9. Dahiyat, B. I. & Mayo, S. L. (1997). *De novo* protein design: fully automated sequence selection. *Science*, **278**, 82–87.
10. Lazar, G. A., Desjarlais, J. R. & Handel, T. M. (1997). *De novo* design of the hydrophobic core of ubiquitin. *Protein Sci.* **6**, 1167–1178.
11. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305–318.
12. Street, A. G. & Mayo, S. L. (1999). Computational protein design. *Struct. Fold. Des.* **7**, R105–R109.
13. Marshall, S. A. & Mayo, S. L. (2001). Achieving stability and conformational specificity in designed proteins *via* binary patterning. *J. Mol. Biol.* **305**, 619–631.
14. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
15. Chothia, C. (1976). The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**, 1–14.
16. Lesser, G. J. & Rose, G. D. (1990). Hydrophobicity of amino acid subgroups in proteins. *Proteins: Struct. Funct. Genet.* **8**, 6–13.

17. Bolon, D. N. & Mayo, S. L. (2001). Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry*, **40**, 10047–10053.
18. Shirley, B. A., Stanssens, P., Hahn, U. & Pace, C. N. (1992). Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry*, **31**, 725–732.
19. Byrne, M. P., Manuel, R. L., Lowe, L. G. & Stites, W. E. (1995). Energetic contribution of side chain hydrogen bonding to thestability of staphylococcal nuclease. *Biochemistry*, **34**, 13949–13960.
20. Yamagata, Y., Kubota, M., Sumikawa, Y., Funahashi, J., Takano, K., Fujii, S. & Yutani, K. (1998). Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six tyrosine to phenylalanine mutants. *Biochemistry*, **37**, 9355–9362.
21. Takano, K., Yamagata, Y., Kubota, M., Funahashi, J., Fujii, S. & Yutani, K. (1999). Contribution of hydrogen bonds to the conformational stability of human lysozyme: calorimetry and X-ray analysis of six Ser to Ala mutants. *Biochemistry*, **38**, 6623–6629.
22. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
23. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747.
24. Dahiyat, B. I., Sarisky, C. A. & Mayo, S. L. (1997). *De novo* protein design: towards fully automated sequence selection. *J. Mol. Biol.* **273**, 789–796.
25. Desmet, J., Maeyer, M. D., Hazes, B. & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, **356**, 539–542.
26. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. (2000). Conformational splitting: a more powerful criterion for dead-end elimination. *J. Comput. Chem.* **21**, 999–1009.
27. Gordon, D. B., Hom, G. K., Mayo, S. L. & Pierce, N. A. (2003). Exact rotamer optimization for protein design. *J. Comput. Chem.* **24**, 232–243.
28. Katti, S. K., LeMaster, D. M. & Eklund, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. Mol. Biol.* **212**, 167–184.
29. Dyson, H. J., Holmgren, A. & Wright, P. E. (1989). Assignment of the proton NMR spectrum of reduced and oxidized thioredoxin: sequence-specific assignments, secondary structure, and global fold. *Biochemistry*, **28**, 7074–7087.
30. Chandrasekhar, K., Krause, G., Holmgren, A. & Dyson, H. J. (1991). Assignment of the ^{15}N NMR spectra of reduced and oxidized *Escherichia coli* thioredoxin. *FEBS Letters*, **284**, 178–183.
31. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl Acad Sci. USA*, **97**, 10383–10388.
32. Langsetmo, K., Fuchs, J. A. & Woodward, C. (1991). The conserved, buried aspartic acid in oxidized *Escherichia coli* thioredoxin has a pK_a of 7.5. Its titration produces a related shift in global stability. *Biochemistry*, **30**, 7603–7609.
33. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA*, **98**, 14274–14279.
34. Marshall, S. A., Morgan, C. S. & Mayo, S. L. (2002). Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* **316**, 189–199.
35. Havranek, J. J. & Harbury, P. B. (1999). Tanford–Kirkwood electrostatics for protein modeling. *Proc. Natl Acad. Sci. USA*, **96**, 11145–11150.
36. Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
37. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
38. Mayo, S. L., Olafson, B. D. & Goddard, W. A., 3rd (1990). DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909.
39. Dunbrack, R. L., Jr & Karplus, M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574.
40. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. (1997). Automated design of the surface positions of protein helices. *Protein Sci.* **6**, 1333–1337.
41. Street, A. G. & Mayo, S. L. (1998). Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.* **3**, 253–258.
42. Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P. & Tidor, B. (1998). Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Protein Sci.* **7**, 206–210.
43. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucl. Acids Res.* **17**, 6545–6551.
44. Holmgren, A. & Reichard, P. (1967). Thioredoxin 2: cleavage with cyanogen bromide. *Eur. J. Biochem.* **2**, 187–196.
45. Kelley, R. F., Shalongo, W., Jagannadham, M. V. & Stellwagen, E. (1987). Equilibrium and kinetic measurements of the conformational transition of reduced thioredoxin. *Biochemistry*, **26**, 1406–1411.
46. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
47. Privalov, G., Kavina, V., Freire, E. & Privalov, P. L. (1995). Precise scanning calorimeter for studying thermal properties of biological macromolecules in dilute solution. *Anal. Biochem.* **232**, 79–85.
48. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. see also pp. 29–32.

Edited by M. Levitt

(Received 26 February 2003; accepted 19 March 2003)

Conclusions

This thesis is a combined effort of technological innovation and scientific discorvery. The author presents novel microfluidic technologies capable of processing single cells to first strand cDNA in a parallel manner. With these devices, the first direct measure of transcriptional noise, as well as the first measurements of noise in single mammalian cells is undertaken.

The technology provides marked improvement on previous devices designed for cell processing for nucleic acid analysis. First, the sieve valve innovated by the author provides a more robust implementation for on-chip separation assays than previous schemes. In Ch. 3 the author shows columns can operate at a significantly higher range of pressures than the previous method. This provides the user flexibility in what pressures to run separation assays, whereas the prior implementation could only functionally operate at one defined pressure. The author also provided quantitative calibrations over six orders of magnitude for each step along the microfluidic process, allowing for absolute copy number quantification in single mammalian cells, or otherwise. First strand synthesis integration was shown to decrease variability of measurements five fold, as compared to product collection after the mRNA isolation step. The decreased variability is most likely due to ambient RNase activity during product collection. Although on-chip reverse transcription efficiency was shown to be less than benchtop, the author showed the ability to perform multiplex microfluidc RT- multiplex qPCR in the template range for single cell measurements, unattainable by benchtop methods. The gain in resolution is due to the utilization of a purified cDNA product, void of dNTPs, reverse transcriptase and RT buffer. Unlike benchtop first strand synthesis

coupled to qPCR, the microfluidic method allows for 100% of the first strand product to be utilized in the multiplex qPCR. Ten percent of product is more the norm for benchtop assays because inhibition of PCR from excess first strand reaction components is an issue at higher percentages. In the author's scheme, excess reaction components are washed away before collection of purified product.

The author utilized the novel technology presented in Ch. 5 to study transcriptional noise in single NIH/3T3 cells. Histograms of gene expression data indicated lognormal distributions for all genes measured, and for each stage of the cell cycle. These results are consistent with the literature on gene and protein distributions, where the end product is the result of many intracellular reactions propagating in a multiplicative fashion. The author calculated coefficient's of variation for the pooled dataset, as well as subsets of the data, to attempt to decrease extrinsic noise. The values obtained in NIH/3T3 cells were significantly higher than values obtained in *E. coli* and yeast. However, because the author quantified in terms of absolute copy numbers, as well as based on lognormal distributions, conclusions based on these comparisons should be made cautiously. The author also demonstrates noise calculations in the literature are misleading because the noise was calculated based on linear distributions, even when lognormal fits are more appropriate. The questions asked and results presented in Ch. 5 are evidence that technological innovation breeds novel science.

Bibliography

1. Bustin, S.A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**, 169-193 (2000).
2. Brady, G. Expression profiling of single mammalian cells--small is beautiful. *Yeast* **17**, 211-217 (2000).
3. Chiang, M.K. & Melton, D.A. Single-cell transcript analysis of pancreas development. *Dev Cell* **4**, 383-393 (2003).
4. Tietjen, I. et al. Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161-175 (2003).
5. Kamme, F. et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J Neurosci* **23**, 3607-3615 (2003).
6. Alsbo, C.W., Wrang, M.L., Moller, F. & Diemer, N.H. Is the AMPA receptor subunit GluR2 mRNA an early indicator of cell fate after ischemia? A quantitative single cell RT-PCR study. *Brain Res* **894**, 101-108 (2001).
7. Danik, M., Puma, C., Quirion, R. & Williams, S. Widely expressed transcripts for chemokine receptor CXCR1 in identified glutamatergic, gamma-aminobutyric acidergic, and cholinergic neurons and astrocytes of the rat brain: a single-cell reverse transcription-multiplex polymerase chain reaction study. *J Neurosci Res* **74**, 286-295 (2003).
8. Henne, J., Pottering, S. & Jeserich, G. Voltage-gated potassium channels in retinal ganglion cells of trout: a combined biophysical, pharmacological, and single-cell RT-PCR approach. *J Neurosci Res* **62**, 629-637 (2000).
9. Klein, D. Quantification using real-time PCR technology: applications and limitations. *Trends Mol Med* **8**, 257-260 (2002).
10. Lindqvist, N., Vidal-Sanz, M. & Hallbook, F. Single cell RT-PCR analysis of tyrosine kinase receptor expression in adult rat retinal ganglion cells isolated by retinal sandwiching. *Brain Res Brain Res Protoc* **10**, 75-83 (2002).
11. Maryanski, J.L., Attuil, V., Bucher, P. & Walker, P.R. A quantitative, single-cell PCR analysis of an antigen-specific TCR repertoire selected during an in vivo CD8 response: direct evidence for a wide range of clone sizes with uniform tissue distribution. *Mol Immunol* **36**, 745-753 (1999).
12. Peixoto, A., Monteiro, M., Rocha, B. & Veiga-Fernandes, H. Quantification of multiple gene expression in individual cells. *Genome Res* **14**, 1938-1947 (2004).
13. Silbert, S.C. Quantitative single-cell RT-PCR for opioid receptors and housekeeping genes. *Methods Mol Med* **84**, 107-140 (2003).
14. Wagatsuma, A. et al. Determination of the exact copy numbers of particular mRNAs in a single cell by quantitative real-time RT-PCR. *J Exp Biol* **208**, 2389-2398 (2005).
15. Bengtsson, M., Stahlberg, A., Rorsman, P. & Kubista, M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res* **15**, 1388-1392 (2005).

16. Rickman, D.S., Herbert, C.J. & Aggerbeck, L.P. Optimizing spotting solutions for increased reproducibility of cDNA microarrays. *Nucleic Acids Res* **31**, e109 (2003).
17. Das, M., Harvey, I., Chu, L.L., Sinha, M. & Pelletier, J. Full-length cDNAs: more than just reaching the ends. *Physiol Genomics* **6**, 57-80 (2001).
18. Wheeler, A.R. et al. Microfluidic device for single-cell analysis. *Anal Chem* **75**, 3581-3586 (2003).
19. Unger, M.A., Chou, H.P., Thorsen, T., Scherer, A. & Quake, S.R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113-116 (2000).
20. Thorsen, T., Maerkl, S.J. & Quake, S.R. Microfluidic large-scale integration. *Science* **298**, 580-584 (2002).
21. Chou, H.P., Unger, M.A. & Quake, S.R. A microfabricated rotary pump. *Biomedical Microdevices* **3**, 323-330 (2001).
22. Hansen, C.L., Sommer, M.O. & Quake, S.R. Systematic investigation of protein phase behavior with a microfluidic formulator. *Proc Natl Acad Sci U S A* **101**, 14431-14436 (2004).
23. Marcus, J.S., Anderson, W.F. & Quake, S.R. Microfluidic single cell mRNA isolation and analysis. *Anal Chem* **78**, 3084-3089 (2006).
24. Hong, J.W., Studer, V., Hang, G., Anderson, W.F. & Quake, S.R. A nanoliter-scale nucleic acid processor with parallel architecture. *Nat Biotechnol* **22**, 435-439 (2004).
25. Hansen, C.L., Skordalakes, E., Berger, J.M. & Quake, S.R. A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc Natl Acad Sci U S A* **99**, 16531-16536 (2002).
26. Fu, A.Y., Chou, H.P., Spence, C., Arnold, F.H. & Quake, S.R. An integrated microfabricated cell sorter. *Anal Chem* **74**, 2451-2457 (2002).
27. Fu, A.Y., Spence, C., Scherer, A., Arnold, F.H. & Quake, S.R. A microfabricated fluorescence-activated cell sorter. *Nat Biotechnol* **17**, 1109-1111 (1999).
28. Liu, J., Enzelberger, M. & Quake, S. A nanoliter rotary device for polymerase chain reaction. *Electrophoresis* **23**, 1531-1536 (2002).
29. Liu, J., Hansen, C. & Quake, S.R. Solving the "world-to-chip" interface problem with a microfluidic matrix. *Anal Chem* **75**, 4718-4723 (2003).
30. Marcus, J.S., Anderson, W.F. & Quake, S.R. Parallel picoliter RT-PCR assays using microfluidics. *Anal Chem* **78**, 956-958 (2006).
31. Alberts, B. et al. Molecular biology of the cell, Edn. 4. (Garland Publishing, New York; 1994).
32. Ovstebo, R., Haug, K.B., Lande, K. & Kierulf, P. PCR-based calibration curves for studies of quantitative gene expression in human monocytes: development and evaluation. *Clin Chem* **49**, 425-432 (2003).
33. Cardullo, R.A., Agrawal, S., Flores, C., Zamecnik, P.C. & Wolf, D.E. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* **85**, 8790-8794 (1988).
34. Becker, A.J. et al. Transcriptional profiling in human epilepsy: expression array and single cell real-time qRT-PCR analysis reveal distinct cellular gene regulation. *Neuroreport* **13**, 1327-1333 (2002).

35. Fink, L. et al. Real-time quantitative RT-PCR after laser-assisted cell picking. *Nat Med* **4**, 1329-1333 (1998).
36. Shi, X. et al. Gene-expression analysis of single cells-nested polymerase chain reaction after laser microdissection. *World J Gastroenterol* **9**, 1337-1341 (2003).
37. Durand, G.M., Marandi, N., Herberger, S.D., Blum, R. & Konnerth, A. Quantitative single-cell RT-PCR and Ca(2+) imaging in brain slices. *Pflugers Arch* (2005).
38. Baro, D.J. et al. Quantitative single-cell-reverse transcription-PCR demonstrates that A-current magnitude varies as a linear function of shal gene expression in identified stomatogastric neurons. *J Neurosci* **17**, 6597-6610 (1997).
39. Dixon, A.K., Richardson, P.J., Pinnock, R.D. & Lee, K. Gene-expression analysis at the single-cell level. *Trends Pharmacol Sci* **21**, 65-70 (2000).
40. Gentile, L. et al. Single-cell quantitative RT-PCR analysis of Cpt1b and Cpt2 gene expression in mouse antral oocytes and in preimplantation embryos. *Cytogenet Genome Res* **105**, 215-221 (2004).
41. Han, S.H. et al. Single-cell RT-PCR detects shifts in mRNA expression profiles of basal forebrain neurons during aging. *Brain Res Mol Brain Res* **98**, 67-80 (2002).
42. Liss, B. Improved quantitative real-time RT-PCR for expression profiling of individual cells. *Nucleic Acids Res* **30**, e89 (2002).
43. Stahlberg, A., Zoric, N., Aman, P. & Kubista, M. Quantitative real-time PCR for cancer detection: the lymphoma case. *Expert Rev Mol Diagn* **5**, 221-230 (2005).
44. Todd, R. & Margolin, D.H. Challenges of single-cell diagnostics: analysis of gene expression. *Trends Mol Med* **8**, 254-257 (2002).
45. Yan, L., Kaczorowski, G. & Kohler, M. One-Tube Protocol for Single-Cell Reverse Transcriptase-Polymerase Chain Reaction. *Analytical Biochemistry* **304**, 267-270 (2002).
46. Osawa, M. et al. Molecular characterization of melanocyte stem cells in their niche. *Development* **132**, 5589-5599 (2005).
47. Al-Taher, A., Bashein, A., Nolan, T., Hollingsworth, M. & Brady, G. Global cDNA amplification combined with real-time RT-PCR: accurate quantification of multiple human potassium channel genes at the single cell level. *Yeast* **17**, 201-210 (2000).
48. Brail, L.H. et al. Gene expression in individual cells: analysis using global single cell reverse transcription polymerase chain reaction (GSC RT-PCR). *Mutat Res* **406**, 45-54 (1999).
49. http://www.premierbiosoft.com/molecular_beacons/index.html
50. http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/guide.html
51. Stahlberg, A., Hakansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the reverse transcription reaction in mRNA quantification. *Clin Chem* **50**, 509-515 (2004).
52. Stahlberg, A., Kubista, M. & Pfaffl, M. Comparison of reverse transcriptases in gene expression analysis. *Clin Chem* **50**, 1678-1680 (2004).
53. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).

54. Swain, P.S., Elowitz, M.B. & Siggia, E.D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* **99**, 12795-12800 (2002).
55. Raser, J.M. & O'Shea, E.K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814 (2004).
56. 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).
57. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415-418 (2004).
58. Blake, W.J., M, K.A., Cantor, C.R. & Collins, J.J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637 (2003).
59. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* **98**, 8614-8619 (2001).
60. Thompson, D.M. et al. Dynamic gene expression profiling using a microfabricated living cell array. *Anal Chem* **76**, 4098-4103 (2004).
61. Wieder, K.J. et al. Optimization of reporter cells for expression profiling in a microfluidic device. *Biomed Microdevices* **7**, 213-222 (2005).
62. 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).
63. Keene, O.N. The Log Transformation Is Special. *Statistics in Medicine* **14**, 811-819 (1995).
64. Koch, A.L. Logarithm in Biology.1. Mechanisms Generating Log-Normal Distribution Exactly. *Journal of Theoretical Biology* **12**, 276-290 (1966).
65. Koch, A.L. Logarithm in Biology.2. Distributions Simulating Log-Normal. *Journal of Theoretical Biology* **23**, 251-268 (1969).
66. Welch, B.L. Distributions in Statistics - Continuous Univariate Distributions - Johnson,Nl and Kotz,S. *Journal of the Royal Statistical Society Series a-General* **135**, 433-433 (1972).
67. Wang, Y. et al. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**, 5860-5865 (2002).
68. Studer, V.H., G; Pandolfi, A.; Ortiz, M.; Anderson, W.F.; Quake, S.R. Scaling properties of a low actuation pressure microfluidic valve. *J Appl Phys* **95**, 393-398 (2004).
69. Dap cDNA was generated by amplifying dap mRNA in an RT-PCR. We then purified the PCR products and quantified the sample by spectrophotometry.
70. Xia, Y. & Whitesides, G.M. Soft Lithography. *Angew Chem Int Ed* **37**, 550-575 (1998).
71. Hansen, C.L. Microfluidic technologies for structural biology. *Ph.D. thesis* (2004).
72. Huang, Y. & Rubinsky, B. Flow-through micro-electroporation chip for high efficiency single-cell genetic manipulation. *Sens Actuators A* **104**, 205-212 (2003).
73. Khine, M., Lau, A., Ionescu-Zanetti, C., Seo, J. & Lee, L.P. A single cell electroporation chip. *Lab Chip* **5**, 38-43 (2005).

74. Olofsson, J. et al. Single-cell electroporation. *Curr Opin Biotechnol* **14**, 29-34 (2003).
75. Huang, Y. & Rubinsky, B. Microfabricated electroporation chip for single cell membrane permeabilization. *Sens Actuators A* **89**, 242-249 (2001).
76. Tegenfeldt, J.O. et al. Micro- and nanofluidics for DNA analysis. *Anal Bioanal Chem* **378**, 1678-1692 (2004).
77. Prinz, C., Tegenfeldt, J.O., Austin, R.H., Cox, E.C. & Sturm, J.C. Bacterial chromosome extraction and isolation. *Lab Chip* **2**, 207-212 (2002).
78. Anderson, R.C., Su, X., Bogdan, G.J. & Fenton, J. A miniature integrated device for automated multistep genetic assays. *Nucleic Acids Res* **28**, E60 (2000).
79. Studer, V. et al. Scaling properties of a low-actuation pressure microfluidic valve. *J Appl Phys* **95**, 393-398 (2004).
80. Liu, R.H., Yang, J., Lenigk, R., Bonanno, J. & Grodzinski, P. Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. *Anal Chem* **76**, 1824-1831 (2004).
81. Obeid, P. & Christopoulos, T. Continuous-flow DNA and RNA amplification chip combined with laser-induced fluorescence detection. *Analytica Chimica Acta* **494**, 1-9 (2003).
82. Yuen, P.K. et al. Microchip module for blood sample preparation and nucleic acid amplification reactions. *Genome Res* **11**, 405-412 (2001).
83. Zhou, X. et al. Determination of SARS-coronavirus by a microfluidic chip system. *Electrophoresis* **25**, 3032-3039 (2004).
84. Unger, M. et al. European Patent. **Vol. EP1463796.** (2004).
85. Vogelstein, B. & Kinzler, K.W. Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241 (1999).
86. Lee, C.-Y., Lee, G.-B., Lin, J.-L., Huang, F.-C. & Liao, C.-S. Integrated microfluidic systems for cell lysis, mixing/pumping and DNA amplification. *J. Micromech Microeng* **15**, 1215-1223 (2005).
87. McClain, M.A. et al. Microfluidic devices for the high-throughput chemical analysis of cells. *Anal Chem* **75**, 5646-5655 (2003).
88. Auroux, P.A., Koc, Y., deMello, A., Manz, A. & Day, P.J. Miniaturised nucleic acid analysis. *Lab Chip* **4**, 534-546 (2004).
89. Galloway, M. & Soper, S.A. Contact conductivity detection of polymerase chain reaction products analyzed by reverse-phase ion pair microcapillary electrochromatography. *Electrophoresis* **23**, 3760-3768 (2002).
90. Galloway, M. et al. Contact conductivity detection in poly(methyl methacrylate)-based microfluidic devices for analysis of mono- and polyanionic molecules. *Anal Chem* **74**, 2407-2415 (2002).
91. Shiddiky, M.J., Park, D.S. & Shim, Y.B. Detection of polymerase chain reaction fragments using a conducting polymer-modified screen-printed electrode in a microfluidic device. *Electrophoresis* (2005).
92. Foote, R.S., Khandurina, J., Jacobson, S.C. & Ramsey, J.M. Preconcentration of proteins on microfluidic devices using porous silica membranes. *Anal Chem* **77**, 57-63 (2005).
93. Gulliksen, A. et al. Real-time nucleic acid sequence-based amplification in nanoliter volumes. *Anal Chem* **76**, 9-14 (2004).

94. Gulliksen, A. et al. Parallel nanoliter detection of cancer markers using polymer microchips. *Lab Chip* **5**, 416-420 (2005).
95. Gao, J., Yin, X.F. & Fang, Z.L. Integration of single cell injection, cell lysis, separation and detection of intracellular constituents on a microfluidic chip. *Lab Chip* **4**, 47-52 (2004).
96. Kleparnik, K. & Horky, M. Detection of DNA fragmentation in a single apoptotic cardiomyocyte by electrophoresis on a microfluidic device. *Electrophoresis* **24**, 3778-3783 (2003).
97. Ling, Y.Y., Yin, X.F. & Fang, Z.L. Simultaneous determination of glutathione and reactive oxygen species in individual cells by microchip electrophoresis. *Electrophoresis* (2005).
98. Munce, N.R., Li, J., Herman, P.R. & Lilge, L. Microfabricated system for parallel single-cell capillary electrophoresis. *Anal Chem* **76**, 4983-4989 (2004).
99. Sun, Y., Yin, X.F., Ling, Y.Y. & Fang, Z.L. Determination of reactive oxygen species in single human erythrocytes using microfluidic chip electrophoresis. *Anal Bioanal Chem* **382**, 1472-1476 (2005).
100. Xia, F., Jin, W., Yin, X. & Fang, Z. Single-cell analysis by electrochemical detection with a microfluidic device. *J Chromatogr A* **1063**, 227-233 (2005).
101. Roper, M.G., Shackman, J.G., Dahlgren, G.M. & Kennedy, R.T. Microfluidic chip for continuous monitoring of hormone secretion from live cells using an electrophoresis-based immunoassay. *Anal Chem* **75**, 4711-4717 (2003).
102. Cheng, S.B. et al. Development of a multichannel microfluidic analysis system employing affinity capillary electrophoresis for immunoassay. *Anal Chem* **73**, 1472-1479 (2001).
103. Lion, N. et al. Microfluidic systems in proteomics. *Electrophoresis* **24**, 3533-3562 (2003).
104. Hellmich, W., Pelargus, C., Leffhalm, K., Ros, A. & Anselmetti, D. Single cell manipulation, analytics, and label-free protein detection in microfluidic devices for systems nanobiology. *Electrophoresis* **26**, 3689-3696 (2005).
105. Li, J. et al. Application of microfluidic devices to proteomics research: identification of trace-level protein digests and affinity capture of target peptides. *Mol Cell Proteomics* **1**, 157-168 (2002).
106. Kartalov, E.P. et al. High-throughput multi-antigen microfluidic fluorescence immunoassays. *Biotechniques* (2006).
107. Ramsey, J.D., Jacobson, S.C., Culbertson, C.T. & Ramsey, J.M. High-efficiency, two-dimensional separations of protein digests on microfluidic devices. *Anal Chem* **75**, 3758-3764 (2003).
108. Rocklin, R.D., Ramsey, R.S. & Ramsey, J.M. A microfabricated fluidic device for performing two-dimensional liquid-phase separations. *Anal Chem* **72**, 5244-5249 (2000).
109. Hansen, C. & Quake, S.R. Microfluidics in structural biology: smaller, faster, better. *Curr Opin Struct Biol* **13**, 538-544 (2003).
110. Tran, L. et al. Agonist-induced calcium response in single human platelets assayed in a microfluidic device. *Anal Biochem* **341**, 361-368 (2005).

111. Yang, M., Li, C.W. & Yang, J. Cell docking and on-chip monitoring of cellular reactions with a controlled concentration gradient on a microfluidic device. *Anal Chem* **74**, 3991-4001 (2002).
112. Li, X. & Li, P.C. Microfluidic selection and retention of a single cardiac myocyte, on-chip dye loading, cell contraction by chemical stimulation, and quantitative fluorescent analysis of intracellular calcium. *Anal Chem* **77**, 4315-4322 (2005).
113. Buhlmann, C., Preckel, T., Chan, S., Luedke, G. & Valer, M. A new tool for routine testing of cellular protein expression: integration of cell staining and analysis of protein expression on a microfluidic chip-based system. *J Biomol Tech* **14**, 119-127 (2003).
114. Li, P.C., de Camprieu, L., Cai, J. & Sangar, M. Transport, retention and fluorescent measurement of single biological cells studied in microfluidic chips. *Lab Chip* **4**, 174-180 (2004).