OSAKA UNIVERSITY

DOCTORAL THESIS

Development of a new high-precision quantification methodology for gene expression analysis and its application in iPS cells

---

Author: Yulia Panina
Supervisor: Professor Toshio Yanagida

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in the

Laboratory for Cell Dynamics Observation
Graduate School of Frontier Biosciences / RIKEN BDR

March, 2019

"It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once. An Average is but a solitary fact, whereas if a single other fact be added to it, an entire Normal Scheme, which nearly corresponds to the observed one, starts potentially into existence. Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalised, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of man."

— Sir Francis Galton
Natural Inheritance (1889), 62-3.

"Whenever you can, count."

— Sir Francis Galton
Quoted in James R. Newman, Commentary on Sir Francis Galton (1956), 1169.

SUMMARY

Quantification of gene expression on mRNA level is one of the most important tasks of modern biology. High precision of such quantification is of utmost importance for drawing correct conclusions about cellular processes. Real-time quantitative polymerase chain reaction (RT-qPCR) is currently considered the most precise and most sensitive method of quantifying mRNA. However, the standard experimental procedure in RT-qPCR experiment requires the use of reference genes for normalization. The behaviour of popular reference genes during long-term biological processes, such as iPS reprogramming, development or aging, has never been investigated. In the initial part of my work, I investigate the behaviour of 12 commonly used housekeeping genes for their suitability in RT-qPCR experiments during a representative long-term process, iPS reprogramming, and find that these genes are unsuitable for normalization procedures due to their fluctuation, making standard RT-qPCR inapplicable to iPS reprogramming. Second, I proceed to develop a new methodology for RT-qPCR experimentation that does not require the use of reference genes. Importantly, my methodology increases the precision of obtained measurements while reducing experiment-associated labor and cost. Third, I go on to apply this new methodology to the investigation of the behaviour of 70 housekeeping genes during the iPS reprogramming, demonstrating high potential of the new methodology for high-throughput use. The results obtained in the course of the analysis reveal previously unknown patterns of gene dynamics during iPS reprogramming. I found a collective pattern in the rise of most ribosomal genes' expression, with the exception of small ribosomal subunits Rps18 and Rps9, during the reprogramming process. Furthermore, I found that cell systems associated with growth inhibition, such as apoptosis-associated genes, ubiquitin system genes, or tumor suppressor genes, are collectively down-regulated. Moreover, the analysis showed that there exists a time-dependent pattern in gene expression dynamics of chosen genes, and hints at the existence of an unknown event early in the reprogramming process. These results showcase successful application of my newly developed methodology for gene expression analysis in long-term biological processes, and its notable precision in detection of gene expression changes.

ACKNOWLEDGEMENTS

A difficult, thorny and winding trail has finally led me to the publication of this doctoral thesis. As any biology PhD student in my time knows, the trip to your doctoral degree is filled with sorrows, regrets, dark thoughts and, at times, despair. This trip would not have been possible without people who supported me throughout my journey.

First and foremost, I need to thank my supervisors, Professors Toshio Yanagida and Tomonobu Watanabe, who have been fully supportive and did everything to ensure that my future and my career become possible.

I want to thank my previous supervisor from my Masters times in Kyoto University, Professor Shige H. Yoshimura, who has come to my rescue at the darkest of times. Indeed, if it was was not for him, I would not have been here today.

I must also thank Professor James Hejna at Kyoto University, who has been the most patient of men, reading my emails and keeping in touch with me when I needed it most, and who supported me in every action I take.

I thank Dr. Jamie Gilmore, who was a postdoctoral fellow at the time, for noticing me and caring about it.

I thank my friends, Dr. Aiko Yoshida, and Mars Toktonaliev, for providing me with advice filled with humour, when I needed it so much.

I thank Professor Kunio Takeyasu for being a representation of the light in the end of the tunnel for me, for he has proven that people matter the most, above all things, and that no matter how gloomy your present may seem to you, you might still get rescued.

Finally, I have to thank my father and my mother for the unconditional love they gave me, as they did everything they could to make the moment of my graduation possible.

Yulia Panina
March 2019

Table of Contents

List of abbreviations

ATP	**Adenosine Triphosphate**
Cq	**Quantification cycle**
Ct	**Cycle threshold**
DNA	**Deoxyribonucleic acid**
E	**Efficiency (of the PCR reaction)**
EGF	**Epidermal growth factor**
ER	**Endoplasmic reticulum**
ESC	**Embryonic stem cell**
FDM	**First derivative maximum**
FGF	**Fibroblast growth factor**
iPS	**Induced pluripotent stem**
LDR	**Linear dynamic range**
MEF	**Mouse embryonic fibroblasts**
MIQE	**Minimum information for publication of qPCR experiments**
NCBI	**National Center for Biotechnology Information**
PBS	**Phosphate buffer saline**
PCR	**Polymerase chain reaction**
PFA	**Paraformaldehyde**
RFU	**Relative Fluorescence Unit**
RNA	**Ribonucleic acid**
RNA-seq	**RNA sequencing**
RT-qPCR	**Real-time polymerase chain reaction**
SD	**Standard deviation**
SDM	**Second derivative maximum**
TCA	**Tricarboxylic acid**

General Introduction

0.1 Background

Life Sciences are rapidly moving into the new era of quantification of biological processes. Life scientists increasingly feel the pressuring need for precise analytical instruments for quantitative data analysis. Exact sciences, such as physics, astronomy or optics, have developed such tools for their own use long ago, and it is no coincidence that these sciences are defined by Oxford English dictionary as "those which admit of absolute precision in their results".

One of the most important fields in need of precise quantification is the the field of gene expression measurements. Gene expression is the conversion of genomic information stored in the DNA into functional RNA species, the primary of which is messenger RNA (mRNA). Recent discoveries and research have brought to light many other functional RNA species, such as microRNA (miRNA), non-coding RNA (ncRNA), small nuclear RNA (snRNA) and others. Due to this rapid development of the field, the demand for fast and precise RNA profiling is increasing.

Quantification of small amounts of nucleic acids, and RNA in particular, first became possible in 1980s with the invention of polymerase chain reaction (PCR) by Kary Mullis (Saiki et al. 1985). Earlier methods of determining the quantity of RNA or DNA were based on end-point detection. The most popular of such methods for RNA quantification, Northern Blot, was designed in 1977 and named after the DNA Southern blot invented by Sir Edwin Southern (Alwine et al. 1977). Northern blotting relies on visualization of RNA using denaturing gel electrophoresis and blotting. RNA levels are quantified directly based on the signal that the bands produce and are compared between samples on a single membrane. The disadvantages of Northern Blotting, however, are the low sensitivity (and thus

requirement for large amounts of RNA), sample degradation, and low accuracy (Lee et al. 2005; Maderazo et al. 2003; Trayhurn 1996; Zhao et al. 1996). The discovery of PCR reaction allowed amplification of small quantities of RNA, making quantification of much less abundant RNA species possible. However, there was a major drawback in the end-point detection methods as they did not account for the reaction amplification efficiency, assuming it to be same for all samples. Since PCR amplification efficiency may largely differ between samples, and the nature of the reaction itself is exponential, the comparison of the end product may give largely misleading results. For instance, if we make a mistake of only 5% in estimating reaction efficiency and assuming it to be 100% instead of actual 95%, the difference in amplification of only one molecule in 20 cycles of PCR will rise to 400,000-fold.

To solve this problem, RT-qPCR technique was invented in 1995 (Riedy et al. 1995). This technique monitors PCR reaction in real-time, assessing the target nucleic acid quantity after each PCR cycle (or, rather, assessing the quantity of the target nucleic acid by the associated signal). The advantage of this method compared to end-point detection is clear: assessment of the quantity of the target after each cycle can provide much more information about the reaction and hence, about the starting quantity or the ratio of two genes under comparison at a given cycle. The disadvantage of this method was the need for excessive manual labor which made the analysis of large number of genes, and in particular whole transcriptomes, problematic.

In the following years microarrays and, lately, RNA-seq were introduced for this purpose. However, their lower accuracy, and, in case of RNA-seq, the reliance on many steps of software-based analysis (which makes the process of quantification a black box for the researcher) (Fig. 1), are still a hindrance for these techniques acceptance as precise tools for gene expression quantification, and they still require the confirmation by RT-qPCR which remains the gold standard (Canales et al. 2006; Bustin et al. 2009).

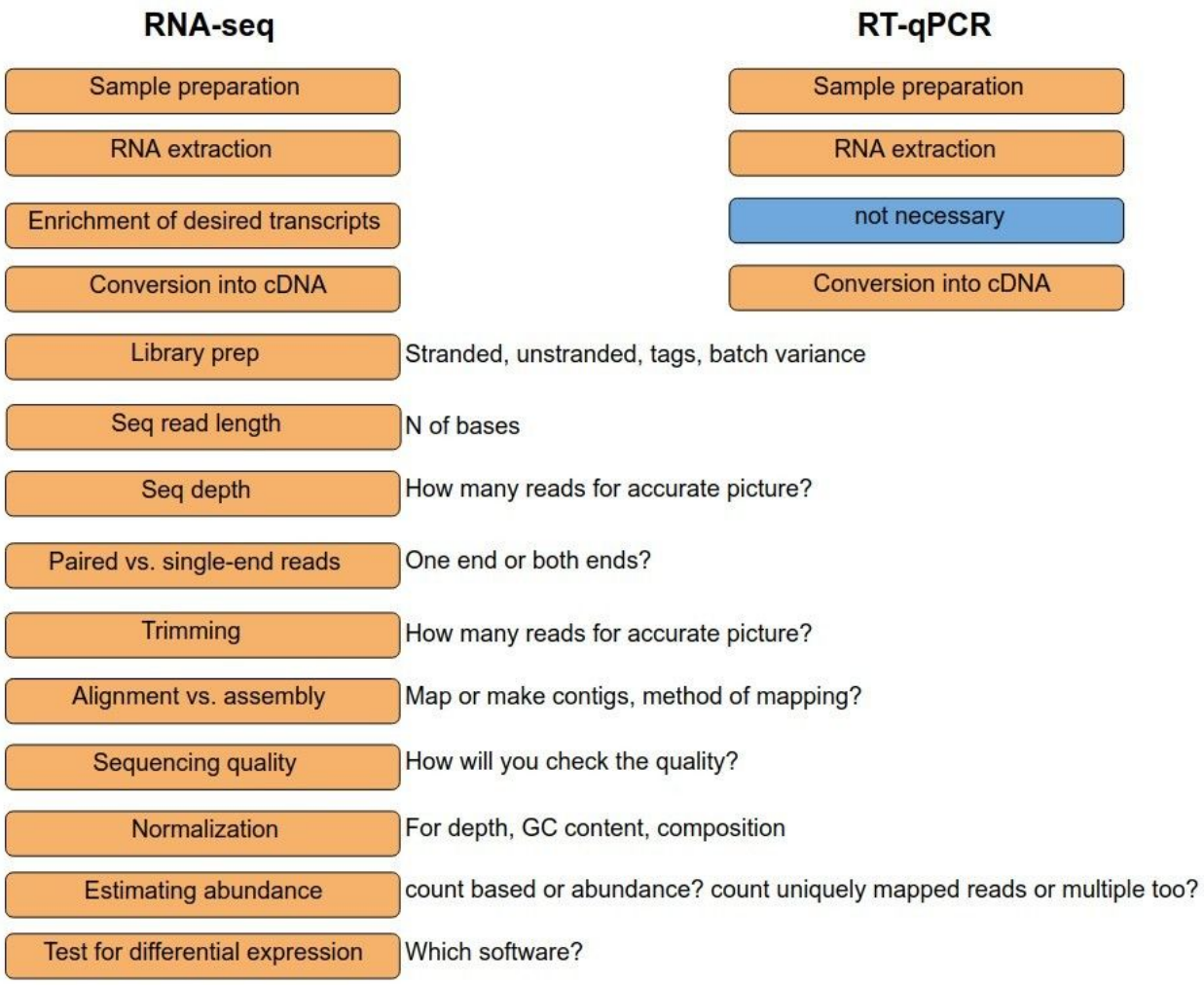


Figure 1. A comparison of high-throughput technology (RNA-seq) with RT-qPCR. The number of steps that may introduce bias or variance *prior to data generation (i.e. before the actual experiment)* differs significantly for RNA-seq and RT-qPCR. In case of RNA-seq at least 10 additional steps are required compared to RT-qPCR. Thus, precision of this technique is considered smaller, and the technique itself is put into the "discovery tools" category rather than "validation tools" category.

Typically, technologies that allow measurements on large number of genes (such as microarrays and RNA sequencing techniques) are considered to be discovery tools and are characterized by lesser reliability and precision. While they allow to screen thousands of samples and transcripts, their drawbacks include high cost, the need for complicated and often subjective computer data analysis

procedures, and technical limitations such as unoptimized protocols, non-specific hybridization problems, data similarity problems stemming from inherent gene sequence similarities and paralogue existence, and high data variance (Jazayeri et al. 2012; Martin et al. 2016; Ozsolak and Milos 2011; Han et al. 2015). On the other hand, increased precision of measurements requires multiple replicas and laborious normalization techniques and are associated with greatly decreased speed of experimentation. The outline of the existing methods and their suitability for particular purpose is shown in Fig. 2.

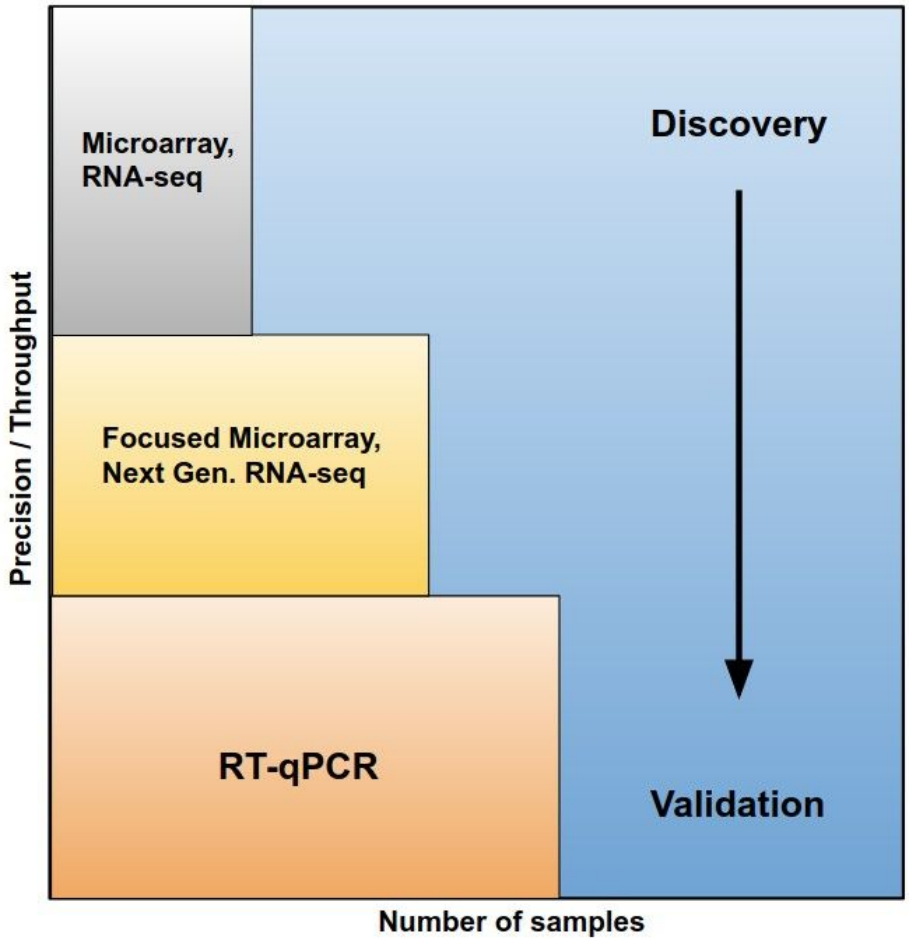


Figure 2. A representation of modern methods of gene expression measurements according to their precision and throughput abilities. High-precision methods, such as RT-qPCR, are used as the validation tools, while less precise methods, such as microarrays and RNA-seq, are used as discovery tools.

0.2 Relevance of this work in its field

Real-time qPCR is currently considered the most sensitive and precise method of gene expression quantification. Both absolute and relative quantification are possible in RT-qPCR. Absolute quantification gives the absolute copy number of RNA fragments per volume and requires construction of so-called "standard curves" (the word standard in this context means "a thing used as a measure or norm in comparative evaluations"). This method is extremely laborious, requiring the amplification and gel-purification of desired standard fragment, subsequent measurement of the concentration, serial dilutions, separate RT-qPCR runs, construction of standard curves and calculations including molecular weight. In addition, the disadvantages include (1) the need to run the standard curve in all subsequent experiments to preserve all experimental conditions, (2) the influence of spectrophotometry or other means of concentration measurement on the resulting concentration values, and (3) most importantly, the assumption that the gel-purified, "clean" fragment will amplify with the same efficiency as the "non-clean" sample (also containing transcribed total RNA). Considering all these problems, the vast majority of researchers use another method, so-called relative quantification.

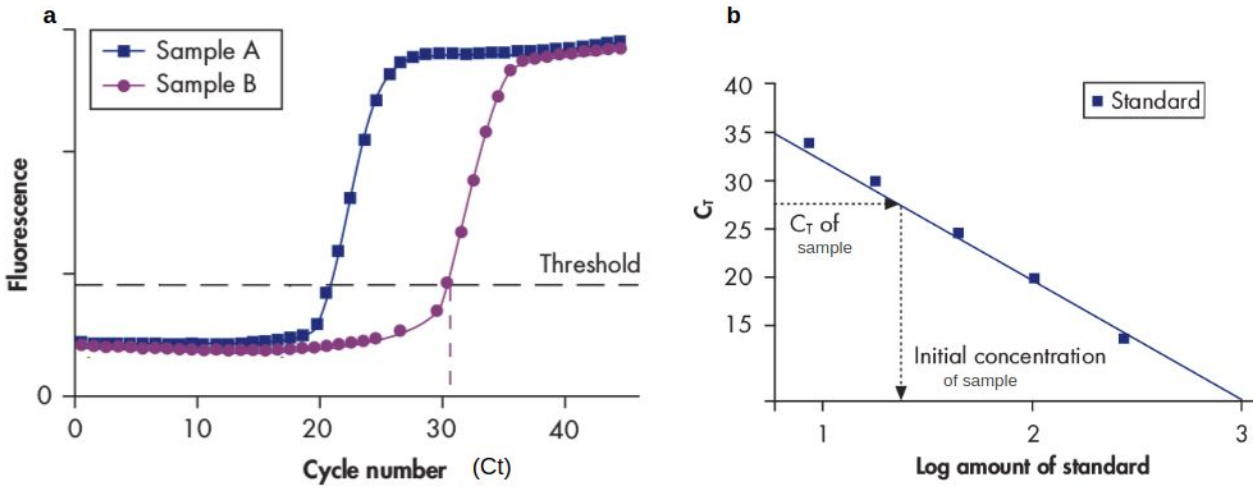


Figure 3. Schematic representation of RT-qPCR absolute quantification method. (a) The RT-qPCR machine produces amplification curves of sample A and sample B based on fluorescence readings that are depicted as squares and circles in the figure. The steep rise in the fluorescence begins around cycle 20 for

sample A, and around cycle 30 for sample B. This means that it takes more PCR cycles to amplify sample B, and thus, its starting concentration must be lower than for sample A. To compare the starting concentrations, the RT-qPCR machine automatically draws a so-called "Threshold" by algorithms that differ depending on the machine maker. The points where the threshold crosses the amplification plot is called "Ct", or Cycle Threshold, and is, in essence, an artificial fractional cycle number created for the purpose of comparison of two samples. **(b)** In order to obtain the absolute copy number of fragments, a standard curve must be constructed as depicted in this figure. The amplified, gel-purified, completely "clean" fragment is measured for exact concentration by spectrophotometry or other methods, and then diluted in at least 6 steps (blue squares). The exact concentration of each diluted sample is thus known. The samples are run in RT-qPCR machine, and the log concentration of each diluted sample is plotted against the resulting Ct value from the machine (blue squares). Thus, we obtain the correlation plot between the amount of the fragment and the Ct value. After that, it is possible to amplify "non-clean" fragments from genomic DNA and total RNA etc., and compare their Ct values with the standard. The efficiency of amplification here is assumed identical for "clean" and "non-clean" samples.

Since absolute quantification is laborious, the standard approach recommended in "The Bible of RT-qPCR", namely The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines, is so-called "relative quantification". Relative quantification is performed with the use of a reference gene or a set of reference genes, and involves normalization of obtained Ct (cycle threshold) values of target genes to the Ct values of reference genes. This method is most frequently used for the analysis of gene expression change after a treatment. The reference genes are used because they are assumed to be constant (non-varied) in all experimental conditions, including the treatment. The expression of target gene(s) is then normalized to the expression of reference genes, making comparison between treated and untreated samples possible.

However, the stability of housekeeping genes used as reference in RT-qPCR has been questioned by several researchers in various organisms, and it is generally recommended to investigate the stability of the reference genes prior to conducting

any experiments with them in the desired system. In particular, the housekeeping genes' stability has never been investigated in long-term dynamic processes that take weeks or months, such as iPS reprogramming, aging, in vivo differentiation etc. Thus, any researchers attempting RT-qPCR analysis in these settings would be required to first conduct several experiments to choose an appropriate reference gene, if such exists at all.

MIQE guidelines recommend normalization to reference genes as the default method of RT-qPCR and presents a list of requirements for conducting experiments using this method (Bustin et al. 2009).

Among these requirements are:

- *Construction of calibration curves with at least three replicates*
- *Measuring slope and y intercept for these curves*
- *Reporting standard deviation (SD) for the replicates of these curves*
- *Reporting r^2 for the curves*
- *Calculating efficiency of amplification from the curves*
- *Reporting Linear Dynamic Range from the curves*
- Reporting method of Ct determination (since it is automated in the machine)
- *Justifying the number and choice of reference genes*
- *Describing the normalization method for reference genes*
- Running at least three replicates for the main experiment and reporting SD

(the requirements concerning reference genes are typed in italics, i.e. this work is needed when performing any RT-qPCR experiment with reference genes)

All of these measures are designed to improve the precision of RT-qPCR measurements in cases when reference genes are used. For example, the calibration curves are run to determine the PCR efficiency because it is the most important parameter that affects the calculation of the ratio of gene expression change. Since this RT-qPCR method is based on Ct values, MIQE requires to run at least three replicates in all cases, including when running calibration curves, due to the fact that a single run gives only one Ct value and is unreliable statistically (Bustin et al. 2009). In addition, since Ct values are automatically defined by the RT-qPCR machine and differ between different machines, MIQE requires to report their determination method. Moreover, since the whole scheme depends on stability of

the reference genes, their choice must be justified, their stability should be confirmed for all experimental conditions, and the normalization procedure should be clearly described. Of course, all experiments involving reference genes and target genes should include at least three technical replicates and, ideally, three inter-assay replicates (same experiment run on different plates at different times of the day). I want to emphasize that all of the above is defined as **essential information that must be published with every RT-qPCR experiment**, according to MIQE guidelines. Needless to say, this constitutes an enormous amount of experimental work. *Eliminating the need to use reference genes without losing precision*, thus, will constitute a vast improvement in RT-qPCR experimental procedure because it will result in economy of hands-on time, reagents, and data analysis.

Several methods have been developed in the past decades to improve RT-qPCR precision, such as FPK-PCR (Lievens et al. 2012), LinRegPCR (Ramakers et al. 2003), Cy0 (Guescini et al. 2013) and others. However, according to a recent analysis, these alternative methods rely on different ways of approximating a single amplification curve and have never yielded acceptable accuracy (Ruijter et al. 2013). Thus, running multiple replicas remains the only way to ensure sufficient precision of RT-qPCR results, and this is associated with greater labor and cost of experimentation. In addition, the classical formula of normalization recommended in MIQE guidelines requires that the efficiency of PCR reaction be measured for both reference gene and target gene, and if the efficiency is not equal, the change of the reference gene is required, which involves considerable time loss and increased labor. Moreover, according to the guidelines, the reference genes must be tested for fluctuation and deemed suitable for experimentation for each particular biological process, which is also a time-consuming and laborious task. All these issues prevent RT-qPCR from becoming a trusted and fully established technique in the biological community.

This work aims at solving these problems by, first, investigating the behavior of commonly used reference genes and then, developing a new, highly precise methodology that requires much less labor and reduces overall hands-on time and cost of experimentation.

0.3 Significance and purpose of this work

My major vision is to establish a new conceptual direction in the field of nucleic acid measurements. The existing methods of experimentation and data analysis in the field of nucleic acid measurements do not possess satisfactory precision unless multiple replicas are run for each experiment, and unless a carefully crafted normalization strategy is applied, as described in the previous section. My goal here is to address this issue, and to move forward the field of nucleic acid quantification. As a first step towards this goal, I have developed a new approach to qPCR data analysis, that is based on solid mathematical principles, such as the Theory of Measurement (described in the next section), has doubled precision compared to the existing methods, and is suitable for high-throughput analysis. I believe that current biology needs its new Francis Galton, and I would be happy if my contribution could move the field in that direction, the direction of exact science.

Goals:

- Investigate the suitability of commonly used gene expression quantification tools for high-precision quantification in long-term processes
- If necessary, develop a new high-precision tool for gene expression quantification
- Apply the said tool in a system that represents a long-term dynamic process

0.4 Originality

The novelty of my approach is in the application of the principles of the Theory of Measurements (Hand 1996) to the solution of the precision problem in gene expression measurements. The Theory of Measurements (alternatively, Measurements Theory) is a branch of applied mathematics that is commonly used in physics, engineering and other exact sciences. It ensures the best performance of data gathering techniques and research techniques, and is considered prerequisite for all measurements conducted in such sciences. The Measurement Theory regards all measurements as observations that are not identical to the attributes

being measured. Rather, all measurements are considered approximations of the said attributes, and are represented by assigned numbers. Thus, when one wants to draw conclusions about a property of an object, he (she) must take into account the nature of the correspondence between the measurements and the property being measured (Krantz et al. 1971). This allows one to use reliable, proven statistical apparatus, including advanced statistical methods, to model complex natural processes.

Since the quantitative research in biology is only gaining momentum recently, the Measurement Theory has not yet been integrated into biological methods. The measurements of gene expression, in particular, are not standardized because there is no international system of units, as is the case for physics which has standard units such as length, weight or speed, the quantities reported in biological studies are nearly always relative and cannot be translated into different experimental systems, and the statistical approaches do not rely on a set of common, firmly defined principles. My application of the classical principles used in exact sciences, namely the Measurement Theory, to the measurements of nucleic acids, is the first step in the direction of standardization of biological procedures.

Thus, the novelty that I introduce is both conceptual and practical, and is based on integration of proven statistical methods from exact sciences into the existing methods in biology. This approach allowed me to increase the precision and reliability of measuring gene expression, greatly reducing labor and cost of experimentation. Moreover, my application of the Measurements Theory to the tracking of 70 genes' expression dynamics during the iPS reprogramming process enabled me to reveal important information about the genes' behavior.

Chapter 1

Investigation of housekeeping gene dynamics during the iPS reprogramming process

1.1 Introduction

As the first step, I investigated whether the conventional method of RT-qPCR analysis is suitable for my research goal, namely, high-precision, medium-throughput gene expression quantification during the iPS reprogramming process.

Induced pluripotent stem cell (iPSC) reprogramming is an artificial, human-induced gradual change in gene expression which allows to turn a differentiated somatic cell into a pluripotent cell. It usually takes a relatively long time (~1 month) (Takahashi and Yamanaka 2006), involves several major transcriptional circuits (Papp and Plath 2011), and is accompanied by many drastic changes in fundamental cell properties and behaviour, such as energy production (Panopoulos et al. 2012), changes in cell cycle progression patterns (Smith et al. 2010), cytoskeletal organization (Li et al. 2010) and others. Currently iPS reprogramming is divided into three major stages, initiation, maturation and stabilization (David and Polo 2014). It is known that the alterations in cell chemistry during these stages include most basic, "housekeeping" functions such as cell metabolism (Panopoulos et al. 2012), speed of the cell cycle (Papp and Plath 2011) and lipid profile (Boraas et al. 2016; Mammoto and Ingber 2009).

Being a highly dynamic, time-dependent process, iPS reprogramming is akin to natural biological processes that humans are interested in investigating, such as cell differentiation, embryo and organ development and aging. As such, iPS reprogramming constitutes an ideal *in vitro* model system to design, build and validate a workflow for quantification of vast, gradual chemical changes in live cells.

RT-qPCR requires a normalization strategy to ensure the reliability of the data (Jacob et al. 2013; De Spiegelaere et al. 2015). One common strategy is to rely on

the comparison of the target gene with an endogenous control (reference gene) in the same sample. When normalizing qPCR data to a reference gene, it is of crucial importance to make sure that the gene of choice is stably expressed throughout all experimental conditions. At present, so-called housekeeping genes are universally used as a reference (Piazza et al. 2017). For example, housekeeping genes such as actin, ubiquitin or ribosomal genes are thought to be universally required for basic cellular functions and to be constitutively and stably expressed in varying physiological and experimental conditions. However, recent works have uncovered that housekeeping genes' expression levels may vary depending on the gene, cell type and experimental conditions. For example, one of the most frequently used housekeeping genes, GAPDH (glyceraldehyde-3-phosphate dehydrogenase) has been found to be unstable depending on the type of tissue (Sullivan-Gunn et al. 2011), metabolic process (Gong et al. 2016) or under certain experimental conditions (Mahoney et al. 2004). Thus, confirming the stability of the normalizing gene of choice in cells under study is a prerequisite for a correct analysis of gene expression of any target gene. The need to confirm stability of expression is even greater in case of iPS reprogramming, as housekeeping genes could be affected by dramatic changes in chemical metabolism of the cells.

The RT-qPCR analysis of common housekeeping genes' stability over the time course of iPS reprogramming has never been performed. Thus, the first step of my analysis was to monitor changes in the expression of common housekeeping genes during iPS reprogramming and determine their suitability for further experimentation.

1.2 Materials and Methods

Cell culture and iPS reprogramming

iPS reprogramming was carried out in a reprogrammable cell system previously described by Hikichi and colleagues (Hikichi et al. 2013). The system consists of a mouse neural progenitor cell line designated N31 which possesses three key characteristics of neural progenitors: (1) fibroblast growth factor and epidermal growth factor-dependent growth, (2) neural stem cell markers' expression and (3) the ability to differentiate into neural lineages (described in detail in Han et.al, 2012 and Hikichi et.al., 2013). To bypass the need for mRNA or virus introduction into the cells, a doxycycline-inducible cassette with four Yamanaka factors, Oct4, Sox2, Klf4 and c-Myc, was permanently integrated into the cell genome. Doxycycline addition results in the activation of the four factors and initiates reprogramming. Cells were seeded and kept on plastic gelatin-coated dishes in RHB neural stem cell media (#Y40000, Clontech Takara, Japan) supplemented with Ndiff (# Y40100, Clontech Takara, Japan) and 10ng/ml FGF and 10 ng/ml EGF until they fully attached and spread. To initiate reprogramming, the medium was changed to Essential 8 iPS reprogramming medium (A1517001, Thermo Fisher, Japan) and 1µg/ml doxycycline was added to the dish. From that point on, the media were changed every day to avoid pH fluctuations. The reprogramming was carried out until day 20, and the cell material samples (whole cell populations) were collected at 8 time points (on days 0, 1, 3, 5, 7, 10, 15, 20). One round of reprogramming thus yielded 8 cell pellets. The experiment (one full round of reprogramming) was repeated three times to obtain three biological replicates for each time point. To confirm the success of reprogramming, cells were analysed on the day 20 for markers of pluripotency: (1) cell morphology, (2) alkaline phosphatase expression, (3) pluripotent genes Nanog and Oct4 expression (Supplementary Fig. 2 and Fig. 14).

MEFs, partial iPS and fully reprogrammed iPS derived from EOS3F-24 line for corroborative experiment were a gift of professor A. Hotta and were maintained as previously described (Hotta et al. 2009).

Candidate genes and primer design

Table 1 contains information about the 12 commonly used housekeeping genes chosen for this study, and the selected primer pairs used for amplification. All primers were designed according to MIQE guidelines with the aid of Primer-BLAST software (NCBI). Primers were designed to be specific preferentially for the longest isoform (transcript variant), allowing complementarity to other transcripts of the same gene (e.g. transcript variant 2, transcript variant X1 etc.), within the coding sequence. To select the best primers, the coding region of a gene was divided into portions spanning approximately 200-400 bp, and primers were designed to each portion using NCBI software. After excluding primer pairs that, according to NCBI Blast, could produce unintended target amplicons, resulting primer-pairs were tested by qPCR and the best pair for each gene was selected for the experiment.

RNA isolation and cDNA synthesis

Total RNA was extracted with RNeasy kit (Cat# 74106, Qiagen, Japan) from each biological sample according to the manufacturer's instructions (on-column genomic DNA digestion was performed as per said instructions), and RNA concentration and absorbance ratios ($A_{260/280}$ and $A_{260/230}$) were measured by spectrophotometer Nanodrop 2000 Spectrophotometer (NanoDrop Technologies, Japan). Only the samples with $A_{260/280}$ and $A_{260/230}$ were used for further analysis. 300 ng of RNA from each sample was reverse-transcribed using Omniscript RT Kit (Cat# 205111, Qiagen) in a total volume of 20µl to produce DNA that was subsequently assessed by spectrophotometric analysis and diluted to 100 ng/µl. Then, individual master mixes with each of the DNA-primer combination (e.g. ‘Day 0 - Atp5f1’, ‘Day 0 - B2m’ etc.) were created for 4 technical replicas, and the mixtures were distributed onto the qPCR plate (8 µl per reaction well).

The reprogramming process (day 0 - day 20) was repeated 3 times, thus 3 replicates were obtained for each time point.

Quantitative real-time PCR

qPCR was performed using a CFX96 Connect apparatus (BioRad, Japan). The reactions were carried out in triplicate using intercalating dye SYBR Green-based

PCR super-mix (BioRad), following the manufacturer's instructions. Each reaction was performed in the final volume of 8 μL, primers were used at the concentration of 300 nM. Thermocycler program consisted of an initial hot start cycle at 95°C for 3 min, followed by 32 cycles at 95°C for 10 sec and 59°C for 30 sec. To confirm product specificity, melting curve analysis was performed after each amplification.

Immunostaining

For immunofluorescence analyses cells were grown on glass bottom 30-mm dishes coated with collagen type I (IWAKI #4970-011). On the day of immunostaining cells were briefly washed with PBS, fixed with 4% PFA (Santa Cruz #sc-281692) for 15 min at room temperature and permeabilized with 0.5% Triton in PBS with 10% FBS addition for 30 min. Primary antibodies were applied: Anti-Oct4 (Santa Cruz #sc-5279, 1/250 dilution), Anti-Nanog (Abcam #ab80892, 1/250 dilution), in PBS with 10% FBS addition, for 1 hour in room temperature. After washing cells were incubated with secondary antibodies: Anti-mouse Alexa Fluor® 594 (Cell Signaling #8890, 1/500 dilution) and Anti-rabbit Alexa Fluor® 488 (Cell Signaling #4412, 1/500 dilution) in PBS with 10% FBS addition for 1 hour in room temperature, then cells were washed 4 times with PBS and 2 mL PBS per dish was added for imaging.

Alkaline phosphatase staining and imaging

For alkaline phosphatase staining cells were briefly washed with PBS, fixed for 5 min with 4% PFA (Santa Cruz #sc-281692) at room temperature and stained with Alkaline phosphatase kit II (Stemgent, #00-0055) according to the manufacturer's protocol. Imaging was carried out on Olympus CKX41 inverted microscope.

Statistical analyses

The assay performance evaluation was carried out as described in MIQE guidelines. Reaction efficiency E was calculated as E =(-1/slope)-1x100 and precision was calculated as the average of all standard deviation values across samples for each gene. Linear dynamic range (LDR) is defined as the highest to the lowest quantifiable copy number established by means of a calibration curve, and covers at least 3 orders of magnitude, as advised. The interval at which the main

experiments were carried out fell into the linear portion of the calibration curve. The linearity was determined by means of correlation coefficients (R^2). Precision refers to intraassay variation and is expressed as standard deviation (SD) of technical replicates, as advised. For BestKeeper (Pfaffl et al. 2004) analysis, Ct values were input directly, and geometric and arithmetic mean as well as standard deviation and coefficient of variance were calculated by the program, according to which genes were subsequently ranked from most stable to least stable. For NormFinder (De Spiegelaere et al. 2015) analysis, Ct values were transformed to linear scale and the normalization factor was calculated as the geometric mean of candidate reference genes included in the dataset. GeNorm software analysis was performed by calculating the expression stability measure as defined in the geNorm paper (Jacob et al. 2013), pairwise variation was determined and genes were ranked according to their positions. RefFinder algorithm was used to produce comprehensive ranking. This algorithm integrates four major programs (geNorm, Normfinder, BestKeeper, and the Delta Ct method) to assign a weight value to an individual gene and calculates the geometric mean of the weights for the overall final ranking. Time-course plot of the gene expression through the reprogramming process were performed using the JMP software (JMP®, Version v11, SAS Institute Inc., Cary, US). Variance analysis between time points were performed using ANOVA test followed by a post-hoc Tukey HSD test at $p < 0.05$.

Results

1.3 Assay performance evaluation

I have chosen twelve commonly used housekeeping genes for the purpose of this initial analysis (Table 1).

Gene symb	Accession No.	Official Full Name (MGI)	Primer Pair (5′-3′)	size (bp)
Actb	NM_007393.5	actin, beta	TCGAGTCGCGTCCACC GGGAGCATCGTCGCCC	157
Atp5f1	NM_009725.4	ATP synthase, mitoch. F0 complex, subunit B1	GTCCAGGGGTATTACAGGCA A TCAGGAATCAGCCCAAGACG	112

B2m	NM_009735.3	beta-2 microglobulin	ACGTAACACAGTTCCACCCG CAGTCTCAGTGGGGGTGAAT	150
Gapdh	NM_001289726.1	glyceraldehyde-3-phosphate dehydrogenase	GCACAGTCAAGGCCGAGAAT GCCTTCTCCATGGTGGTGAA	151
Gusb	NM_010368.2	glucuronidase, beta	AACAACACACTGACCCCTCA ACCACAGATCGATGCAGTCC	140
Hprt	NM_013556.2	hypoxanthine guanine phosphoribosyl transferase	CAGTCCCAGCGTCGTGATTA TGGCCTCCCATCTCCTTCAT	168
Pgk1	NM_008828.3	phosphoglycerate kinase 1	GGGTGGATGCTCTCAGCAAT GTTCCTGGTGCCACATCTCA	160
Ppia	NM_008907.1	peptidylprolyl isomerase A	CCCACCGTGTTCTTCGACAT CCAGTGCTCAGAGCTCGAAA	116
Rps18	NM_011296.2	ribosomal protein S18	AAGCAGACATCGACCTCACC CTAGACCGTTGGCCAGAACC	171
Tbp	NM_013684.3	TATA box binding protein	AGTTGGGCTTCCCAGCTAAG GCTACTGAACTGCTGGTGGG	160
Tfrc	NM_001357298.1	transferrin receptor	AAGAGCTGCTGCAGAAAAGC ACGGTCTGGTTCCTCATAACC	190
Ywhaz	NM_011740.3	tyrosine 3-monooxygenase /	GATTGGAGGAAACCCCGTGT CCTTCTGCACCAGCTCATTT	190

Table 1. Summary of twelve housekeeping genes evaluated in this study. Accession numbers, gene descriptions, primer sequences and product sizes are shown.

To evaluate the performance of the qPCR assay, I generated calibration curves using tenfold serial dilutions and assessed the PCR efficiency denoted *E* (see Materials and Methods), linear dynamic range (LDR) and precision, as described in MIQE guidelines (Bustin et al. 2009). Results are shown in Table 2. The mean amplification efficiency values ranged from 95% (Actb) to 163% (Gusb), corresponding to slopes of -3.45 and -2.39, respectively. Nine genes out of twelve fell within "good" range of PCR efficiency defined as 90%< *E* <110%, while Tfrc, Hprt and Gusb produced 114%, 143% and 163%, respectively. To ensure that this result was due to the gene behavior rather than primer design, I evaluated 4 primer pairs for each of these genes, designed to cover different regions of the genes, as

well as tested these pairs both in the parental cell line before reprogramming (N31 cells) and after the reprogramming (iPS cells). The results for these primer pairs can be found in Supplementary Table 1.

The LDR values were lowest for Hprt and Gusb and were in the range of 7-700 ng of template. Correlation coefficients, on the other hand, varied from 0.991 (Gusb) to 1.0 (Pgk1) and fell within acceptable range for all genes as all of them were >0.99. Precision values ranged from 0.12 (Hprt) to 0.38 (Gusb). Overall these results show good performance of the qPCR assay except for Gusb and Hprt that performed less well in the lower concentration ranges (less than 7 ng).

Gene	E (%)	Slope	LDR (ng)	Precision	R^2
Actb	95	-3.45	0.07-700	0.18	0.999
Atp5f1	104	-3.23	0.07-70	0.31	0.999
B2m	100	-3.31	0.07-70	0.22	0.999
Gapdh	96	-3.42	0.07-70	0.29	0.998
Gusb	163	-2.39	7-700	0.38	0.991
Hprt	143	-2.60	7-700	0.12	0.995
Pgk1	101	-3.29	0.07-700	0.26	1.0
Ppia	100	-3.32	0.07-70	0.15	0.997
Rps18	99	-3.34	0.07-700	0.16	0.998
Tbp	106	-3.18	0.7-700	0.21	0.997
Tfrc	114	-3.02	0.07-700	0.36	0.999
Ywhaz	102	-3.27	0.7-700	0.17	0.999

Table 2. Assay performance characteristics showing PCR efficiency *E*, linear dynamic range (LDR), slope, precision and associated correlation coefficient R^2 (see Materials and Methods).

1.4 Analysis of candidate reference genes' stability in iPS reprogramming process

The stability of candidate reference genes was analyzed according to four statistical methods of assessment, namely, the Delta Ct method, the estimation of the intra- and intergroup variation (NormFinder), the basic descriptive statistics evaluation (BestKeeper), and pairwise comparison (geNorm). The comprehensive ranking of the genes was also evaluated (see Materials and Methods), giving a total of five evaluation methods. The analysis revealed that Atp5f1 was unanimously chosen as the most stably expressed gene by all four algorithms, while Rps18 was designated as the least stable gene. Pgk1 was chosen as the second most stable gene by 4 out of the 5 algorithms. Gapdh was designated as the third most stable gene, except by the Delta Ct method which designated it as the second most stable gene. Thus, the order of stability for the best three genes was summarized as follows: Atp5f1 > Pgk1 > Gapdh. On the other hand, the three least stable genes were Rps18 > Hprt > Tbp / Actb (Table 3).

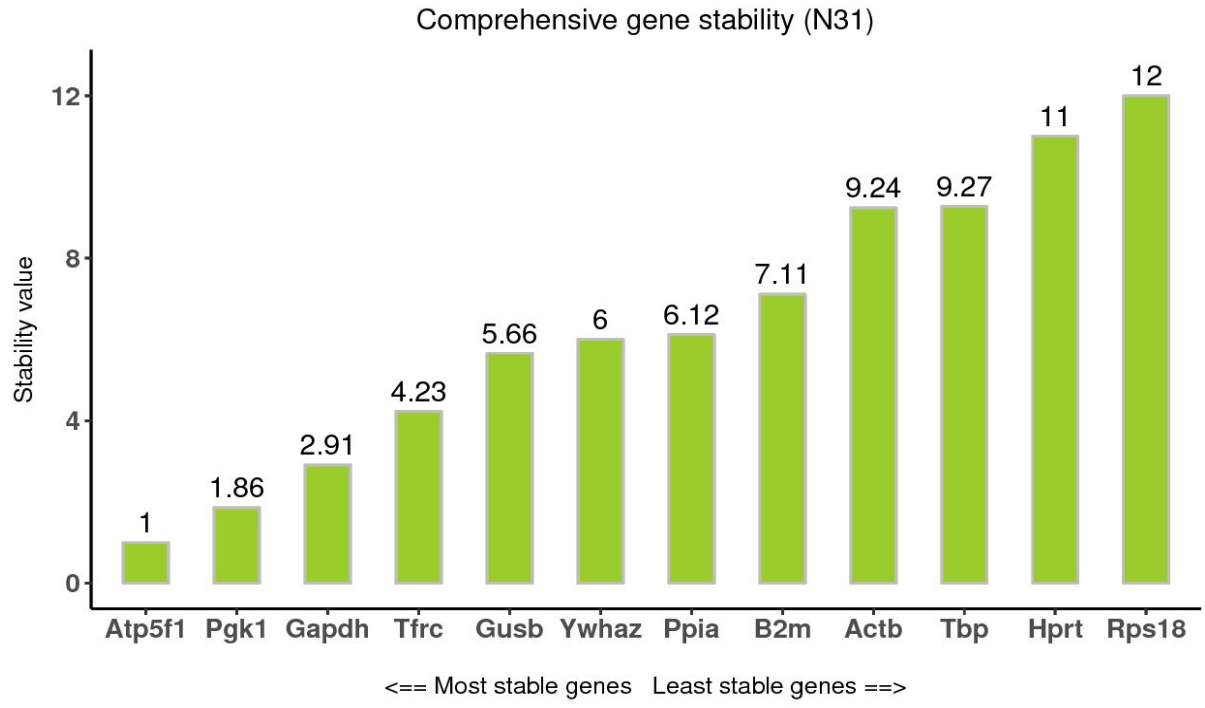
Gene	Comprehensive Ranking		Delta Ct		geNorm		NormFinder		BestKeeper	
	Value	Rank	SD aver.	Rank	M value	Rank	Stability	Rank	SD	Rank
Atp5f1	1.00	1	1.03	1	0.362	1	0.288	1	0.40	1
Pgk1	1.86	2	1.12	3	0.362	1	0.489	2	0.46	2
Gapdh	2.91	3	1.11	2	0.504	2	0.516	4	0.48	3
Tfrc	4.23	4	1.14	4	0.560	3	0.640	5	0.52	4
Ppia	6.12	7	1.18	5	0.580	4	0.742	7	0.67	8
Gusb	5.66	5	1.22	7	0.729	6	0.502	3	0.60	7
Ywhaz	6.00	6	1.21	6	0.643	5	0.716	6	0.55	6
B2m	7.11	8	1.31	8	0.790	7	0.784	8	0.54	5
Actb	9.24	9	1.45	9	0.847	8	1.140	9	0.89	10

Tbp	9.74	10	1.59	10	0.952	9	1.194	10	0.86	9
Hprt	11.00	11	2.07	11	1.154	10	1.767	11	1.35	11
Rps18	12.00	12	2.88	12	1.441	11	2.750	12	2.11	12

Table 3. Ranking of the candidate reference genes according to five different evaluation methods. Atp5f1, Pgk1 and Gapdh were ranked as the most stable candidate reference genes, while Rps18, Hprt, and Tbp / Actb were designated as the least stable ones.

Then, to corroborate these results in a more commonly used cell line, I assessed the stability of the 12 housekeeping genes using mouse embryonic fibroblasts (MEFs) described by Hotta et. al. (Hotta et al. 2009). The stability was assessed using 3 time points for each gene, corresponding to the non-reprogrammed state (MEFs), partially reprogrammed state (partial iPS), and fully reprogrammed state (iPS). In fibroblasts, the five statistical algorithms also selected Atp5f1, Pgk1 and Gapdh as the best reference genes, while B2m, Actb and Hprt showed the lowest stability (Fig. 4). The exact ranking as obtained by the five statistical algorithms can be found in Supplementary Table 2.

a



b

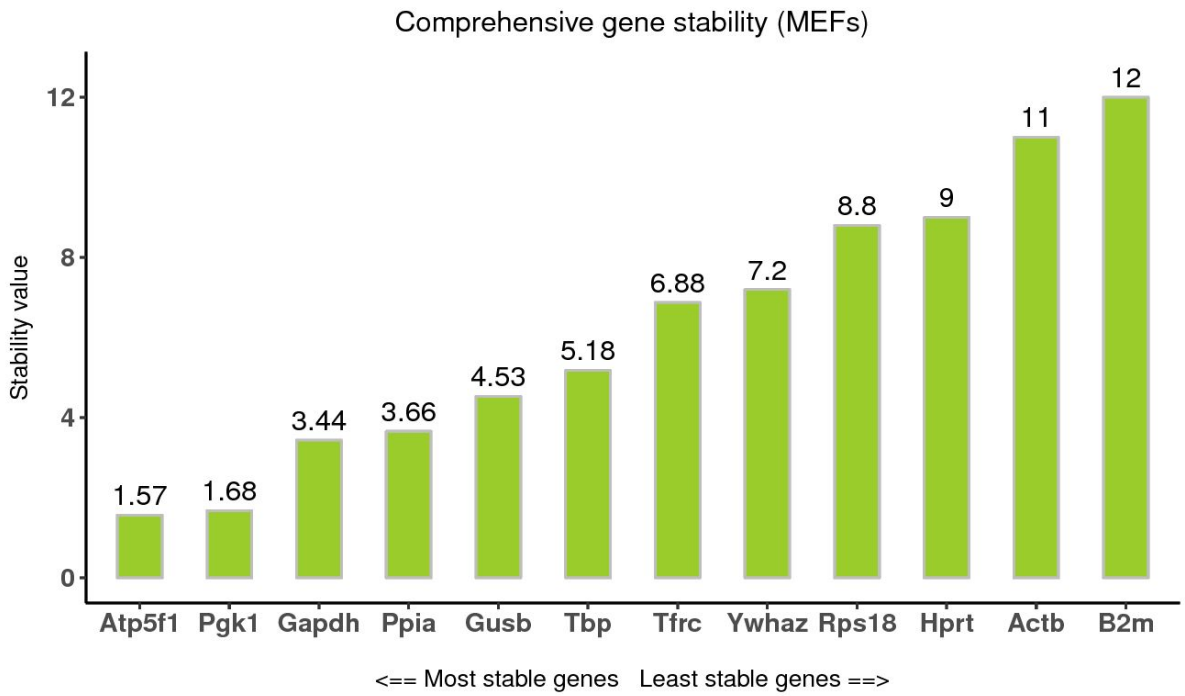


Figure 4. Comprehensive stability ranking of candidate reference genes during iPS reprogramming of neural stem cells (N31) or mouse embryonic fibroblasts (MEFs). The stability is expressed as a comprehensive value for five algorithms: Genorm, Normfinder, Bestkeeper, Delta Ct, and RefFinder. Atp5, Pgk1 and Gapdh are showing the best stability in both cell lines. The lowest stability values differed in two cell lines, the lowest score belonging to Rps18, Hprt and Tbp in N31 cells (a) and to B2m, Actb and Hprt in MEFs (b).

1.5 Expression variability of candidate reference genes during the iPS reprogramming process

To assess the expression variability of chosen candidate reference genes during the reprogramming process, qPCR was performed and the relative Ct values for each gene across 8 time points were obtained throughout the reprogramming process, from day 0 to day 20. Figure 5 shows that the mean Ct values for 12

candidate genes varied from 9.99 to 24.21 cycles. The highest Ct value was observed for Rps18 (9.99) while the lowest value was observed for Hprt (24.21). To provide an initial estimation of the variability for each gene, I calculated standard deviation and coefficient of variation (CV). The least variable gene as expressed by SD value was Atp5f1 (n = 3, SD = 0.52), and the most variable was Rps18 (n = 3, SD = 2.70). Atp5f1 exhibited the lowest CV value, and Actb the highest one. For a more comprehensive analysis, the difference between 25th and 75th percentile was calculated in order to estimate the amplitude fluctuation. According to this analysis, Rps18 showed the highest variability, with an amplitude fluctuation of 3.89, and the least variable genes were Atp5f1 and Pgk1 with the amplitudes 0.58 and 0.61, respectively. Thus, according to the initial analysis, Atp5f1 was identified as the most stable gene while Rps18 was identified as the least stable gene across the whole reprogramming process.

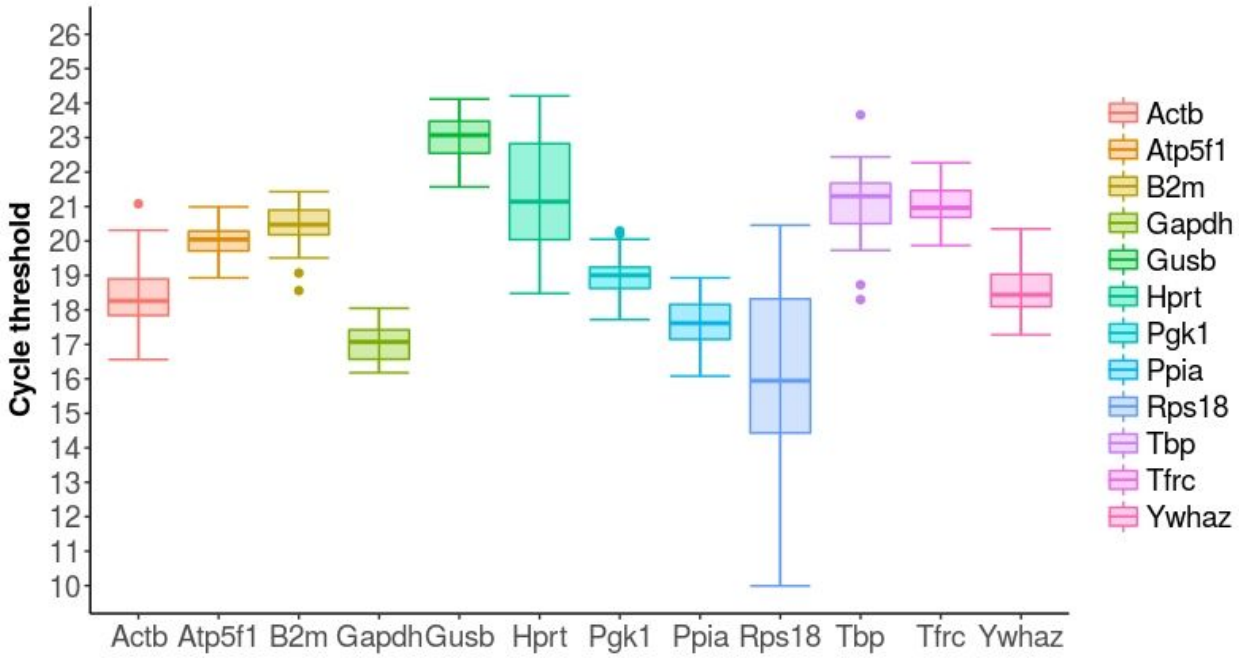


Figure 5. Box-and-whisker plot indicating range of Ct values of candidate reference genes throughout iPS reprogramming. Values of three biological replicates taken as averages of 4 technical replicates are given. The whiskers represent standard deviation of n samples (n=24).

1.6 Time-course expression profiles of candidate reference genes

Figure 6 shows Cycle threshold values (Ct) of candidate reference genes plotted against time. The cycle threshold is inversely proportional to the gene expression of the considered gene. Results showed that Actb, Hprt, and Rps18 displayed the strongest variation over time. According to this analysis, Actb expression decreased during the reprogramming process (R^2 = 0.85), and the analysis of variance was found significant ($p < 0.0001$). Hprt and Rps18 expression decreased in the first week of reprogramming, increased around day 10, then decreased again. R^2 values for Hprt and Rps18 were R^2 = 0.71 and R^2 = 0.20, respectively. At Day 20, the genes Gusb, Rps18, Tbp, and Ywhaz showed the strongest variation.

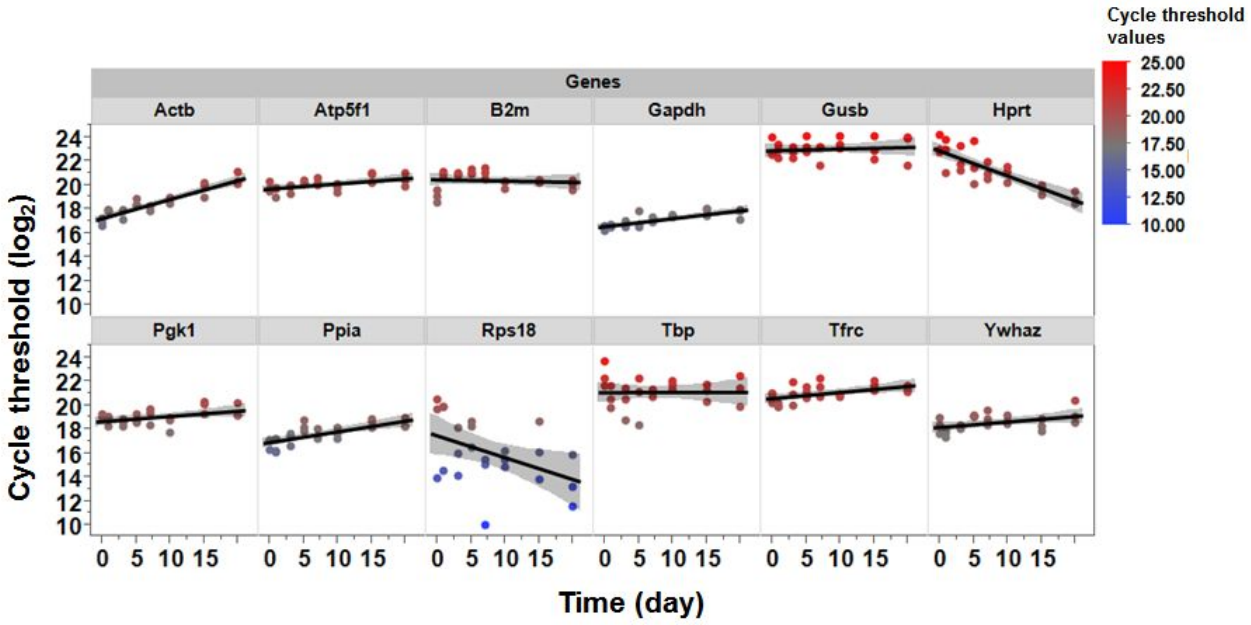


Figure 6. Expression profile of the 12 candidate reference genes throughout the 20 days of the reprogramming process. Measurements were performed in triplicate for each day. For each gene, linear fits were applied (black lines) and the displayed grey areas represent the 95% confidence intervals. For visualization purpose, a color bar representing the log2 values of Cycle threshold was added.

1.7 Discussion

In this chapter I have performed iPS reprogramming in murine cells and measured the expression patterns of the most commonly used housekeeping genes by conventional RT-qPCR method. I have analysed 12 most frequently used reference genes for suitability in RT-qPCR experiments during iPS reprogramming, and found that all genes analysed fluctuated by 1-2 fold throughout the reprogramming process.

Using the expression data obtained by the conventional RT-qPCR method and five statistical algorithms published in available literature, I have identified Atp5f1, Pgk1 and Gapdh as the most stably expressed genes. Notably, these genes were grouped together by the algorithms, and they all belong to the ATP production process. First of all, the Atp5f1 gene, which had the highest stability rank, is a B subunit of the proton channel of mitochondrial F0 complex, and is a part of mitochondrial ATP synthase. ATP synthase is composed of F0 and F1 complexes and is linked by the peripheral stalk, of which B subunits are part (Ko et al. 2000). The function of the subunits in this context is, apart from linking the complexes, to act as a stator to prevent other subunits from rotation in relation to the central rotary element. Atp5f1 is, thus, an essential structural element of the ATP synthase. The second most stable gene as determined by four algorithms was phosphoglycerate kinase 1 (Pgk1). It is an ATP-generating enzyme that catalyzes the reversible conversion of 1,3-diphosphoglycerate and ADP to 3-phosphoglycerate and ATP and is considered an important part of the glycolytic pathway (Li et al. 2016). The third most stable gene was Gapdh, or glyceraldehyde 3-phosphate dehydrogenase, which also belongs to the glycolytic pathway, and catalyses the conversion of glyceraldehyde 3-phosphate to D-glycerate 1,3-bisphosphate. These results can be explained from the point of view of requirement for glycolysis. Recent research has shown that glycolysis is required for iPS reprogramming (Zhang et al. 2012) and that inhibition of glycolysis can impede the reprogramming process (Folmes et al. 2011). The so-called “glycolytic switch” is suggested to play a major part in the pluripotency switch (Teslaa and Teitell 2015), and it would be interesting to follow up on the dynamics of ATP-related genes throughout the iPS reprogramming.

On the other hand, my analysis has revealed that the expression pattern of the actin gene (Actb), often considered as a reliable reference gene in other studies (Ruiz-Villalba et al. 2017), steadily decreased throughout the reprogramming. As a result, the gene was consistently ranked as one of the most unstable genes across different statistical methods. This change in actin mRNA expression may reflect the cytoskeletal remodelling which is normally associated with iPS reprogramming (Boraas et al. 2016) and plays a central role in the cell fate change (Mammoto and Ingber 2009).

Hprt was marked as second least stable gene in the analysis. This result is in agreement with previous work on reference genes in pluripotent stem cells that also marked Hprt unsuitable for use as a reference gene (Murphy and Polak 2002). The authors conducted differentiation of embryonic stem cells and measured housekeeping gene expression change at different time points. Induction of pluripotency can be viewed as a process opposite to differentiation, with pluripotency features gradually emerging instead of disappearing.

The ribosomal gene Rps18 was found to vary greatly and stand out as the most unstable gene among all, being ranked last by all algorithms unanimously. Previous investigations of Rps18 have shown that this gene can be stable (Scharlaken et al. 2008) or unstable (Najafpanah et al. 2013) as a reference for qPCR experiments. My study has found that, in addition to high variability, the expression level of Rps18 was very high compared to the majority of other genes (the average of 16.09 cycles, compared to other genes having around 20 cycles on average), and it increased during reprogramming. This level and the increase may reflect the growing need of the cell in protein synthesis because of metabolic alterations and increased proliferation rate. Such large differences in basal expression and high variability make Rps18 an unsuitable candidate for normalization, and its use as a reference should be avoided in future works on iPS reprogramming.

Overall, I concluded that the standard method of RT-qPCR, namely, relative quantification with the use of reference genes, is unsuitable for the analysis of gene expression during iPS reprogramming due to statistically significant fluctuations both in the housekeeping genes' expression and in reaction efficiency.

Chapter 2

Development of a new methodological tool for high-throughput quantitative analysis of gene expression during the iPS reprogramming process

2.1 Introduction

Since one of the goals of my work was to perform a medium-throughput analysis of the expression of genes throughout the reprogramming process, and the conventional method of analysis proved unsuitable for this purpose, the next step was to devise a satisfactory methodological tool for gene expression analysis during the iPS reprogramming process. To be suitable for the said purpose, the new method should satisfy the following requirements:

1) Does not require the use of reference genes
2) Does not require running a separate calibration curve analysis
3) Increases precision while decreasing pipetting workload and cost.

To satisfy the above-mentioned requirements, the new method should accomplish two goals in one experiment: 1) effectively determine the reaction efficiency (normally measured by calibration curve analysis, as required by MIQE guidelines), and 2) produce a single value of gene expression with increased precision for each sample.

The efficiency of the reaction is defined as the increase of product per cycle as a fraction of the amount present at the start of the cycle (Bustin et al. 2009; Ruijter et al. 2013). It is assumed that the efficiency of a qPCR reaction is stable and maximal before reaction saturation. Due to the exponential nature of PCR, the reaction efficiency can have dramatic effects on quantification measurements. It has been estimated that an uncorrected 0.05 difference in amplification efficiency between a reference gene and a target gene can lead to false estimation of the target

gene expression change of 432% (Rao et al. 2013). The calibration curve method is widely considered the most precise method for qPCR efficiency estimation (Svec et al. 2015). The calibration curve is built by creating a serial dilution of known DNA concentration and plotting the quantification cycle (Cq) values on the y-axis against the logarithm of the sample concentrations on the x-axis. The efficiency (E) is then estimated from the slope of this curve using the classical formula E=10-1/slope – 1; the estimation in this case is based on knowledge of the concentrations of all diluted samples. However, due to the insufficient precision of single dilution sets that could be caused by pipetting errors etc., it is often recommended to run at least three PCR reaction replicates for each sample to have three dilution sets for a single calibration curve.

In this chapter, I develop a new mathematical approach, Pairwise Efficiency, that improves the precision of estimations of qPCR efficiency, while reducing the workload. This approach does not rely on Cq values or amplification curve approximations. Instead, this new method applies pairwise approach to fluorescence data by calculating efficiency (E) using all possible pairwise combinations of fluorescence readings on several amplification curves of a dilution set. One pair of fluorescence readings allows to calculate a single E, while pairing up all fluorescence values allows to produce hundreds of E values and enables extensive statistics. I employ three statistical steps to increase precision: 1) first, I introduce a new formula for E estimation from a pair of fluorescence readings which allows me to use pairwise approach and produce hundreds of E estimations; 2) second, instead of using a single threshold, I define the wider boundaries for all curves from a dilution series, including more fluorescence values into the analysis; and 3) third, I use the resulting hundreds of E estimations as a large statistical dataset to perform more extensive and more mathematically accurate statistical analyses, such as analysis of value distributions, outlier removal and others. Because this approach is based on commonly used, robust statistical methods, it is systematic and can be applied in any setting and on any instrument as long as basic statistical principles are conserved.

In the later part of this Chapter, I compare the new Pairwise Efficiency method to the current “gold standard method” from the points of view of precision and

accuracy. Since these terms are of exceptional importance for demonstration of the value of my work, I find it necessary to define these terms here.

Precision is a measure of *random error* that arises in the process of measurement. It is related to reproducibility and repeatability, and constitutes the degree to which repeated measurements under unchanged conditions show the same results. This error always occurs when using sophisticated instruments, such as fluorescence readers etc., and can be attributed to measurement noise, sensor sensitivity or other such random factors.

Accuracy is a measure of *systematic error* (error that is "built into measurement system"). This type of error arises due to systematic factors, such as a defect in the measurement instrument (e.g. the 1-meter ruler is longer than 1 meter by mistake).

"True" value is the reference, or "*measurement standard*" to which a comparison is made. **In practice, such standards may not exist**. For example, a 1 meter standard exists in France in The International Bureau of Weights and Measures in Paris. In this sense, the only "true" 1 meter is in France (because it has been designated as such by humans). All measurements are, in this sense, trying to get as close as possible to that 1 meter standard, which is considered "true". However, in case of biological measurements, for example, efficiency of PCR reaction E, **such standard sample does not exist**. Thus, it is impossible to measure the E of the sample whose E is previously known because any measurement of "true" E needs a standard sample (akin to 1 meter) which currently does not exist in biology. In fact, the lack of such internationally recognized standard samples is one of the major problems of modern quantitative biology.

Thus, it is only possible to measure accuracy indirectly, by comparing the results of the measurements to other, existing standards (for example, microliters of dilution, or a chosen standard sample), which is what I do in the later part of the Results.

My results show that the application of Pairwise Efficiency makes it possible to nearly double the precision in qPCR efficiency measurements without increasing the pipetting workload and minimizing cost. In addition, I demonstrate a 2.3-fold improvement in precision of the estimation of gene expression ratios.

2.2 Materials and Methods

DNA sample

Mouse embryonic stem cell line E14Tg2a was purchased from RIKEN Cell bank, JP (AES0135) and was maintained as previously described. Total RNA was extracted using RNeasy kit (Cat# 74106, Qiagen, Japan) following the manufacturer's instructions. Genomic DNA digestion was performed on-column according to said instructions. RNA concentration and absorbance ratios ($A_{260/280}$ and $A_{260/230}$) were checked with a spectrophotometer Nanodrop 2000 Spectrophotometer (NanoDrop Technologies, Japan). To produce cDNA for qPCR analysis, 300 ng of total RNA were reverse-transcribed with an Omniscript RT Kit (Cat# 205111, Qiagen) in a total volume of 20 µl. The resulting DNA was assessed by spectrophotometric analysis and diluted to 100 ng/µl.

Quantitative real-time PCR setup and reagents

qPCR was performed using a CFX96 Connect apparatus (Bio-Rad, Japan). Hard-Shell® 96-Well PCR Plates (Cat # HSP 9601, Bio-Rad) sealed with optically clear adhesive seals (Microseal® 'B' seal, Cat # MSB1001, Bio-Rad) were used in all experiments. The thermocycler program consisted of an initial hot start cycle at 95°C for 3 min, followed by 33 cycles at 95°C for 10 sec and 59°C for 30 sec. Mouse actin beta (Actb) was amplified using the following primers: F-5'-AACCCTAAGGCCAACCGTGAA-3',
R-5'-ATGGCGTGAGGGAGAGCATA-3' (with estimated product length 194bp). The primers were used at a concentration of 300 nM. SYBR Green-based PCR supermix (Bio-Rad) was used for all reactions according to manufacturer's instructions. Each reaction was performed in a final volume of 8 µL. To confirm product specificity, a melting curve analysis was performed after each amplification, and agarose gel analysis was performed to ensure the amplification of the right product (Supplementary Fig. 3).

Experiment design and PCR dataset generation

For the assessment of precision of my method and comparison with the classical calibration curve method, I produced 16 replicas of a 6-step dilution series. Two datasets were generated from this experiment and processed using Bio-Rad CFX Manager 2.0 (2.0.885.0923). Dataset 1 consists of relative fluorescence data obtained from the aforementioned experiment: 6 serial dilution wells * 16 replicas = 96 wells. Fluorescence data in Dataset 1 are expressed as RFU (Relative Fluorescence Units) which is a term specific to Bio-Rad software. It is important to note that, since my goal was to improve the accuracy of the classical calibration curve, all RFU values were taken as already processed by Bio-Rad software with the same settings that were applied to the generation of Cq values, as follows: Baseline Setting set to Baseline Subtracted, Cq Determination Mode set to Single Threshold. Dataset 2 contains automatically generated Cq values corresponding to Dataset 1. The threshold was automatically set at 31.07 by the Bio-Rad software.

Determination of the exponential region

The most suitable bounds of the exponential region of the respective amplification curves were determined experimentally (see Results). However, prior to the experimental estimation, I conducted an initial estimation using well-known conventional techniques, namely, the "first outlier" method, the First Derivative Maximum (FDM) and Second Derivative Maximum (SDM) approaches (Lievens et al. 2012; Tichopad et al. 2003). Since the initial estimation was done solely in order to provide a general range for experimental testing, I chose the approaches mentioned above, even though other more sophisticated approaches have been suggested (Rao et al. 2013). The lower boundary of the exponential region has been defined as the point at which the signal significantly rises above the baseline level as determined by the formula of "first outlier" detection (Tichopad et al. 2003). The results of the formula application to the first calibration curve replica (wells A1 through A6) are provided in Supplementary Table 7. In agreement with these data, the tentative lower boundary of the exponential region was set at 10-40 RFU.

I also calculated the FDM and SDM values for all calibration curves. As expected, the values differed for samples with different initial DNA concentration, and were in the range of 17-23 cycles for SDM, and 18-25 cycles for FDM values.

Supplementary Figure 1a shows the FDM values for the whole Dataset 1 plotted against cycle numbers. The earliest FDM was encountered at cycle 18 in the most concentrated sample. The latest FDM of the dataset came at cycle 25. As shown in Supplementary Figure 1b, the RFU values for cycles corresponding to calculated FDMs fall in the range of 120-230 RFU. Thus, in accordance with these data, the tentative initial estimation of the upper boundary of the exponential region to use in the experimental test was set between 120-230 RFU.

Baseline treatment

Baseline is a software parameter inside any qPCR machine software. Since the goal of my analysis was to directly improve the precision of the classical calibration curve method, the same software settings were applied to fluorescence data as to the generation of Cq values. The Bio-Rad software was set to Baseline Subtracted, and the baseline was subtracted automatically by the software producing Relative Fluorescence Unit values. This Bio-Rad subtraction method is based on either adding a constant value, or a linearly growing value to the raw fluorescence and thus does not eliminate the noise.

Evaluation of the noise influence

Every qPCR machine produces technical noise. To determine the properties of noise and the scale of noise influence, I examined the fluorescence readings in the beginning cycles of the Dataset 1. As shown in Supplementary Fig. 4a, the fluorescence readings in the beginning cycles (up to cycle 13-18, depending on the starting concentration) were distributed close to 0, with inclusion of negative readings. The minimal value of the whole dataset was -9.44 RFU. To demonstrate the noise distribution, I show three histograms which contain fluorescence readings from the following cycles: 1) Cycles 1 through 5; 2) Cycles 1 through 10; and 3) Cycles 5 through 10. The data were taken from dataset 1 and two more 96-well plates replicating serial dilutions, with the Actb gene as target (raw data of these two plates are available on request). The total number of data points resulted in 2880 fluorescence readings (first 10 cycles from 96 wells in 3 plates). The result is shown in Supplementary Fig. 4b. The noise in the beginning cycles appeared to have a nearly normal distribution with a non-zero peak. The positions of the peaks

and the distribution did not change depending on the number of included cycles, which indicated that there was no detectable signal at this stage - because the increasing signal would have produced a shift to the right in the noise distribution if it existed. Thus, I concluded that the initial fluorescence readings in our system contain noise, and the noise has the approximate range of -10 RFU to 10 RFU. To ensure that all data points that I would take for analysis contain the non-noise signal, I concluded that the lower boundary should not be lower than 10 RFU which is in accordance with the boundary set by the 'first outlier'.

Results

2.3 A new formula for estimation of amplification efficiency

Since PCR amplification efficiency E is one of the most important parameters of the reaction which can have a dramatic effect on RT-qPCR measurements (as stated in the Introduction), I first approached the question of how to reduce the uncertainty in the estimation of E.

For this purpose I introduced a new formula (3) for E estimation from a dilution set. This formula describes the relationship between two independent fluorescence readings in any given dilution set. The fluorescence readings are represented by data points on six amplification curves, in the case of one six-step serial dilution experiment (Fig. 7b). The E estimation in my case is based on a relationship between a pair of actual fluorescence readings, as opposed to the slope of the calibration curve, which is based on cycle fraction values (Ct or Cq).

When devising the formula, I used the same basic assumptions that the calibration curve method uses (Ruijter et al. 2013; Guescini et al. 2008) when calculating the efficiency on a calibration curve, namely:

1) The kinetics of a PCR reaction with a given DNA-primer set is the same irrespective of the initial template concentration.

2) The kinetics of the PCR reaction are assumed to be classical (described by the classical formula F=F0*(1+E)i)

3) The efficiency is maximal and constant before the reaction saturation.

4) Fluorescence readings and double-stranded DNA concentration are linearly related to each other, and the increase in fluorescence is directly proportional to the increase in target concentration.

Given these assumptions, any single fluorescence reading F on any one of the amplification curves in the dilution set can be described by the following equations:

$$F_i = \frac{F_0}{2^{D1}} \times (1+E)^i \qquad (1)$$

$$F_j = \frac{F_0}{2^{D2}} \times (1+E)^j \qquad (2)$$

where i and j are cycle numbers for a particular fluorescence reading, F_i and F_j are the fluorescence readings in cycle i or cycle j, F_0 is the initial fluorescence of the undiluted sample, D1 and D2 are dilution factors for curve 1 and curve 2 (if the pair of data points are on the same curve, then D1=D2), and E is the amplification efficiency for the qPCR reaction for the given DNA-primer set. The dilution factor D is defined as the logarithm of the fold-dilution, compared to the undiluted sample whose logarithm of the fold-dilution, by definition, is 0. Since I applied twofold dilutions for mathematical clarity, D values in this case were integers from 0 to 5. In the case of tenfold dilutions, the corresponding '2' values in the formulae will become 10, and the dilution factors will remain unchanged.

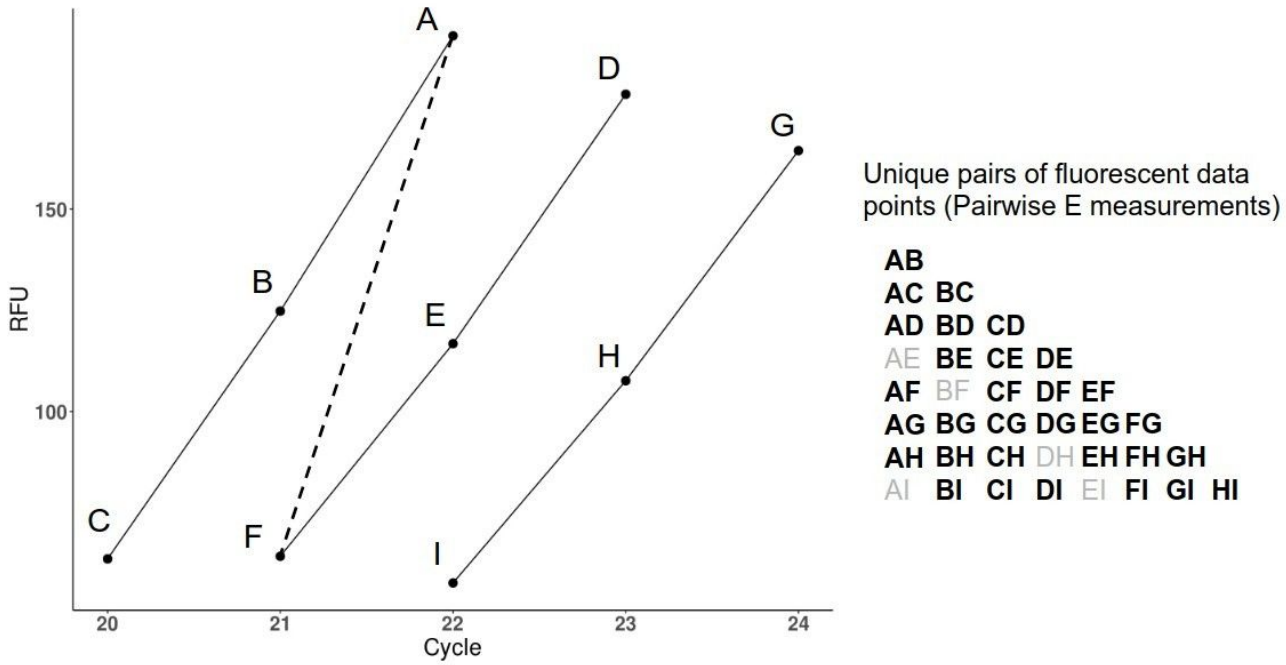
The equations 2 and 3 allow me to calculate the efficiency E for a given pair of fluorescence readings, such as:

$$E_{i,j} = 2^{\frac{(log_2(F_j) - log_2(F_i) + (D2 - D1))}{(j-i)}} - 1 \qquad (3)$$

Thus, while the estimation of efficiency across a dilution set by the calibration curve method is based on a single curve and produces a single E value, my new method, Pairwise Efficiency, calculates an array of E values based on all possible pairwise combinations of fluorescence readings from this dilution set, producing about 50-400 individual pairwise E measurements (depending on the number of fluorescence readings included in the exponential region taken for analysis), and then estimates the average efficiency from this array of E measurements.

In one of the classical pairwise approaches, namely Walsh Averages, a set of data points is treated the following way. Each data point is paired up with all other data points in this set (including itself), thus creating all possible combinations of pairs. For example, the set [3,8] has three pairs: (3,3), (8,8) and (3,8). The difference in my approach is only that I exclude self-paired values (such as (3,3)). Each pair in this approach is unique, and in Walsh averages gives a unique average value (E value produced from this pair in my case).

a


A
B
C
D
E
F
G
H
I
RFU
150
100
20
21
22
23
24
Cycle
Unique pairs of fluorescent data points (Pairwise E measurements)
AB
AC BC
AD BD CD
AE BE CE DE
AF BF CF DF EF
AG BG CG DG EG FG
AH BH CH DH EH FH GH
AI BI CI DI EI FI GI HI

b

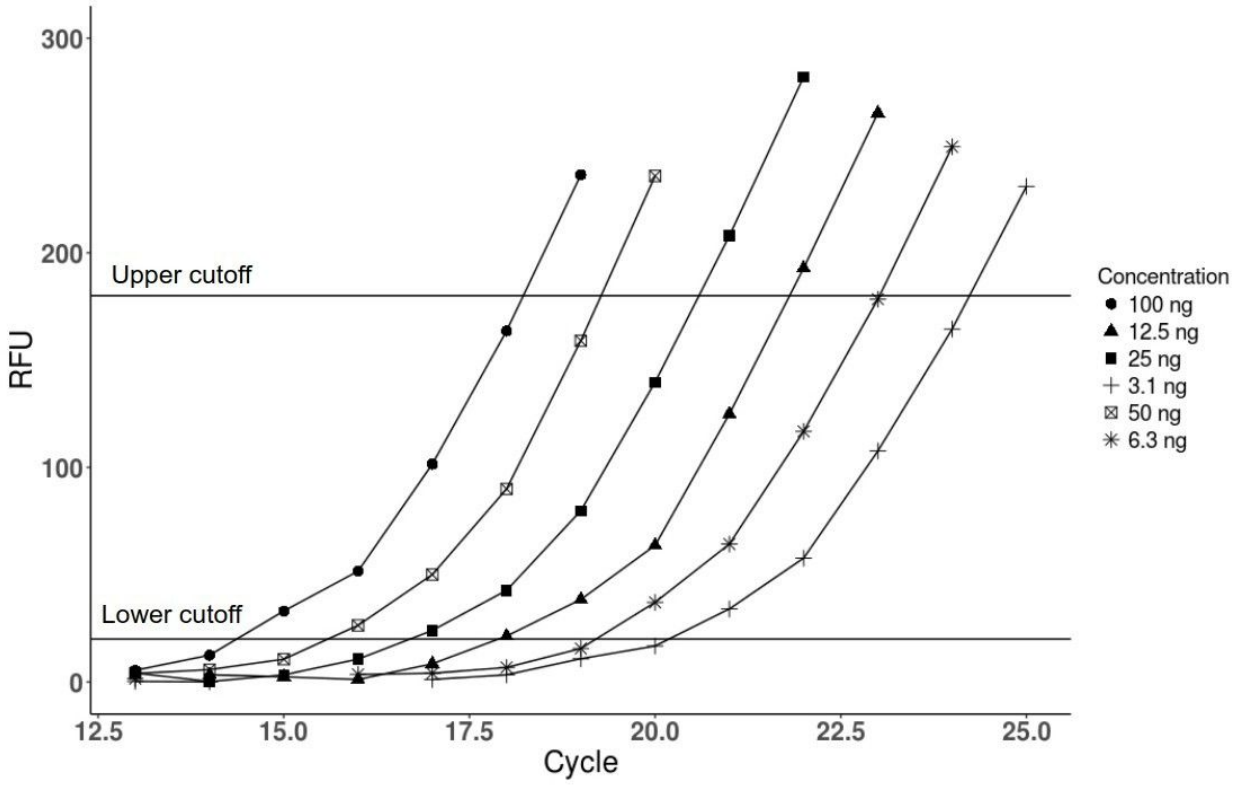

300
200
100
0
RFU
Upper cutoff
Lower cutoff
12.5
15.0
17.5
20.0
22.5
25.0
Cycle
Concentration
100 ng
12.5 ng
25 ng
3.1 ng
50 ng
6.3 ng

Figure 7. Graphical representation of the principle of Pairwise Efficiency method and its application to six dilution curves. (a) A graphical illustration of the Pairwise Efficiency method. Small portions of three amplification curves, with three fluorescence data points on each, are shown. Dashed line connects point A to point F on separate curves, and represents a single, unique pairwise E measurement (pair AF). All possible pairs, each one representing a unique pairwise E value, are shown on the right. Since some of the values occur on the same cycle (for example, AE, BF), and thus are excluded from the measurements, and are denoted in gray. (b) The amplification curves from the wells C1 through C6 are shown (RFU data taken from Dataset 1). Different shapes (circles, squares, triangles etc.) represent fluorescence readings taken by the machine after each PCR cycle. Horizontal lines denote the region of amplification curves from which the fluorescence data points were taken for analysis. Upper cutoff was set at 180 RFU, and lower cutoff was set at 20 RFU. In this experiment, the total of 24 fluorescence data points fall inside the denoted region, and unique pairs formed by these 24 points, excluding repetitive values occurring on the same cycle, are taken for analysis.

In other words, after gathering all fluorescence RFU readings that fall within boundaries for each amplification curve, I further treat them as a statistical set (usually a set of 4x6=24, on my qPCR machine, or more if the machine sensor has higher sensitivity) and calculate all possible unique pairwise combinations from this set (24 RFU values) using my formula (3). As a result, I obtain a statistical population of unique E measurements which I can further analyse using chosen statistical methods. Thus, my first difference from classical approach is collecting more fluorescence values from a single curve, and **an additional step** is performing pairwise calculations to produce a statistical population of unique E measurements.

2.4 Assessment of the detectability of stable amplification efficiency in the exponential phase

Next, I approached the question of defining the exponential region of the reaction data to ensure that non-exponential values would be properly excluded from the analysis. According to the mainstream view, any PCR reaction proceeds with stable efficiency until end-stage reagent depletion and the accumulation of reaction products cause a steep decline in the efficiency, and the reaction gradually slows down (Bar et al. 2012; Archer 2017). The calibration curve method aims at measuring the stable efficiency of the reaction before the saturation occurs, and this maximal efficiency is assumed to be identical across all dilution samples. However, it has been argued that the sensitivity of some qPCR machines does not allow detection of a weak fluorescent signal in the exponential phase of the PCR reaction, where the efficiency is still stable, and the signal first appears when the efficiency is already declining (Lievens et al. 2012; Rutledge and Stewart 2008; Tellinghuisen and Spiess 2014). It has also been pointed out that the analyses based on stable efficiency should be conducted strictly at the region before efficiency decline, if such a region is detectable.

To determine if my system allows to detect the theoretical stable efficiency, I analyzed the fluorescence readings data from Dataset 1 (see Materials and Methods for description) using the following formula for the calculation of efficiency E.

$$E = 2^{\frac{log_2 F_i - log_2 F_0}{i}} - 1 \qquad (4)$$

where i is the cycle number for a particular fluorescence reading F, and F_0 is the initial fluorescence value of the sample. The logarithms, base 2, are used because the series contains 2-fold dilution sets.

The formula (4) cannot be used directly for E calculation because the fluorescence level of the starting material F0 is unknown. The purpose of the analysis described below was to confirm the detectability of the stable exponential E region with varying F0 values. To obtain initial approximation of F0 value to test with formula (4), I used E values calculated using calibration curve method (Supplementary Table 3). Knowing the efficiency of the reaction (around 80%)

allowed me to produce initial F0 estimations by the standard formula. The resulting F0 values were in the range of 0.007 to 0.0002. I then substituted these F0 values in the formula (4) and analyzed the resulting E values at each cycle of the reaction (Fig. 8). As shown in the figure, I found that in the first cycles where non-background signal is detected by the machine, E displays a relatively constant pattern (SD=0.01), while in the later cycles it starts to decline steadily (Supplementary Table 4). The initial region with the small standard deviation lasted from cycle 13 until cycle 17 for the most concentrated sample. Varying the F0 value did not affect the detection of this region of relatively constant E, as other curves also produced a similar pattern with small variation of E in the initial 3-5 cycles where the signal was already detected, and a steady decline after that.

According to these data, my experimental system allowed the detection of approximately 4 fluorescence values from the exponential phase of amplification where the variation of efficiency does not exceed ±0.01. This result overall shows that the theoretical stable efficiency is detectable and can be quantified.

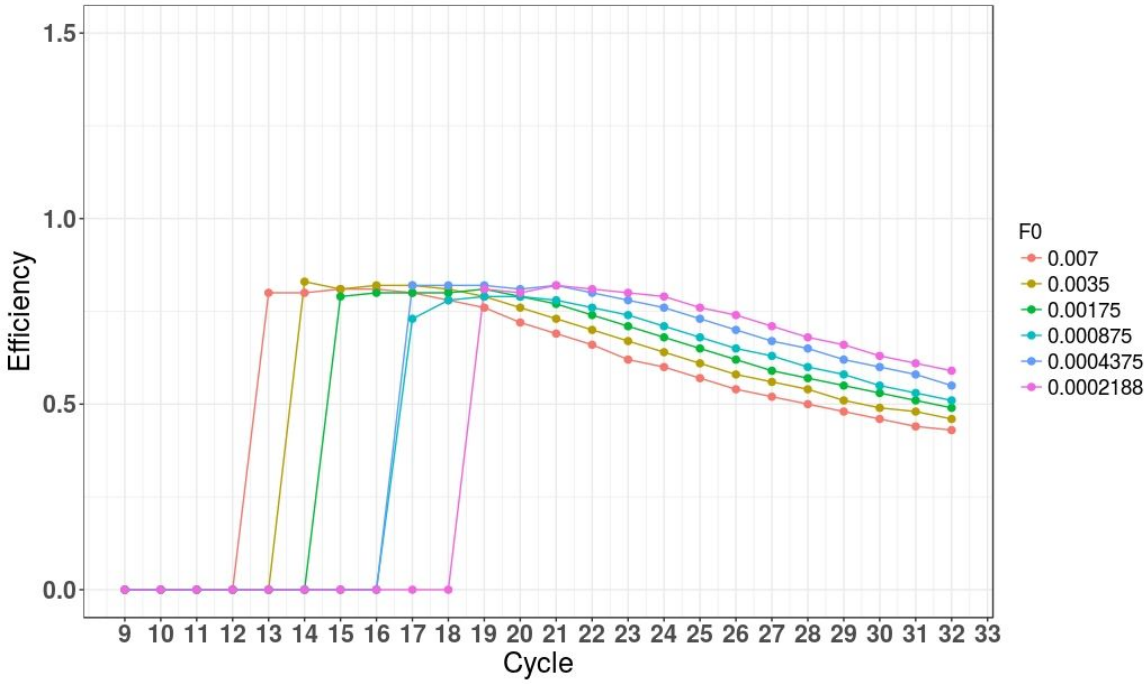


Figure 8. A graphical representation of the efficiency (E) values across all cycles taken from a 6-step dilution set. Efficiency is calculated using the formula $E = 2^{\frac{log_2 F_i - log_2 F_0}{i}} - 1$. The F_i and i values for calculation are taken directly from Dataset 1, wells A1 through A6. Since F_0 value is unknown, it was selected from

the range of theoretically possible F_0 values (covering 0.007 - 0.0002) and used in the formula.

2.5 Experimental determination of lower and higher boundaries

Because each dilution set produces amplification curves based on different concentrations of starting material, one can expect that the exponential region of each curve should start at a different cycle. Thus, it is necessary to experimentally determine the most suitable upper and lower boundaries of the exponential region for all curves taken together. An incorrect determination of the boundaries and subsequent inclusion of non-exponential values would be a major source of error in E estimation. To determine the most suitable boundaries for my system, I experimentally tested at what fluorescence range (i.e. what portion of each of the amplification curves) the application of Pairwise Efficiency method produces results with the highest precision. For the estimation of precision I applied modified Monte Carlo approach that was previously described by Svec et.al. for the evaluation of precision of the calibration curve method (Svec et al. 2015). The essence of Monte Carlo method is described below. To find the value *a* of a certain quantity, one chooses a random variable *X*, the mathematical expectation of which equals *a*:

$$M(X) = a$$

In practice, this method is applied by conducting *n* real measurements and obtaining *n* values of *X*, and then calculating their average. Mathematically, this method leads to the increase in precision of measurement due to the increased number of performed measurements and is based on the central limit theorem; with repeated measurements, the error of the arithmetic mean decreases depending on the number of measurements:

$$\sigma_{cp} = \frac{\sigma}{\sqrt{n}}$$

Where σ_{cp} is the error of the average value of multiple measurements;
σ is a single measurement error;
n is the number of measurements.

The calibration curve method described in MIQE guidelines requires researchers to perform at least three replicative measurements of the Ct value for this very reason, which is to increase precision by increasing the number of measurements. However, since it always uses Ct values which represent only one data point on an amplification curve, each additional measurement must be obtained by manually pipetting one more replicate of the sample. In contrast, Pairwise Efficiency uses the whole array of fluorescence data available from the exponential phase of the amplification curve defined by lower and upper boundaries, and operates hundreds of measurements without additional pipetting load.

In case of Monte Carlo simulations, for the purpose of precision estimation it is sometimes permissible to apply computer simulation of pseudo-random numbers to increase the number of measurements. Such application is described in Svec et.al. for assessment of precision of the calibration curve method depending on the number of taken replicas (one, two or three). The authors calculated PCR efficiency by standard calibration curve method of 6 dilution steps. For each of the step they performed four technical replicates. Then they randomly formed data sets using either one, or two, or three replicas to calculate the efficiency, and these replicas were taken from the set of available four replicas for each dilution step. Thus, the four replicas represented a sample from a "pseudo-general" population from which random samples were taken.

Similarly to the method described above, I took random samplings from pseudo-general population, with the difference being that my population was much larger and contained 16 replicas instead of 4. To evaluate the precision for different

boundaries, I randomly drew 100 different six-sets from the general population of 16 (Fig. 9), and calculated the precision for each combination expressed as standard deviation (SD). The results of this operation are displayed in Fig. 10. Exact SD values and other specifications can be found in Supplementary Table 5.

The lower boundary was tested at the range of 10 RFU - 80 RFU, and the higher boundary was tested at the range of 120 RFU - 230 RFU (see Materials and Methods).

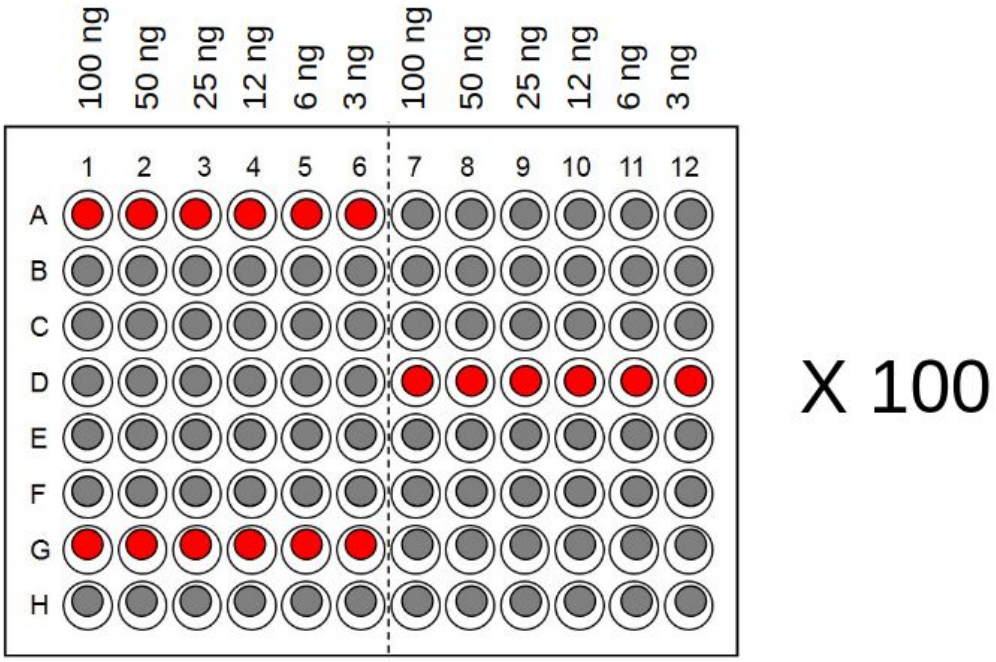


Figure 9. Schematic representation of Monte Carlo simulation for assessment of precision. All 96 wells contain samples of different concentration (written above the plate). Technically, this constitutes 16 sets of six-step dilution series, 8 on the left half of the plate, and 8 on the right half of the plate. These 16 sets are identical and represent the general population on which Monte Carlo simulation is based. For each assessment of SD of a given boundary set (for example, for boundaries 10 RFU-180 RFU), 100 pseudo-random measurements are performed using Pairwise Efficiency method. Each pseudo-random measurement consists of averaged 3-set of randomly chosen dilution series (represented by red circles in the figure). The SDs of different boundaries are then compared to each other.

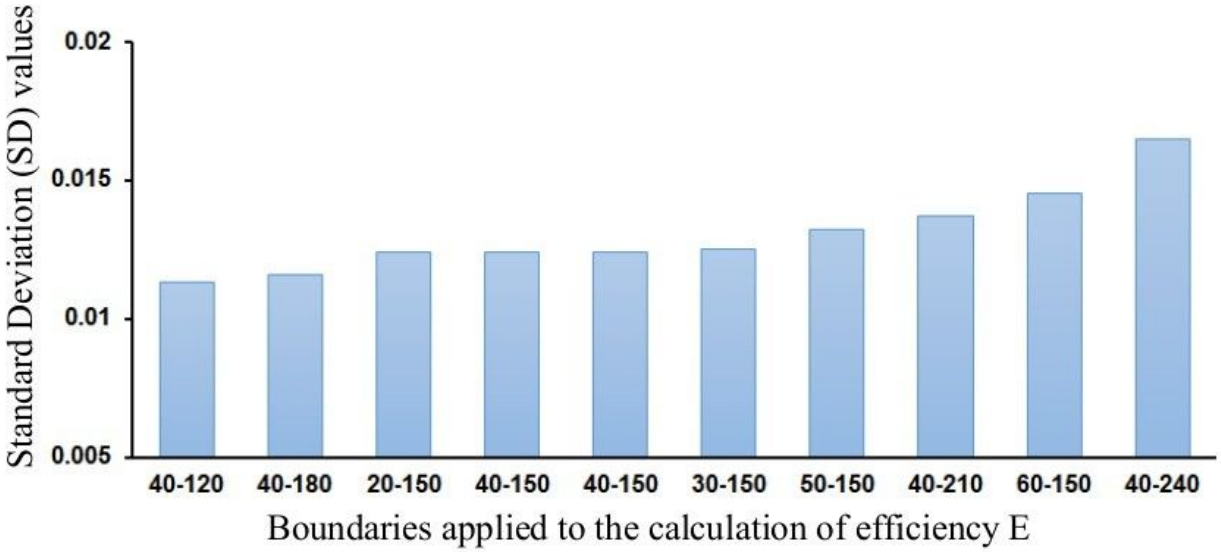


Figure 10. **Determination of the most suitable RFU boundaries for a 6-step dilution series.** Standard deviations (SD) of the efficiency values calculated by Monte Carlo approach using different regions of amplification curves. The average efficiency (E) of the 6-step dilution set was calculated based on randomly selecting data out of 16 set replicas. Each time different portions of the amplification curves were included in the calculations, defined by lower and upper boundaries. The lower boundary varied between 20 RFU and 80 RFU, while the upper boundary varied from 120 RFU to 240 RFU. The lowest SD was obtained when applying the following boundaries: lower at 40 RFU and upper at 120 RFU. The SD tended to rise when boundaries were raised.

While varying the boundaries within the exponential region (for definition of exponential region see Materials and Methods) did not produce a significant difference in SD values, and the best result was obtained at the lower portion of the curve (40-120 RFU) (Fig. 10). The variation in the SD value did not exceed 0.001 for the lower portion (40-120 RFU, 40-150 RFU, 20-150 RFU). This result is in agreement with previous studies reporting that the threshold for measurements is best set at the lower portion of the amplification curve because the declining efficiency in later cycles might significantly affect the results (Archer 2017). I

conclude that slight variation of lower and upper boundaries does not significantly affect the precision of E estimation. This is well in agreement with my data on exponential region estimation (see Materials and Methods) which puts the upper boundary at the fluorescence range 120-230 RFU and the lower boundary at the fluorescence range 20-40 RFU (Supplementary Figure 1). To include as many values as possible in my case, I decided to use 20-180 RFU boundaries, which allows the inclusion of approximately 4 fluorescence data points (Fig. 7b).

2.6 Statistical elimination of outliers

Statistical outliers are data points that are numerically distant (far removed) from the rest of the data points in the general population. Outliers occur in many statistical analyses and can be a chance phenomenon, a measurement error or an experimental error. The origin of the outliers in any particular case could be separately investigated if needed; however, it is not necessary to know their origin to conduct statistical operation on the outliers. The origin of the outliers in any RT-qPCR experiments could be:

1) Electrical noise in the measurement instrument inherent to the machine
2) Pipetting mistakes of the operator in one or more samples
3) Dilution errors of the operator in one or more samples
4) Other factors.

Excluding such values from the calculations is an important additional way to increase precision of measurements, and in my case it is possible exactly because of the availability of the vast array of measurement data (hundreds of points). Elimination of outliers is statistically impossible when only three or four data points (replicates) are available, as is the case of the standard method. Thus, mathematically increasing the number of data points by using Pairwise Efficiency formula allows both direct improvement in precision, and more robust statistical analysis such as outlier elimination.

In my case, I utilized the elimination of outliers technique that is often used in statistics to exclude unreasonable values that occur due to random measurement errors, and to increase the precision of measurements. First, I analyzed the

distribution of pairwise E values for normality in each group of pairwise E measurements. This analysis is necessary in order to decide which kind of method to use for outlier exclusion (parametric, such as three sigma rule, vs. non-parametric). To assess the distribution normality in a mathematically objective way, I used standard tools, namely, skewness, kurtosis, and chi-square test. As shown in Table 4, the majority of skewness values significantly deviated from 0, confirming distribution asymmetry.

Dilution set (wells)	Skew	Kurtosis	N of E measurements
A1-6	1.064	7.357	167
B1-6	0.615	4.085	168
C1-6	0.221	3.556	170
D1-6	1.051	6.305	183
E1-6	0.473	5.524	168
F1-6	1.880	6.769	152
G1-6	2.012	10.079	182
H1-6	1.379	12.177	168
A7-12	-0.337	2.160	168
B7-12	0.098	4.508	149
C7-12	0.215	2.838	204
D7-12	0.739	2.514	168
E7-12	0.563	3.555	188
F7-12	-0.034	3.843	171
G7-12	1.429	7.023	152
H7-12	-0.148	5.319	188

Table 4. Estimation of distribution normality. Pairwise E values of 16 dilution sets were analyzed for skewness and kurtosis. Skewness values that deviate from 0 indicate asymmetry of the distribution, making it a non-normal distribution. Positive kurtosis values also imply deviation from normal distribution and indicate that the distribution is sharp (more values are close to mathematical expectation, and precision is higher than would be expected in the case of normal distribution). The right column contains the total data points for each dilution set that were taken for this analysis.

Thus, for example, the dilution set in wells A1 through A6 had 167 individual pairwise E measurements, skewness=1.06 and kurtosis=7.36. The frequency of E values below 5 was first encountered at E=0.6 (60% efficiency) on the lower end, and at E=1.15 (115% efficiency) on the higher end (for more information see Supplementary Table 6 and the next passage on Chi-square criteria). Based on Chi-square criteria, all pairwise E measurements that exceeded 115% and did not reach 65% were excluded from the calculation of average E for this dilution set. E value for wells A1 through A6 prior to outlier analysis was E=0.79, and after the removal of outliers became E=0.816. Other E values for the remaining 15 sets were processed on the basis of the same algorithm.

In addition, all kurtosis values were positive, indicating that calculated pairwise E measurements from these dilution sets had leptokurtic distribution (Fig 11).

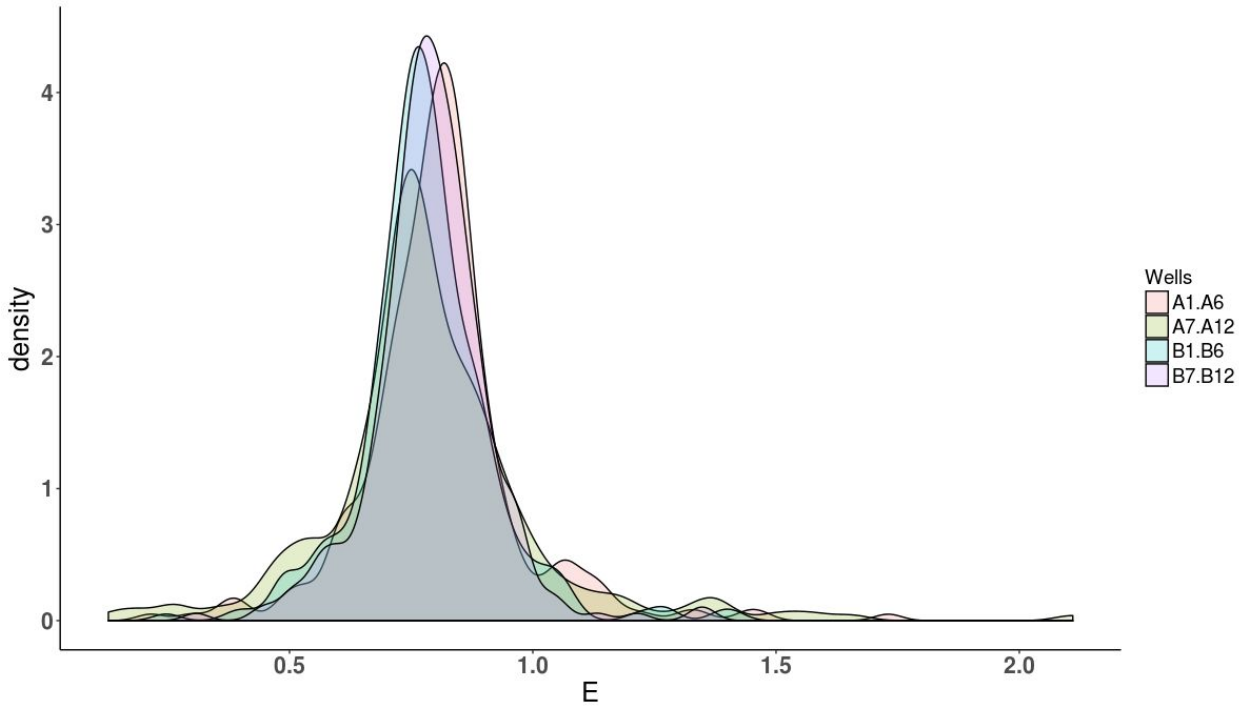


Figure 11. A graphical representation of the distribution of pairwise E values for the wells A1-A12 and B1-B12. The distribution of pairwise E values is leptokurtic in all sets, and has a sharp appearance, indicating that the values are closer to mathematical expectation, and precision is higher than would be expected

in the case of normal distribution. In addition, the distributions are skewed and possess larger tail areas, indicating significant deviation from normality.

Next, I used the Pearson's chi-squared test to test the goodness of fit of the frequency distribution of calculated pairwise E values. The application of this test is considered valid if there are at least 50 values analyzed for distribution (which is the case of Pairwise Efficiency), and no more than 20% of the values have expected frequencies below 5. The values whose frequency is less than 5 are considered statistically unreliable and are designated as outliers. An analysis by the Chi-square test showed that the majority of the distributions (12 out of 16) significantly deviated from normal (Supplementary Table 6 and Fig. 12). Thus, parametric tools designed for normally distributed values, such as quartile ranges or sigma rules, could not be applied in this case. Instead, when the distributions do not follow a fixed set of parameters (e.g. are not normal), non-parametric statistical tools are used; however, the selection of specific tool is left to the researcher and is decided case-by-case. Since Pearson’s chi-square test is a universal tool that can be applied to any kind of distribution (both parametrized and non-parametrized), I chose to use the criteria of this test to exclude outlier E values in our case. As mentioned above, according to the principles of the Pearson's chi-square test, the values whose frequency is less than 5 are considered statistically unreliable. Based on this criterion, the pairwise E measurements with frequency less than 5 were considered outliers and were excluded from the calculation of the average E value of the dilution set.

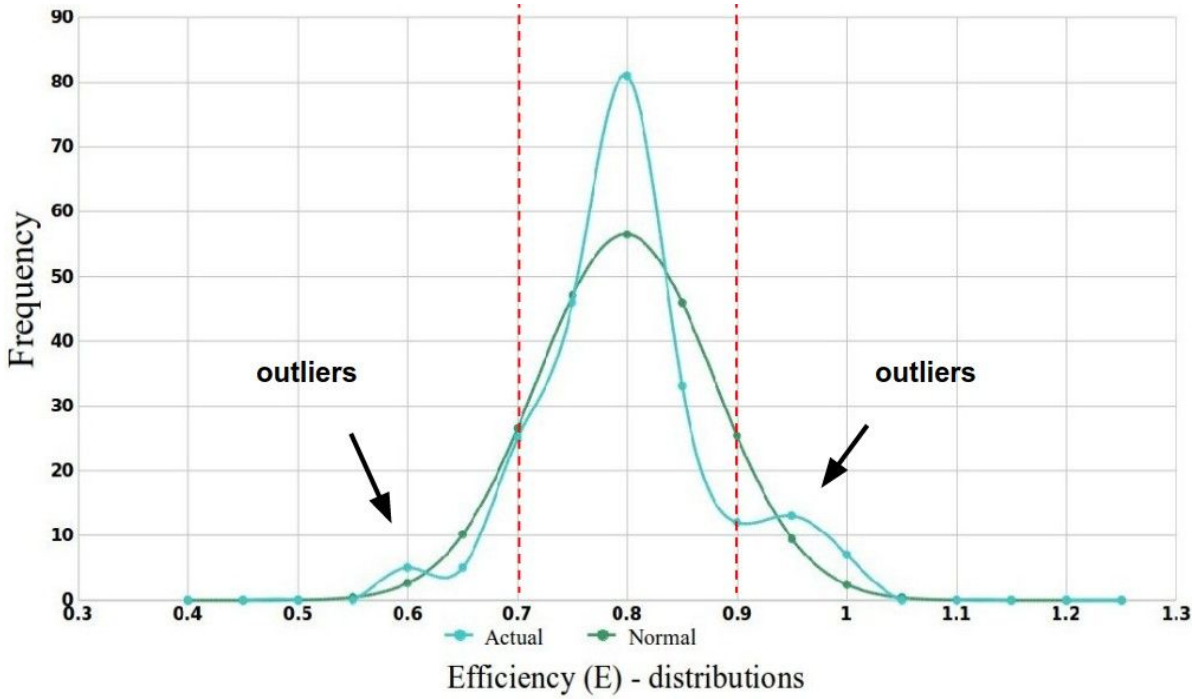


Figure 12. **A graphical representation of the distribution of pairwise E values for the wells H7-H12 compared to normal distribution, and the principle of the outlier exclusion.** The distribution of pairwise E values is leptokurtic (has a sharp peak), indicating that the values are closer to mathematical expectation, and that the precision is higher than would be expected in the case of normal distribution. In addition, skewness is present in this distribution as compared to normal distribution, indicating significant deviation from normality.

2.7 Comparison of Pairwise Efficiency method with the calibration curve-based E estimation by precision

Next, I set out to compare the **precision** of my method to the classical calibration curve method. Since precision is defined as a measure of random error, it can be measured by the same Monte Carlo approach that was used for comparison of different boundaries described above in section 2.4. Again, as shown in Fig. 9, I took more than 100 samplings of three replicas of dilution sets at random, and, for the classical calibration curve, used them to generate calibration

curves. Then, individual E estimations were made based on these randomly produced calibration curves, and standard deviation (SD) was calculated for the produced E values. The SD value for E estimation found to be 0.019. Next, I applied the same approach to the corresponding RFU values by Pairwise Efficiency method and calculated SD for it as well. The results are shown in Table 5. Pairwise Efficiency produced a decrease in SD (increase in precision) of E estimation from SD=0.019 to SD=0.010, thus nearly two-fold. While the average E values were found to be 80% in both methods, Pairwise Efficiency produced a smaller standard deviation and a smaller difference between maximal and minimal E values. The dispersion of E values obtained by Pairwise Efficiency method, expressed as Max E - Min E, did not exceed 0.045, as opposed to 0.072 obtained by the calibration curve method. This means that the magnitude of random error in the E estimation was approximately two times lower in the case of Pairwise Efficiency compared to the calibration curve method.

Approach	SD	Max E	Min E	Max-Min difference	Average E
Calibration curve	**0.019**	0.83	0.76	0.072	0.80
Pairwise Efficiency	**0.010**	0.82	0.78	0.047	0.80

Table 5. Comparison of the calibration curve method with the Pairwise Efficiency method. Standard deviations (SD) obtained from the Monte Carlo test, maximal and minimal efficiency values, the range between maximal and minimal values, and the average efficiencies are shown. While the average E value was the same for both methods (E=0.80), the precision of E estimation obtained by the Pairwise Efficiency method, expressed as standard deviation (SD), was nearly two times higher, and the dispersion, expressed as the difference between maximal and minimal calculated E values, was 1.6 times smaller.

Next, I investigated whether this increased precision in the efficiency estimation would translate into increased precision of gene expression ratio measurements. To do that, I calculated the magnitude of possible error for the calibration curve method and for the Pairwise Efficiency method, using the same assumptions as

described in Materials and Methods. For the calculation of expression ratios in the case of calibration curve, I used the equations described by M. Pfaffl (Pfaffl 2001). The mathematical model presented in his publication is, in principle, equivalent to the model previously designed by Roche Diagnostics and takes into account the efficiency of both target and reference genes. The formula presented by Pfaffl has the following appearance:

$$ratio = E_{target}^{\Delta Ct}$$

where ΔCt is the difference between Ct of the sample and Ct of control at the same threshold. This formula was devised by Pfaffl for calculating the efficiency and relative expression ratio based on dilution curve for only the target gene. Since my dataset of 16 dilution replicas contained exactly the same amount of target gene (Actb) in wells with the same concentration, theoretically the calculated ratio between these wells should be 1. Thus, I could measure the magnitude of error in the determination of the ratio by measuring maximal difference between each one of these 16 replicas. In this case, the error would be maximal when the efficiency value is maximal.

First, I determined which one of the 16 dilution sets gives the highest efficiency value. The analysis using the calibration curve method showed that wells D1 through D6 produced the highest efficiency (E=0.882). Next, using this efficiency, I applied the aforementioned formula for the undiluted samples, considering the Ct sample the highest Ct from all 16 replicas, and Ct control the lowest of all. This resulted in a ratio = 1.606. Thus, the maximal possible error in the estimation of gene expression ratio when using the calibration curve method can reach up to 60%. Similarly, I used the maximal efficiency calculated by Pairwise Efficiency method to estimate the magnitude of error on Dataset 1 with 16 replicas. The maximal efficiency value was obtained in the same wells (D1 through D6) as for the calibration curve, which indicates robustness of both methods for E estimation. Using this maximal efficiency value, I estimated F0 in all wells using my modified formula:

$$F_0 = \frac{F_i}{(1+E)^i}$$

based on actual fluorescence values. The estimation of F0 in Pairwise Efficiency method in this case was analogous to the calibration curve method, while the way I estimate efficiency differed. I obtained the following result: Max F=0.00435436, Min F=0.00345735. Then I calculated the difference between maximal F0 and minimal F0 which yielded a ratio=1.26. Thus, the magnitude of possible error in ratio estimation using Pairwise Efficiency method amounts to 26%, which amounts to an **improvement of about 2.3 fold in the precision of gene expression ratio estimation** compared to the calibration curve method.

2.8 Comparison of Pairwise Efficiency method with the calibration curve-based E estimation by accuracy

Accuracy is a measure of systematic error, and can only be determined by comparing the sample to a known standard. Biological standards for RT-qPCR do not exist. Thus, it is only possible to determine accuracy indirectly, for example, by comparing the measurements to other known values (such as dilution proportions etc.).

Thus, I compared Pairwise Efficiency method to the classical method by their ability to reflect dilution proportions. In my case, the known values were the dilution proportions represented by cDNA concentrations (100 ng, 50 ng, 12.5 ng, 3 ng) which should result in the following proportions: 1, 2, 8, 32. The closer the ratio result to the known dilution value, the higher the accuracy. I calculated the ratios and the error for the known concentrations using Pairwise Efficiency method and the classical calibration curve method, and then determined the error for each method (Table 6).

Wells	Conc.	Efficiency	F0	Ratio	Error (%)	Ratio (Ct)	Error (%)
A1-A6	100ng	0.73130	0.00800	1	N/A	1	N/A
A7-A12	100ng	0.76200	0.00780				

B1-B6	100ng	0.77170	0.00660				
B7-B12	100ng	0.77230	0.00710				
C1-C6	50ng	0.83530	0.00280	2.513	20%	2.47	19%
C7-C12	50ng	0.79550	0.00290				
D1-D6	50ng	0.81870	0.00290				
D7-D12	50ng	0.82390	0.00300				
E1-E6	12ng	0.75780	0.00060	8.519	6%	12.73	37%
E7-E12	12ng	0.68420	0.00110				
F1-F6	12ng	0.72470	0.00090				
F7-F12	12ng	0.70420	0.00100				
G1-G6	3ng	0.76180	0.00020	35.455	10%	57.41	44%
G7-G12	3ng	0.66870	0.00020				
H1-H6	3ng	0.72810	0.00020				
H7-H12	3ng	0.66640	0.00020	Average	28%		67%

Table 6. Comparison of the accuracy between Pairwise Efficiency and the standard calibration curve method based on the ability to detect known dilution proportions. The ratio between different concentrations of Actin beta was measured using Pairwise Efficiency or the standard calibration curve method. The average error for Pairwise Efficiency is 28%, while the average error for standard method is 67%.

Pairwise Efficiency could detect known dilution proportions with much better accuracy compared to standard RT-qPCR method overall. It is also interesting to note that the error for Pairwise Efficiency was largest at the smallest volume (x2 dilution, 20% error), while for standard method the error tended to increase with the dilution volume (19%, 37%, 44%). Overall, Pairwise Efficiency outperforms the standard method by a large margin on accuracy, being able to detect small quantities of RNA with much less error.

2.9 Discussion

Quantitative PCR is an affordable and widely used technique for nucleic acid quantification. However, despite its popularity, this method has yet to gain full acceptance in the research community due to limitations in its ability to provide precise measurements, which may lead to low reproducibility. In chapter 1 of this thesis, I have shown that the housekeeping genes analyzed by conventional qPCR approach display high variability and are unsuitable for use in the relative (standard) gene expression quantification by qPCR. Instead, I introduce a new approach to qPCR data analysis, Pairwise Efficiency, which consists of three elements. First, it introduces a formula describing the relationship between two fluorescence readings on amplification curves, and does not rely on Cq values or a calibration curve for the estimation of reaction efficiency. Second, it estimates the boundaries of the exponential region for a group of amplification curves in order to determine reliable data boundaries. And third, it eliminates outliers during the process of calculating E values, as opposed to at the end.

The most important advantage of the new Pairwise Efficiency method is the increase in data points available for analysis due to the introduction of the new formula for efficiency estimation. Such increase in available data enables one to use sophisticated statistical instruments. **No other current method of RT-qPCR operates hundreds of data points at once**, and the standard method (recommended by MIQE guidelines) operates 16 data points at most. This advantage of Pairwise Efficiency includes the following factors:

1. While the standard method uses only one point on the amplification curve (Ct, or fractional cycle at the threshold), Pairwise Efficiency uses the whole array of available fluorescence points (RFU) in the exponential region of the amplification curve. This increases available statistical data from 16 to hundreds, and improves precision.

2. The calculations in Pairwise Efficiency do not depend on the Ct value which is automatically set by the machine. Ct values depend on the algorithm, machine, maker, data etc, and differ for different conditions, thus introducing more fluctuation into the data. In contrast, Pairwise Efficiency operates only the actual

fluorescence data which do not depend on the settings of the machine or the experimenter and always stay the same.

3. The outlier elimination process in the case of the standard method is not optimal: to determine "bad" reactions it uses the efficiency value obtained by the calibration curve. MIQE stipulates that "robust and precise qPCR assays are usually correlated with high PCR efficiency", and considers the efficiency to be an indicator of assay quality. In cases in which the E value exceeds the theoretical maximum of 100%, it is taken to be the result of reaction inhibition in one of the wells, generally meaning that the entire assay needs to be repeated or redesigned (Bustin et al. 2009). In contrast, Pairwise Efficiency allows one to analyze each amplification curve separately, detect outliers in the process of analysis, and exclude potentially bad data from the final calculation of gene expression ratio. This significantly reduces the amount of workload (Fig. 13).

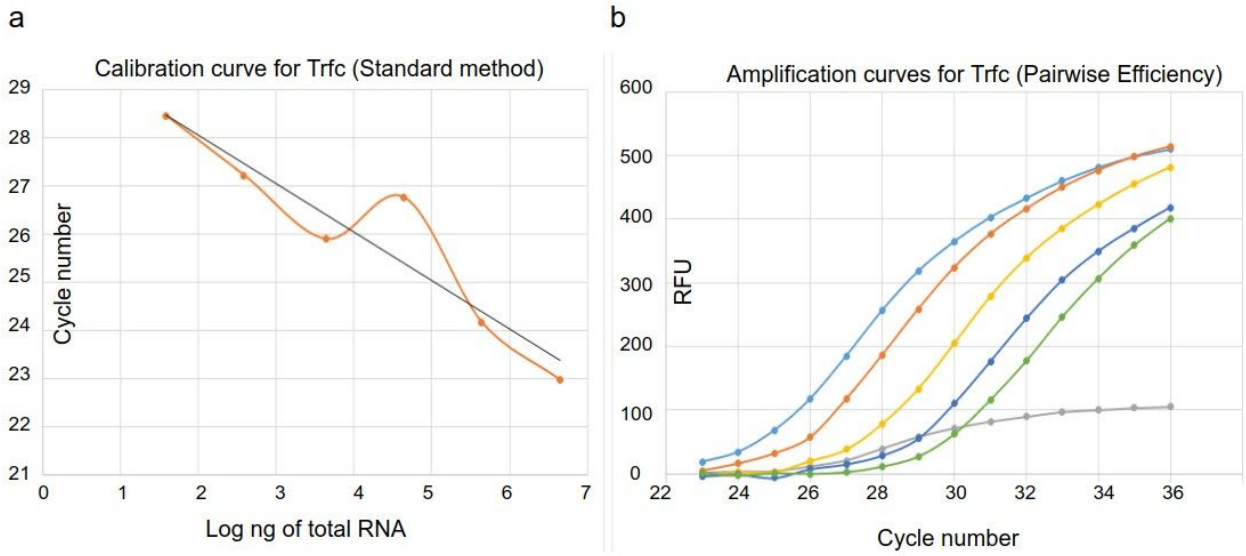


Figure 13. Demonstration of one of the important advantages of Pairwise Efficiency using data obtained from running RT-qPCR experiment on Tfrc gene. (a) Six-well dilution series was pipetted and RT-qPCR was run for Tfrc gene. Then, Ct values for each of the concentrations were plotted against the logarithm of known concentration. This allowed to draw a calibration curve (orange). This calibration curve significantly deviates from linear shape, and it is impossible to determine what was the cause of it. According to MIQE, this experiment is considered "bad", and the entire experiment must be re-performed.

(b) The six amplification curves obtained from six-well serial dilution have been generated based on fluorescence (RFU) values. Each curve corresponds to one well on the plate (one PCR reaction). The grey curve can be seen deviating significantly from normal reaction pattern. Thus, since Pairwise Efficiency operates the data taken from amplification curves directly, it can detect the abnormal curve and exclude it from calculations. Other curves remain valid and this allows the researcher to use available experimental data without the need to re-perform the entire set.

A significant advantage of Pairwise Efficiency is that it relies on actual fluorescence readings rather than implied data. It has been previously pointed out that the estimation of efficiency by the means of a calibration curve, as required by MIQE guidelines, is based not on existing, but rather on implied data: "the data from a tube is discontinuous; fluorescence is measured at the end of each cycle, and there is no such thing as a fluorescence after a fractional number of cycles as implied by the continuous functions [that the classical Cq approach involves]" (Jones et al. 2014). I agree with this point of view. One of the advantages of Pairwise Efficiency is that it is based on the analysis of actual fluorescence readings produced after each cycle, and does not rely on fractional cycles.

In Pairwise Efficiency, not only can we obtain more than 150 data points from a single dilution set (six wells), but replication of the calibration curve three times could potentially increase this number up to 2556 (72 fluorescence readings, all in cross-pairwise relationships). This allows the use of powerful statistical instruments, and represents a marked advantage over other methods.

Overall, the new method, Pairwise Efficiency, allows a nearly two-fold increase in the precision of efficiency estimation and a 2.3-fold increase in the precision of the gene ratio estimation. Thus, I have successfully devised a new method for qPCR data analysis that 1) does not require the use of reference genes, 2) increases the precision of measurements, and 3) reduces labor and cost associated with qPCR experimentation (for further discussion of this last point see General Discussion at the end of this thesis).

Chapter 3

Application of Pairwise Efficiency to the analysis of gene expression dynamics during the iPS reprogramming process

3.1 Introduction

After I developed a new, high-precision RT-qPCR method, Pairwise Efficiency, I set out to demonstrate the effectiveness of the method in revealing biological information in iPS reprogramming as a model system of a long-term dynamic process. I chose to test the new method on a set of selected genes in a time-course of iPS reprogramming. My goals for this part of my work were as follows:

1) Test the applicability of Pairwise Efficiency in medium-throughput setting
2) Reveal the patterns of gene expression during iPS reprogramming
3) Determine whether Pairwise Efficiency can produce significantly better results than previously applied methods.

Previous works in the field of quantification of gene expression in stem cells have centered on mouse embryonic stem cells (ESCs), and have been mostly conducted in the low-throughput settings. For example, a study by E. Willems and colleagues in 2006 has identified Actin Beta and Gapdh as the most stable genes in mouse embryos and in differentiating mouse and human ES cells (Willems et al. 2006) among 10 genes. This study applied classical tools geNorm and NormFinder to draw a ranking on the genes, similarly to what I did in iPS cells in Chapter 1, and did not focus on time-course or range analysis. A subsequent study by S. Mamo in 2007 focused on 12 reference genes in mouse oocytes and embryos, and pointed out the instability of housekeeping genes, while suggesting that Ppia, H2afz and Hprt I as the most stable genes in the embryos (Mamo et al. 2007). The same team has later published a 2008 study in rabbit oocytes and preimplantation stage embryos. The team identified H2afz, Hprt I and Ywhaz as the most stable reference genes, while indicating that Ubc, Tbp and B2m were the least stable and unsuitable for normalization in qPCR experiments in pluripotent stem cells (Mamo

et al. 2008). A 2011 study subsequently analyzed mouse embryonic stem cells and found that Sdha, Tbp and Ywhaz were the most stable genes during the differentiation of embryonic stem cells *in vitro* (Veazey and Golding 2011). Another study, performed in human embryonic stem cells in 2013, on the other hand, identified B2m and Rpl13a as the most stable genes during differentiation (Vossaert et al. 2013). A large-scale study conducted in human ESCs was done by Synnergren and colleagues (Synnergren et al. 2007). This study used microarrays to analyse housekeeping gene dynamics, and has also arrived at the conclusion that conventionally used housekeeping genes, such as HPRT, Actin, Gapdh, fluctuate in differentiating human ESCs. This study has identified a special set of housekeeping genes for use as a reference in pluripotent stem cell experiments. The next largest study, judging by scale, was done in 2015 by the same team and again in differentiating human ESCs, and has expanded the set of analyzed genes because of rising concerns about variability of the housekeeping genes. This study used large-scale datasets to perform global transcriptional analysis. It included 9 different datasets and 144 microarray to identify a set of non-varying genes, while highlighting the fact that commonly used Actb, Gapdh, Hprt1, Ppia, Sdha and B2m varied substantially during human ESC differentiation. This study put Hprt1 and B2m in the group of highly varied genes (Holmgren et al. 2015).

Only one study so far has been conducted on iPS cells, where the team investigated the stability of commonly used housekeeping genes during iPS differentiation (not reprogramming). This small-scale study on 16 genes has identified Actb, C1orf43, PSMB4, Gapdh and HMBS as the most stable genes during iPS differentiation. There are no studies on housekeeping genes during the iPS reprogramming process, and the existing studies are limited in their scale. Thus, the stability of common housekeeping genes and the newly suggested housekeeping genes (Synnergren et al. 2007; Holmgren et al. 2015), as well as and their performance compared to conventional housekeeping genes, still needs to be investigated in iPS systems, especially during the reprogramming process.

In this chapter, I **demonstrate the application** of the developed high-precision qPCR method, Pairwise Efficiency, to uncover gene expression patterns in 70 housekeeping genes during the iPS reprogramming. I have included a portion of

recently discovered genes, mentioned above in relation to large-scale studies, as well as conventionally used genes, such as Gapdh, ActB and ribosomal genes.

3.2 Materials and Methods

Cell culture

The iPS reprogramming was carried out in a reprogrammable cell system previously described in Hikichi et. al., 2012, and applied in Panina et.al., 2018. The details of the experimental procedures are described in Materials and Methods section for Chapter 1. In short, neural progenitor cells were cultured in a suitable medium and iPS reprogramming was carried out without cell split for 20 days. Samples were collected on Days 0, 5, 10 and 15 for this experiment.

RNA isolation and cDNA synthesis

The RNA isolation and DNA synthesis was carried out as described in Materials and Methods section for Chapter 1 and Chapter 2.

Experiment design and PCR dataset generation

The experiment was designed according to the requirements of the newly developed Pairwise Efficiency method. Since four time points during the reprogramming process were to be analyzed, I divided the 96-well PCR plate into four parts. Every part out of the four could host four target genes for analysis. Thus, 70 genes analyzed resulted in 17.5 plates. No other replicas were necessary because Pairwise Efficiency allows the analysis based on six-well approach for each gene-DNA combination (for pipetting layout principle see Fig. 19). The Baseline Subtracted PCR datasets were generated from each PCR run and processed using Bio-Rad CFX Manager 2.0 (2.0.885.0923). These datasets were imported for analysis into the Pairwise Efficiency software (unpublished).

Quantitative real-time PCR

RT-qPCR was performed with a CFX96 Connect apparatus (BioRad) and the reagents as described in Materials and Methods section for Chapter 1 and Chapter 2.

Results

3.3 The preparation of cell samples and the standard pluripotency check

To monitor gene expression during the iPS reprogramming process, I prepared mouse neural progenitor cells and reprogrammed them to pluripotency as described in the Materials and Methods. For gene expression measurements cells were harvested at Day 0, Day 5, Day 10 and Day 15, and the completion of reprogramming by Day 15 was confirmed by alkaline phosphatase staining and immunostaining for pluripotency markers Nanog and Oct4 (Fig. 14). While the alkaline phosphatase has started to appear, albeit very little, at Day 5 (Fig. 14A), the pluripotency markers Nanog and Oct4 were expressed on the last day of the reprogramming (Fig. 14B) but could not yet be detected by immunofluorescence on Day 10 (data not shown). The levels of pluripotency markers Nanog and Oct4 in RNA samples was also confirmed by qPCR in the process of gene expression analysis (see the following sections of the Results).

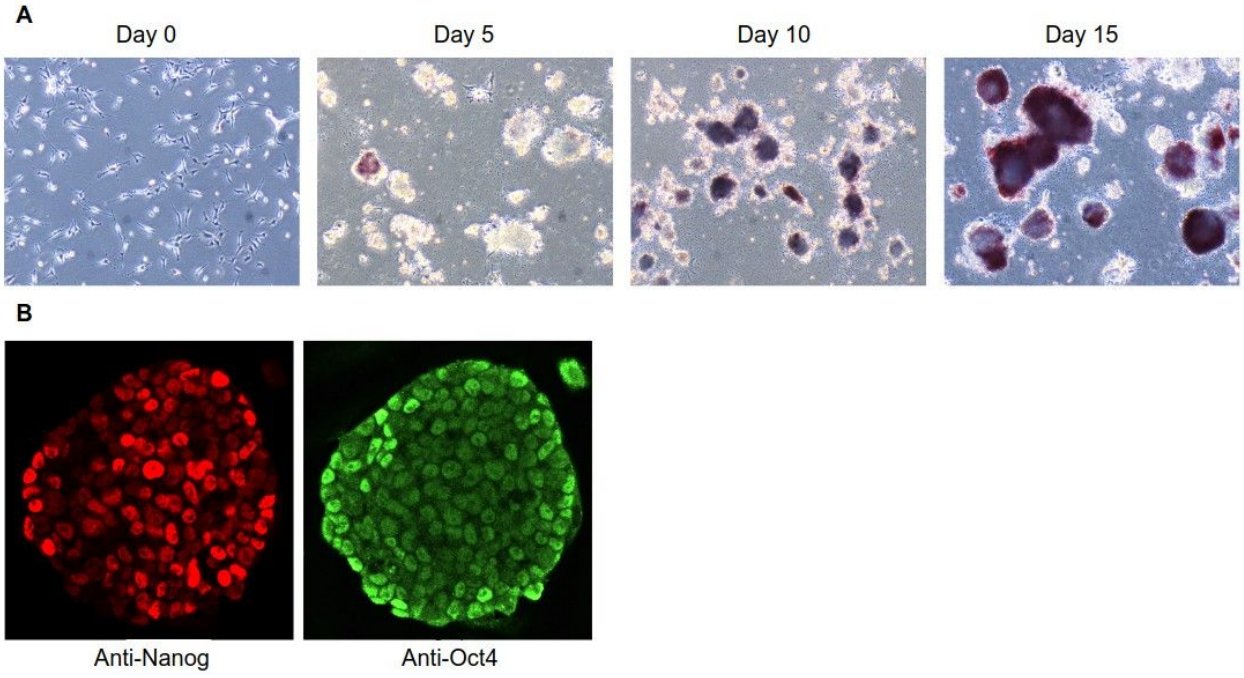


Figure 14. Standard tests confirming successful reprogramming of neural progenitors into the pluripotent iPS cells. A. Alkaline phosphatase staining of the neural progenitor cell line at different days throughout the iPS reprogramming process. Nascent colonies possessing expressed alkaline phosphatase are depicted in dark red. **B.** The results of the immunofluorescent analysis of nascent colonies for pluripotency markers Nanog and Oct4 at the end of the reprogramming (Day 15).

3.4 The choice of housekeeping genes for the investigation of iPS reprogramming

Next, using previously published articles, I assembled a list of 70 housekeeping genes for expression pattern analysis (Table 7). The list contained a portion of commonly used genes and the genes recently identified by microarray analysis as stable and suitable for qPCR normalization in differentiating human embryonic stem cells (Synnergren et al. 2007; Holmgren et al. 2015).

Gene	ID	Function
Aasdh	NM_173765.3	Unknown function, possibly post-translational modification
Actb	NM_007393.5	Cytoskeleton
Ada	NM_001272052.1	Purine metabolism, possibly immune
Alas1	NM_001291835.1	Mitochondrial, porphyrin metabolism
Alb	NM_009654.4	Blood serum albumin
Atp5f1	NM_009725.4	ATP generation
B2m	NM_009735.3	Immune system
Car6	NM_009802.2	Carbonic anhydrase, only in salivary glands
Cdc14a	NM_001080818.2	Centrosome separation, cytokinesis
Cox4i1	NM_009941.3	Electron transport chain
Cpne2	NM_153507.2	Calcium-mediated intracellular processes

Crebbp	NM_001025432.1	Acetylates histones and other proteins
Cript	NM_019936.3	PDZ-binding protein
Def8	NM_001253783.1	Lysosome peripheral distribution, possibly bone resorption
Dtwd2	NM_026854.3	Unknown function (location only cervix)
Eef1d	NM_029663.2	Translation
Eln	NM_007925.4	Cytoskeleton
Fbxl12	NM_013911.3	Ubiquitin ligase component
Fh1	NM_010209.2	ATP (TCA cycle)
Foxp4	NM_001110824.1	Transcriptional repressor that represses lung-specific expression, possibly other?
G6pdx	NM_008062.2	Oxidative pentose-phosphate pathway (addition to glycolysis)
Gapdh	NM_001289726.1	ATP generation
Got1	NM_010324.2	ATP generation
Gtf2h3	NM_181410.3	Transcription
Gusb	NM_010368.2	Digestion system
H13	NM_001159551.1	Immune system
Hddc2	NM_027168.2	Unknown function (only in brain)
Hmbs	NM_001110251.1	Heme biosynthesis
Hprt	NM_013556.2	Nucleotide salvage
Idh3a	NM_029573.2	ATP generation
Kiaa0141	NM_024179.5	Apoptosis
Ldha	NM_001136069.2	ATP generation
Mdh1	NM_001316675.1	ATP generation
Mlh3	NM_001304475.1	mutL
Mpi	NM_025837.2	Glycosylation
Nubp1	NM_011955.2	Assembly of an Fe-S cluster, centrosome duplication, negative regulator of cilia
Pdha1	NM_008810.3	ATP generation

Pfkp	NM_001291071.1	ATP generation
Pgam1	NM_023418.2	ATP generation
Pgk1	NM_008828.3	ATP generation
Plekha1	NM_001346515.1	Membrane signaling
Pole	NM_011132.2	DNA polymerase
Ppia	NM_008907.1	Possibly immunosuppression
Pten	NM_008960.2	Stopping cell cycle
Ripk3	NM_001164107.1	Signaling, apoptosis
Rnd1	NM_172612.3	Cytoskeleton (Rho GTPase)
Rnf7	NM_011279.3	Component of ubiquitin ligase
Rpl13a	NM_009438.5	ribosome
RPL15	NM_001359897.1	ribosome
Rpl7	NM_011291.5	ribosome
Rplp1	NM_018853.3	ribosome
Rps11	NM_013725.4	ribosome
Rps18	NM_011296.2	ribosome
Rps3	NM_012052.2	ribosome
Rps9	NM_029767.2	ribosome
Sdha	NM_023281.1	Electron transport chain
Slc4a1ap	NM_001347328.1	Kanadaptin in h., mRNA export
Slc5a11	NM_146198.2	Sodium-depend. transport across membranes
Snrpb	NM_009225.2	Pre-mRNA splicing
Srp72	NM_025691.1	ER protein traffic
Srsf7	NM_001195485.1	Pre-mRNA splicing
Stim1	NM_009287.4	Ca2+ entry, Ca2+ sensor in ER
Tbp	NM_013684.3	Transcription
Tfrc	NM_001357298.1	Iron uptake
Tmem41b	NM_153525.5	Motor neuron development, autophagy
Tubb5	NM_011655.5	Cytoskeleton

Ubc	NM_019639.4	Ubiquitin
Vim	NM_011701.4	Cytoskeleton
Vsnl1	NM_012038.4	Insulin secretion
Ywhaz	NM_001253805.1	Signaling

Table 7. 70 housekeeping genes assembled from literature. The list includes commonly used housekeeping genes as well as newly suggested genes from the works on embryonic stem cells. The description of the genes' function is based on Gene Ontology database.

3.5 Housekeeping genes' expression dynamics during the iPS reprogramming

I then performed qPCR of these genes at all four time points using Pairwise Efficiency approach. First, I have found that out of 70 genes, 8 were not expressed in neural progenitors or reprogrammed iPS cells. Even though their expression was reported in the previous works by Synnergren group among newly suggested genes, my Gene Ontology analysis revealed that they, in fact, are very unlikely to be expressed in pluripotent stem cells. For example, according to NCBI expression and Gene Ontology data, carbonic anhydrase Car6 is only found in salivary glands, the Alb gene coding for blood serum albumin would naturally be found only in samples containing blood, etc., and thus the absence of these genes from iPS cells in my experiment is rather logical, or least not implausible. The discrepancy with the Synnergren's data may be due to lower precision of the methods they used for identification of these genes (please also note that Synnergren group did not perform Gene Ontology analysis to check for gene function).

The changes in expression levels of the remaining 62 genes are shown in Figure 15. For clarity, the initial analysis on gene expression change was conducted as follows. The fold-change of gene expression during the reprogramming process was estimated and assigned to one of the three groups: 1) The expression level equal to that of Day 0 (i.e. no change compared to non-reprogrammed cells), 2) the expression level is at least 1.5 times greater (because the expression change under 1.5 times fold can potentially be attributed to random error in measurements), but

less than 3-times fold, and 3) gene expression change of more than 3-times fold (Fig. 15).

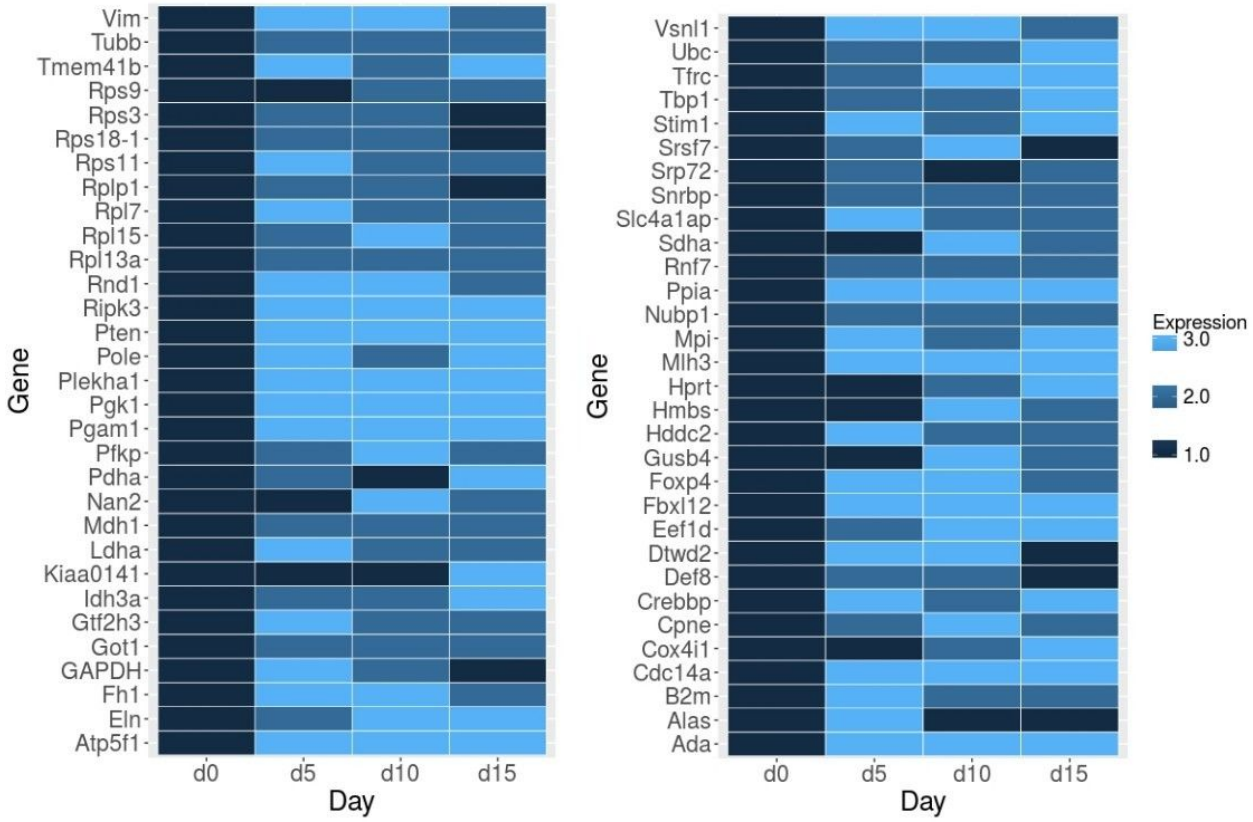


Figure 15. A heatmap representing overall changes in 62 housekeeping genes' expression throughout the iPS reprogramming process. Four time points during the reprogramming are indicated under the heatmaps, and the gene names are written on the left. The dark blue color represents no expression change, and Day 0 is taken as the "time point zero" before initiation of reprogramming. Thus, all tiles on Day 0 are colored dark blue. Blue color represents the expression change of more than statistically significant 1.5-fold, but less than 3-fold. Light blue represents gene expression change greater than 3-fold, compared to time point zero (Day 0).

I have found that, out of 62 genes, no genes showed constant expression levels (that would be shown as dark blue on all four time points) during the iPS reprogramming. Moreover, 80% of all genes displayed changes in gene expression levels that were more than 3-times fold at least on one of the days of

reprogramming (light blue tiles). Only 20% of all genes stayed within the 3-times fold change in expression levels (medium blue tiles), and fluctuated more than 1.5-times fold, but not greater than 3-times.

3.6 Clustering analysis of the gene expression dynamics during the iPS reprogramming

To reveal patterns in gene expression change, I applied standard clustering approach (Euclidean distance analysis) based on the expression data. The rising and falling levels of the genes were visualized by the dual-color heatmap tiling, where green represented rising levels, and red represented falling levels (Figure 16).

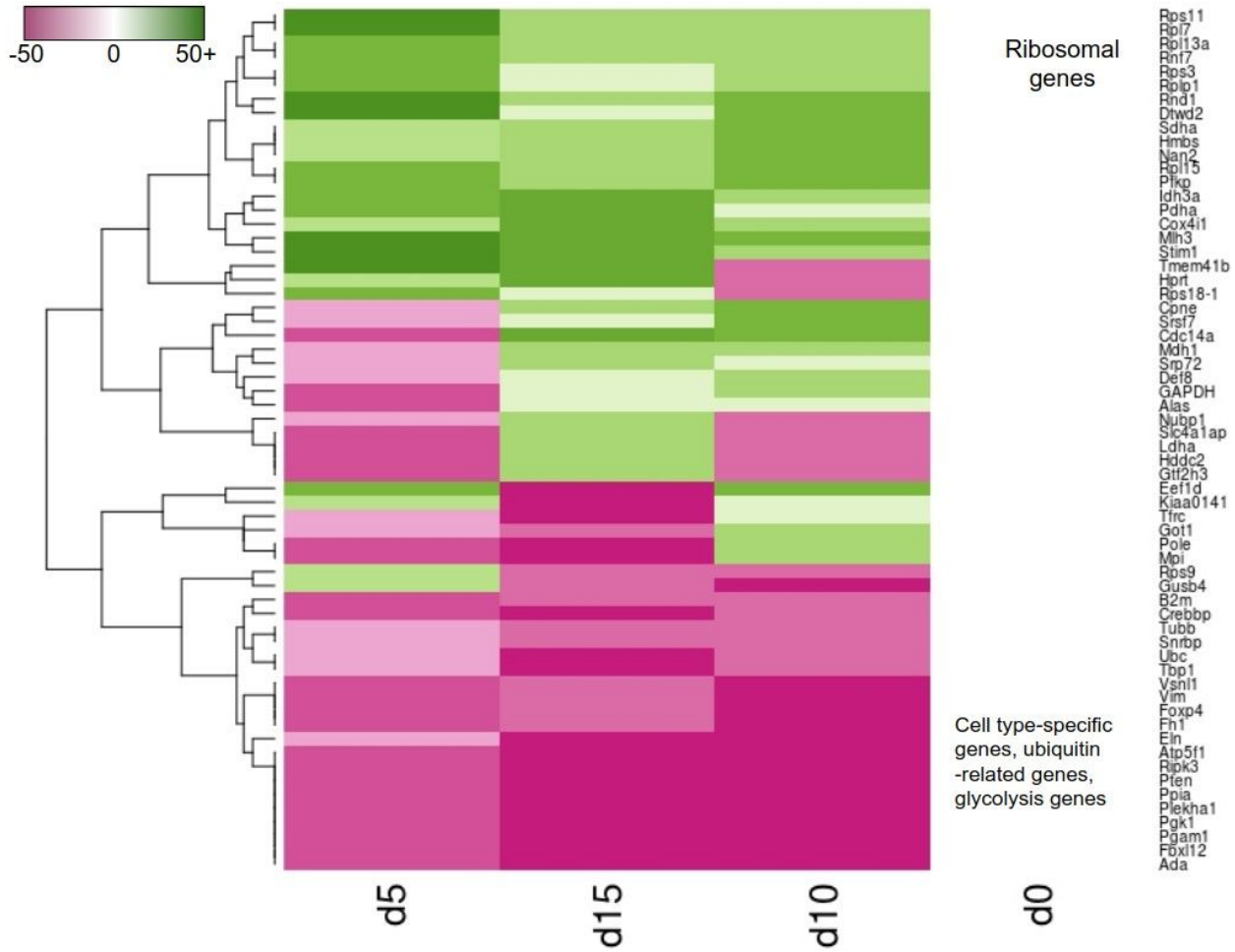


Figure 16. A heatmap with the application of standard clustering analysis, Euclidean distance measure, representing the grouping of 62 housekeeping

genes by their expression change throughout the iPS reprogramming process. Four time points during the reprogramming are indicated under the heatmap, and the gene names are shown on the right. The green color represents positive expression change (increase in the expression), and red color represents negative expression change (decrease). White color represents no change in expression, and all tiles on Day 0 are colored white. The standard dendrogram on the left shows the grouping of the genes by their similarity of expression dynamics.

First of all, the clustering analysis revealed that the biggest change in genes' levels occurred early in the reprogramming process, represented in my experimental system by Day 5. On that day, the biggest variation in expression level was found. These expression changes subsided at later stages (Day 10 and Day 15), and these two days were found to be more similar to each other than to the Day 5. The dark green color (which indicates greater change compared to point zero than the light green color) appeared only on Day 5 in all genes analyzed, and the green became lighter as the reprogramming progressed. This indicates that the positive gene expression change is greatest at the earlier stages of reprogramming, and is flattened out closer to the end of the process. On the other hand, the negative gene expression change is greatest (judging by fold-change) in the later stages of the reprogramming, because the red-colored tiles become darker as the reprogramming progresses to its final stages represented by Day 15.

Second, the algorithm revealed high similarity between gene expression pattern of all ribosomal genes, which were grouped together by the algorithm. The ribosomal genes Rps11, Rpl7, Rpl13a, Rps3, Rplp1 and Rpl15 were on the far end of the rising spectrum, and the biggest rise in these genes' level occurred early in the reprogramming, represented in this experiment by Day 5. On the other, falling side of the spectrum, I have found cell type-specific genes such as Vimentin, Elastin, Pten, Plekha1 and Ada. In addition, ubiquitin-related genes Ubc and Fbxl12, and glycolysis-related genes Pgk1 and Pgam1 were also grouped at the low end of the spectrum. For the interpretation of these data see Discussion.

3.7 Analysis of the overall tendency of selected genes to fluctuate during the iPS reprogramming

Finally, to analyze the tendency of selected genes to rise or decline throughout the whole reprogramming process, I have counted the number of rising and declining genes on each day of the reprogramming and plotted these numbers against the day of reprogramming (Fig. 17).

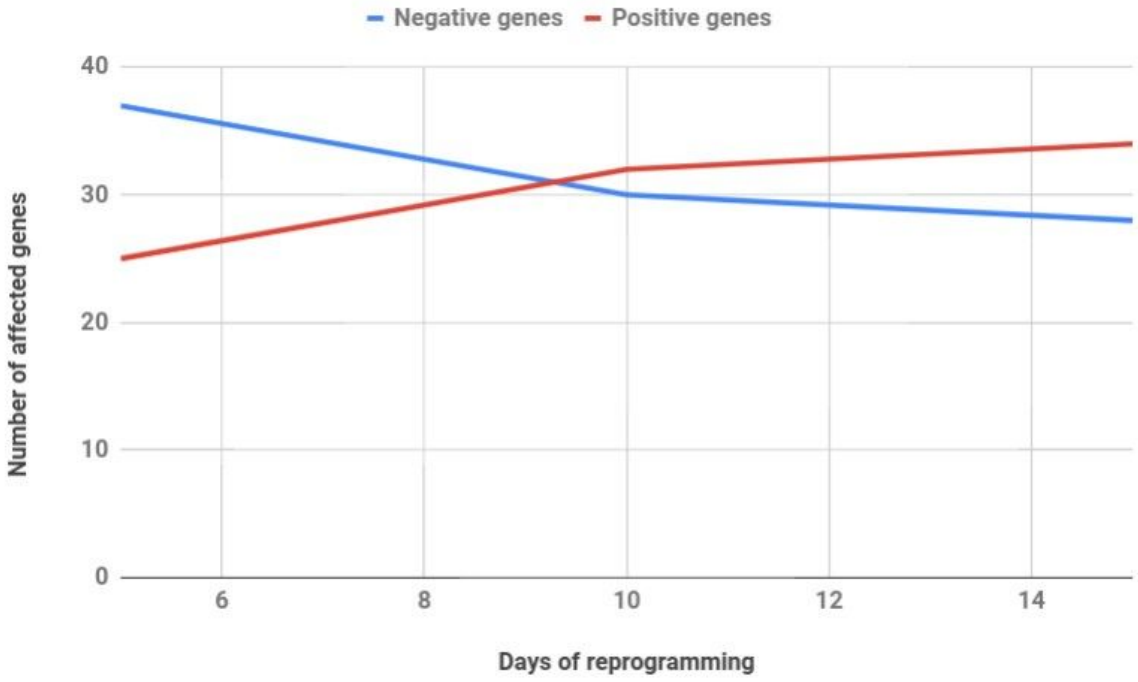


Figure 17. The tendency of selected genes to fluctuate throughout the whole reprogramming process. Red line represents the number of positively affected genes, blue line indicates the number of negatively affected genes. Day 0 is excluded from the graph because the value of each gene on that day is taken as a reference and is compared against. The number of positively affected genes continues to rise during the reprogramming, while the number of negatively affected genes is falling approximately after day 6, if we take into account the inherent standard deviation.

This analysis has shown that, first of all, in the group of selected housekeeping genes, there is a tendency to downregulate gene expression in the beginning of reprogramming. This tendency is greatest in the beginning of the process, and reaches its peak approximately on Day 5. As the reprogramming progresses, this tendency is diminished, and the number of downregulated genes decreases with time. On the other hand, the number of positively regulated housekeeping genes whose expression is rising during the reprogramming, is constantly growing, with the lowest point falling on Day 5, and continuing to rise further into the reprogramming process. These data are in agreement with the heatmap clustering analysis (Fig. 16) that shows the number of green tiles, corresponding to positively regulated genes, rising, while the number of red tiles is falling.

3.8 Comparison of the results for 10 housekeeping genes obtained in Chapter 1 with the results obtained in Chapter 3

As an additional way to validate the new Pairwise Efficiency method I performed the comparison of general tendencies for 10 housekeeping genes that were analyzed both by standard method (in Chapter 1) and by Pairwise Efficiency (Chapter 3). Theoretically, if the new method were to improve precision of measurements, the general tendency ("increase" or "decrease") should be preserved; in other words, the differences in gene expression analysis should be quantitative rather than qualitative.

Gene	Tendency by Classical	Tendency by Pairwise Efficiency
Atp5f1	Decrease	Decrease
B2m	Unclear	Fluctuate
Gapdh	Fluctuate	Fluctuate
Gusb	Unclear	Fluctuate
Hprt	Increase	Increase
Pgk1	Decrease	Decrease

Ppia	Decrease	Decrease
Rps18	Increase	Increase
Tbp	Unclear	Fluctuate
Tfrc	Unclear	Fluctuate
Tendency Match (%)		100%

Table 8. Comparison of the general tendencies of the gene expression change results obtained by the standard calibration curve method (Chapter 1) and Pairwise Efficiency method (Chapter 3). The general tendency for 10 genes was recorded in the table, using either "decrease", "increase" or "fluctuate" (meaning that on some days the expression increased, and on some days it decreased). The label "unclear" indicates that the linear fit applied to the genes in Chapter 1 could not identify a clear tendency due to large dispersion of obtained Ct data. As shown in the table, the data obtained by both methods match in 100% of the cases *by tendency*.

The analysis revealed that 60% of the genes measured by the standard method had a clear tendency to either decrease or increase, according to linear fit performed in Chapter 1, and 40% of the genes did not have a clear tendency due to large dispersion of Ct values which was reflected in a horizontal line in the linear fit analysis. Such cases were designated as "unclear" (because a linear fit would not identify fluctuation if it existed). In case of Pairwise Efficiency, the gene expression tendency was labelled as "increase" (in cases where all of the days of reprogramming had an increase in gene expression), "decrease" (when the gene expression decreased on all days) and "fluctuate" (in cases when on some days the expression increased, while on other days it decreased). The analysis in the Table 8 shows that the "fluctuation" tendency matched with the "unclear" tendency in all of the cases, and in all cases the "decrease" and "increase" tendencies matched. This comparison is not quantitative but rather qualitative because linear fit cannot be directly quantitatively compared with euclidean distance clustering analysis.

To further demonstrate that the results in Chapter 1 correspond well to the results in Chapter 3, I plotted the Ct values of eight genes for direct comparison of

tendencies (Fig. 18). This analysis has shown that the changes in gene expression in Ct values were very similar, and the characteristic drops and rises were conserved, as expected. For example, a characteristic rise in Ct value for B2m gene between Day 0 and Day 5 was observed in both cases, a significant drop in Ct values for Pgk1 gene on Day 10 was observed in both cases, the rise on Day 5 and drop on Day 10 for Ppia gene was observed in both cases, etc. It is necessary to remember that Ct values themselves are NOT COMPARABLE between different experiments (this is another disadvantage of the standard qPCR method), because Ct threshold is set **separately for each experiment** by qPCR machine.

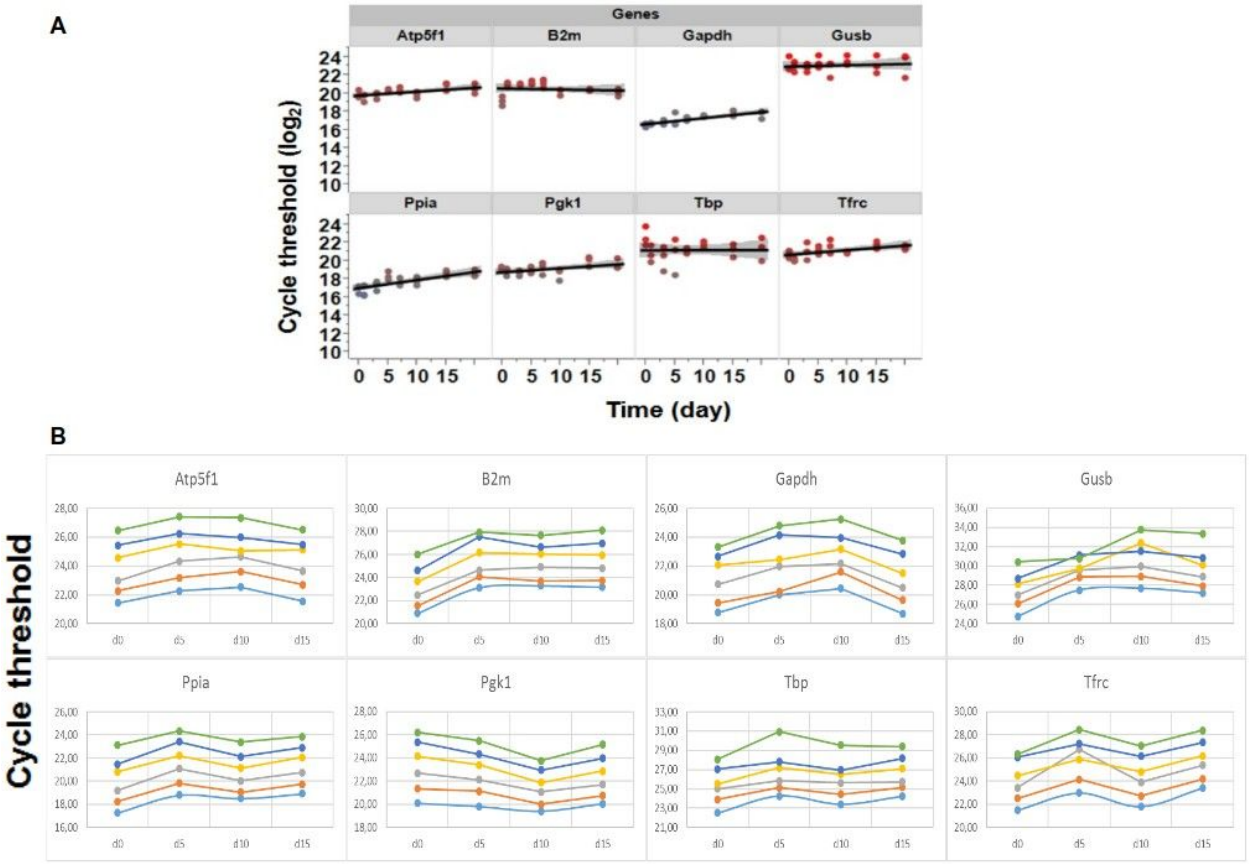


Fig. 18. A direct comparison by Ct values of the results obtained in Chapter 1 and the results obtained in Chapter 3. A. An excerpt from the linear fit analysis conducted in Chapter 1 for eight genes. In this case, the cells were collected on days 0, 1, 3, 5, 7, 10, 15 and 20, and each day had three technical replicas, all of which is depicted in dotted circles. **B.** Ct values for eight genes taken from reprogramming experiment in Chapter 3. In this case, the samples were collected only on days 0, 5, 10 and 15. Each day has six dots because it was a

six-step serial dilution. The drops and rises of the genes are similar in part A and part B for the days that can be compared (0, 5, 10, 15).

3.9 Discussion

In this chapter, I have applied the newly developed method, Pairwise Efficiency, to the measurement and analysis of gene expression change throughout the iPS reprogramming process. The selection of four time points, which resulted in four separate DNA samples in the RT-qPCR experiment, and 70 housekeeping genes amounts to 280 separate DNA-primer combinations. Since six replicas were run for each combination, as required by Pairwise Efficiency method, a total of 1680 amplification curves were obtained, making it a medium-throughput analysis. This showcases successful application of Pairwise Efficiency method in medium-throughput settings.

Moreover, the analysis of RT-qPCR data produced by Pairwise Efficiency, could reveal previously unnoticed patterns in gene expression change during the reprogramming process.

First of all, the heatmap clustering analysis has detected a surge in housekeeping gene expression change early in the reprogramming process which was represented by Day 5, that subsided at later stages, represented by Day 10 and Day 15. The clustering results are supported by the analysis of changes in affected gene numbers. According to this analysis, the beginning stage of the reprogramming is associated with a surge in negative gene regulation, which subsides at later stages, and positive gene regulation, the growth of which is also diminished with time. Thus, it can be said that housekeeping genes are affected by iPS reprogramming **mostly at the beginning of the process**. This fact was previously unknown, as the reprogramming was seen as a continuous, uniform process that gradually progressed from “non-pluripotency state” to “pluripotency state” (Buganim et al. 2013). There is no research that I am aware of that points out at non-uniform patterns of change in gene expression during the reprogramming, especially for housekeeping genes.

Second, the applied clustering algorithm has detected high similarity in the rise of ribosomal genes’ levels, and placed these genes grouped together on the positive end of the spectrum. The rise in the expression of ribosomal genes, and the

similarity was previously unknown, as the ribosomal genes were assumed to be housekeeping and constantly expressed in all cell types. However, the rise in the ribosomal genes' expression would logically follow from the concept of iPS reprogramming that assumes the increase in cell viability and cell cycle progression.

Third, the algorithm has placed cell type-specific genes on the negative side of the spectrum. For example, such genes as the Elastin (Eln) and Vimentin (Vim) which are components of the cytoskeleton specific to certain cell types or states (particularly, Vimentin is implicated in epithelial-to-mesenchymal transition), Plekha1 (pleckstrin homology domain-containing gene related to membrane signaling), Pten (tumor suppressor), Ripk3 (receptor-interacting serine-threonine kinase implicated in apoptosis), Ada (immune system-related gene) and others were found to be downregulated. In addition, ubiquitin system-related genes Ubc and Fbxl12 were also found on the negative side of the spectrum, and their expression decreased. It is worth noting that many of these genes are considered housekeeping, however, their downregulation also logically follows from the reprogramming concept. For example, tumor suppressor Pten inhibits cell cycle progression, and its downregulation should be vital to the progression of reprogramming. Similarly, the ubiquitin system and apoptosis genes can also be expected to be down-regulated during the reprogramming process, especially at the initial stages of forced expression of pluripotency-related transcription factors. Previous works (such as Synnergren et al. 2007) have failed to notice these patterns presumably because their experimental tools lacked the precision of Pairwise Efficiency.

I further discuss the implications of these findings in the General Discussion.

General Discussion

The purpose of this work was to 1) investigate the suitability of commonly used tools for gene expression quantification with high-precision in long-term processes; 2) to develop a high-precision tools if deemed necessary; and 3) to apply the said tool in a system representative of a long-term dynamic process. In the first chapter, I conducted basic analysis of the expression of 12 housekeeping genes by the conventional qPCR method, as required in MIQE guidelines. This analysis has revealed high variability of housekeeping genes during the reprogramming process, if analyzed by conventional RT-qPCR. Thus, in the second chapter, I developed a new tool that 1) would not require the use of reference genes, 2) would be high-precision and high-throughput, and 3) would decrease the hands-on time for RT-qPCR experimentation as well as the cost of the experiments. In the third chapter, I have successfully applied this tool to the quantification of gene expression during the iPS reprogramming process and **revealed new, previously unknown patterns in gene expression** change during this process. All of this demonstrates the general usefulness and scientific advantage of my method, Pairwise Efficiency, over the existing methods of gene expression analysis. Below, I will discuss the implications of this work, and further possible directions of its development.

First of all, in the first chapter I have applied a common method of high-precision gene expression analysis, RT-qPCR, as described in MIQE guidelines (Bustin et al. 2009). Then, I have put the resulting Ct values to the statistical analysis using five previously established algorithms. I have found, that, according to this method, the expression of housekeeping genes varied throughout the reprogramming process more than 2-fold for all genes, and reached variation of 4-cycle-fold for some genes, notably the ribosomal gene Rps18. I have also found that a group of genes associated with ATP production (Atp5f1, Pgk1 and Gapdh) stood out among the 12 genes analyzed, and was deemed the most stable group by the algorithms applied in the analysis. The results of this chapter prompted me to

conclude that a new, more precise method without the use of reference genes was needed to properly analyze long-term dynamic processes.

In the second chapter, I have developed such method by the application of the principles of the Measurement Theory to the RT-qPCR data analysis. This new method, Pairwise Efficiency, has allowed me to improve the precision of measurements nearly two-told, while reducing labor and cost of experimentation. In addition, this new method does not require the use of reference genes. In this chapter, I have also conducted a comparative analysis of Pairwise Efficiency with the classical calibration curve method using common approaches previously described in literature. Notably, the procedure of classical calibration curve with the use of reference genes requires the researcher to determine PCR efficiency coefficients for each gene analyzed, including the reference genes (Bustin et al. 2009). Thus, the classical calibration curve method would require at least 42 wells to properly analyse gene expression, while Pairwise Efficiency would only require 12 wells. In addition, if the efficiency of PCR reaction for the reference gene and the target gene differ, the calibration curve approach requires a change of reference gene or primer because it is not applicable in cases where the efficiency is different. In contrast, my new method, Pairwise Efficiency, does not require the use of reference genes at all, and accounts for efficiency in every sample, thus **removing the need for new primer design** or the search for a new reference gene. In other words, while the classical calibration curve approach would require multiple runs of the RT-qPCR machine, **Pairwise Efficiency requires only one run** for each gene, and allows to correct for efficiency differences in the process of data analysis (Fig. 19). This opens wide possibilities for high-throughput use of Pairwise Efficiency, including automation of the experimental process.

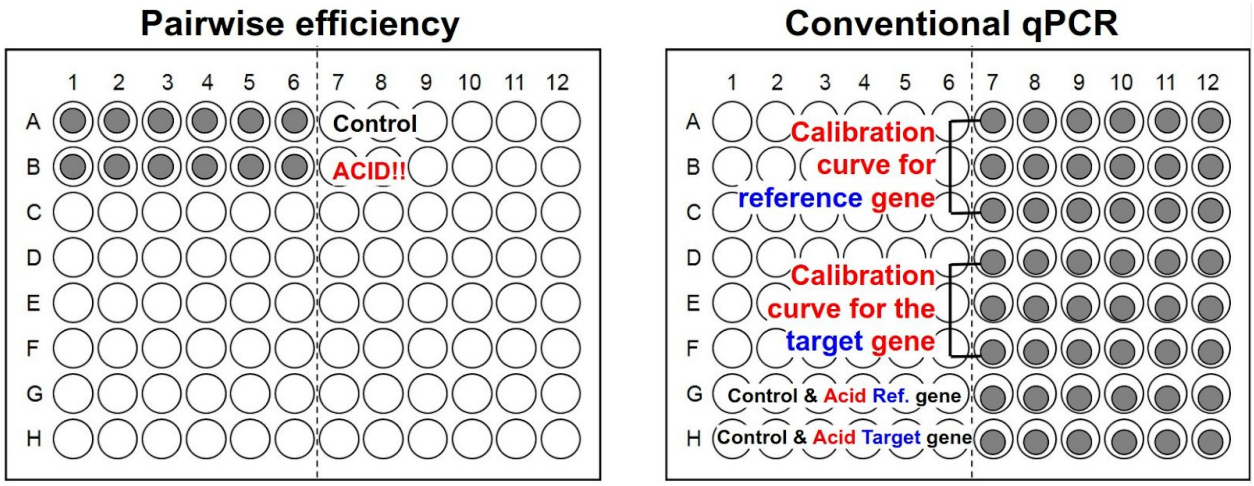


Figure 19. A schematic representation of the differences in the experimental procedure for Pairwise Efficiency approach compared to the standard widely used calibration curve method. Pairwise Efficiency requires only one run with 12 wells for the analysis of one target gene in two samples (untreated sample, labelled "Control" in the figure, and treated sample, labelled "ACID" or "Acid" in the figure). However, the standard calibration curve method requires to separately run the calibration curves for the target and reference genes (one run of the PCR machine with 36 wells), and THEN additionally run the actual experiment with the target gene, reference gene, and Control and Acid samples (second run of the PCR machine with 12 wells). Furthermore, since the precision of the classical approach is two times lower than the Pairwise Efficiency approach, it frequently requires additional runs to account for insufficient precision.

As the final part of my work, I have applied Pairwise Efficiency to the analysis of gene expression dynamics during the iPS reprogramming process. For the purpose of this analysis, I have searched the available literature and selected 70 housekeeping genes that have been previously implicated in the stem cell research (Synnergren et al. 2007; Holmgren et al. 2015). I have included both commonly used and newly suggested genes in my analysis. Pairwise Efficiency allowed me to **identify previously unnoticed patterns of gene expression** change during the iPS reprogramming process.

First of all, the clustering analysis of gene expression pattern indicated that the biggest change in gene expression occurs early in the reprogramming process. This fact was previously unknown. The iPS reprogramming was commonly seen as progressing through three stages, initiation, maturation and stabilization (David and Polo 2014). My analysis has shown that the gene expression patterns do not change equally, and the biggest change occurs early (represented in my system by Day 5), while the later phases show less expression change (Fig. 16). Since the immunostaining for pluripotency markers and the alkaline phosphatase expression analysis has shown no marker presence on Day 5, I conclude that the biggest change in the gene expression occurs prior to the appearance of the markers of pluripotency alkaline phosphatase, Nanog and Oct4.

Second, the clustering analysis of RT-qPCR data obtained by Pairwise Efficiency method revealed that ribosomal genes have similar expression change pattern, are grouped together, and exhibit positive expression change (ribosomal gene increase) overall. This result was previously unknown, but it is well in agreement with the concept of iPS reprogramming, where the change in pluripotency state from less pluripotent to more pluripotent is associated with the increase in cell cycle progression and speed (Ghule et al. 2011; Kapinas et al. 2013; Ruiz et al. 2011; Xu et al. 2013). The increase in cell cycle speed would logically be associated with the increased need for ribosomal RNA synthesis, and the ribosomal gene expression would go up, as was uncovered by my analysis. The confirmation of this fact, as well as further investigation (for example, the measurements of nucleoli size or modern RNA staining techniques) would shed more light on these cellular events.

Third, and most interestingly, the clustering analysis has grouped several ATP-generating genes into two categories: the genes whose expression decreases following a similar pattern ("decreasing genes"), and the genes whose expression fluctuates throughout the reprogramming. For example, such genes as Pgk1 and Pgam1 have been grouped together as decreasing, and placed next to each other on the clustering tree. Interestingly, Pgk1 and Pgam1 represent consequent steps in glycolysis, as depicted in Fig. 20. Pgk1 is a phosphoglycerate kinase that catalyses the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate, and

Pgam1 is a phosphoglycerate mutase that catalyzes the conversion of 3-phosphoglycerate to 2-phosphoglycerate (2-PGA) in the glycolytic pathway.

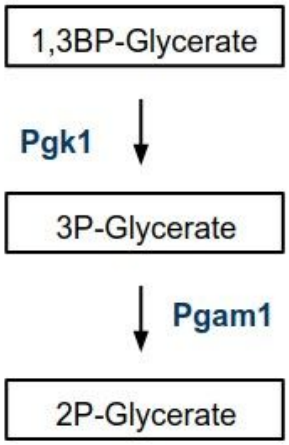


Figure 20. A scheme of the part of the glycolytic pathway, in which the enzymes Pgk1 and Pgam1 are the subsequent steps in the occuring set of reactions, and are situated next to each other in the pathway. It is interesting to see that the clustering analysis based on Pairwise Efficiency data has identified similarity in the pattern of their change during iPS reprogramming and placed these genes next to each other in the clustering tree.

On the other hand, Pfkp, Idh3a and Pdha1 were clustered next to each other on the other, "increasing" part of the clustering spectrum. **Pfkp** is a key regulatory enzyme in glycolysis, a phosphofructokinase that catalyzes the irreversible conversion of fructose-6-phosphate to fructose-1,6-bisphosphate, **Idh3a** is an isocitrate dehydrogenase which is a key enzyme in the tricarboxylic acid (TCA) cycle that takes place in the mitochondria, and **Pdha** is pyruvate dehydrogenase, a part of mitochondrial matrix multienzyme complex that provides the primary link between glycolysis and TCA cycle. Thus, the clustering of these three enzymes together on the increasing spectrum is extremely interesting and may indicate **a switch step** occurring between glycolysis, and oxidative phosphorylation in mitochondria. More research on the gene expression change that would include all

genes involved in ATP generation may shed more light on the regulation of these switches during iPS reprogramming.

It is also interesting to note that, while most ribosomal genes were found to decrease during the iPS reprogramming process, some ribosomal genes displayed a fluctuating pattern (Rps18 and Rps9). Both of these genes belong to the small ribosomal subunit. It is compelling to speculate that these two genes might play a distinctive role in regulating the production of ribosomes and by that, regulate the cell growth and division state. Indeed, recent research has also pointed at such possibility, simulating the behaviour of ribosomal genes and identifying them as a major speed-regulating hub for cell cycle progression (Lin and Amir 2018).

Since this analysis has only been done on a small portion of genes, namely so-called housekeeping genes, it cannot reveal the changes in expression patterns of other genes. However, some conclusions can be drawn about the genes analyzed. First of all, in case of housekeeping genes, there exists a strong tendency to downregulate some genes and upregulate others in the beginning of the reprogramming, and **the strength of down- or up-regulation reaches its peak approximately on days 5-7 of the reprogramming** in this cell system. After that, the number of negatively regulated genes starts to decline, and the speed at which the number of positively regulated genes rises slows down. According to the obtained data, the overall reprogramming process related to housekeeping genes is "finished" by day 10 (Fig. 17), after which day the tendencies of the genes to fluctuate flatten out. This conclusion is supported by the clustering analysis that has placed Day 10 and Day 15 as similar to each other, while Day 5 was found to be dissimilar in the gene expression pattern (Fig. 16). Moreover, another interesting finding is that the number of positively regulated genes continues to rise, and the number of negatively regulated genes rises sharply in the beginning of the reprogramming but starts falling after the "break point" that could be approximately placed around day 7. The existence of such a "break" in the course of the reprogramming is speculative, however, my multiple unpublished data on the speed of colony formation also indicates that fully formed colonies of truly round shape appear approximately between days 5-7 (Fig. 21).

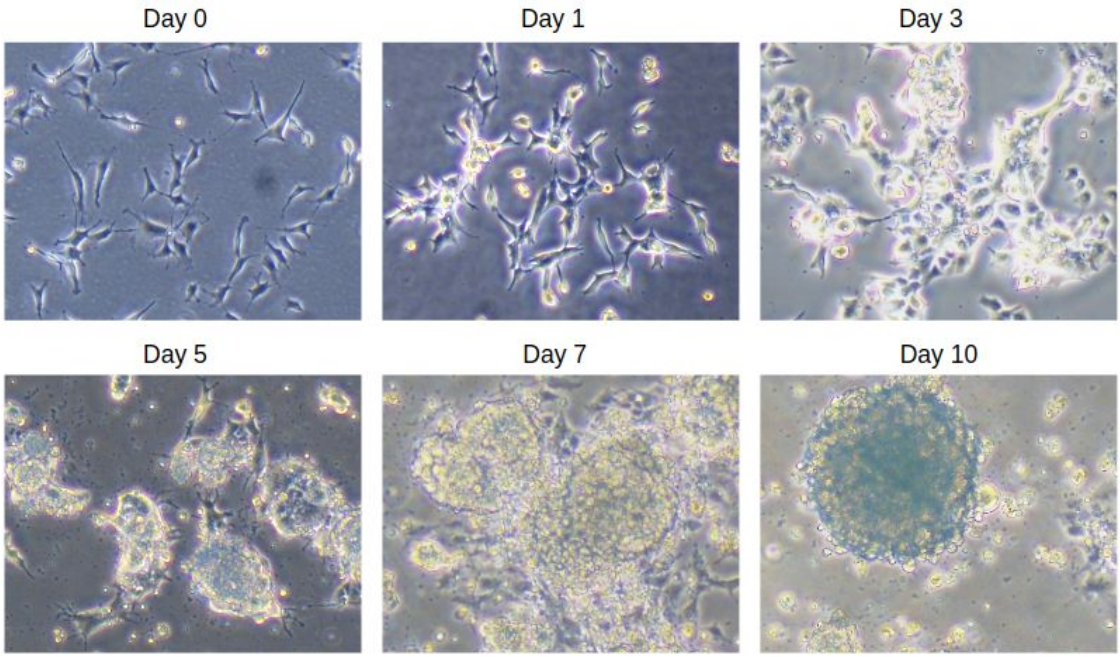


Figure 21. Colony formation during the iPS reprogramming of neural progenitors. The truly round-shaped colonies with a characteristic "glow" indicative of reprogrammed cells appear between days 5-7.

The gene expression data obtained in this work also **hints at the existence of an unknown event**, after which the tendencies of genes to fluctuate flatten out, and this event, according to the data, happens between days 5-8, prior to the expression of pluripotency markers (Fig. 14). More research that would include other days of the reprogramming process, and other gene groups, is needed to draw further conclusions and confirm the existence of the said event and reveal more information about it.

Finally, the obtained data on the number of positively and negatively regulated genes shows overall increase in rising genes, while the number of falling genes first increases sharply at Day 5, and then decreases back. This sharp increase and subsequent decrease in the number of down-regulated genes could depend on the selection of genes. However, another possibility is that some genes that were downregulated in the beginning, "want to return back", as if the process has bounced back. Among such genes we can name glutamic-oxaloacetic transaminase Got1, which is a part of glycolysis pathway, DNA polymerase epsilon Pole, which

is a central catalytic subunit of DNA polymerase, or mannose phosphate isomerase Mpi, a part of glycosylation pathway. It is interesting to speculate that there may be an existence of a negative feedback loop, an inherent tendency of the cell to preserve general homeostasis, i.e. to make genes return to the balanced state.

The comparison of the results obtained in Chapter 1 and Chapter 3 showed that the general tendencies of gene expression match, as expected. It is worth noting that the classical qPCR method gives the result in Ct values (cycle numbers). It is necessary to take into account that the growth of Ct indicates a *decrease* in the amount of target gene in a sample, and a drop in Ct indicates an *increase* in the amount of the gene in the sample. The results for the reference genes obtained in Chapter 1 correspond to the results obtained in Chapter 3 of the study not only in terms of general trends, but also in terms of the oscillations happening on different days. Moreover, both chapters' results suggest that the main changes occur at the initial stage of the process, which is also confirmed by the results of processing the initial data using the Euclidean distance measure (clustering) method.

Overall, the application of Pairwise Efficiency has uncovered several previously unnoticed patterns of gene expression during the iPS reprogramming that were impossible to identify by the existing methods of gene quantification. This shows that Pairwise Efficiency is superior in both precision and throughput to the existing data analysis approaches, and has a great potential as a new method of high-precision gene expression measurements.

Thus, I have successfully accomplished the goals of this work by developing a new high-precision of method of gene expression measurements and successfully applying it to a representative long-term process, the iPS reprogramming.

Supplementary Information

Supplementary Tables

Cell type	Gene	Pair #	Primer sequence	Slope	E (%)	R^2
Fully reprogrammed iPS cells	Gusb	1	F: AACAACACACTGACCCCTCA R: ACCACAGATCGATGCAGTCC	1.054	93	0.953
	Gusb	2	F: TGGCTGGGTGTGGTATGAAC R: GGTGACCTCCCTCATGTTCC	0.722	161	0.931
	Gusb	3	F: GGTGGAACATGAGGGAGGTC R: AGGGTATGAGGGGTCAGTGT	0.926	111	0.983
	Gusb	4	F: GGCCTCTAGATAGCCTTGAGC R: ACACGCACTCCATTTTAGGGA	0.342	660	0.736
	Hprt	1	F: GTTGGGCTTACCTCACTGCT R: TAATCACGACGCTGGGACTG	0.526	274	0.923
	Hprt	2	F: GATCAGTCAACGGGGGACAT R: GGTCCTTTTCACCAGCAAGC	0.529	271	0.886
	Hprt	3	F: ACAGGCCAGACTTTGTTGGA R: ACTTGCGCTCATCTTAGGCT	0.699	170	0.962
	Hprt	4	F: CAGTCCCAGCGTCGTGATTA R: TGGCCTCCCATCTCCTTCAT	1.01	99	0.995
	Tfrc	1	F: AAACTGGCTGAAACGGAGGA R: AGATCCAGCCTCACGAGGAG	1.157	82	0.997
	Tfrc	2	F: AAGAGCTGCTGCAGAAAAGC R: ACGGTCTGGTTCCTCATAACC	1.049	94	0.997
	Tfrc	3	F: GTTCGTACAGCAGCGGAAGT R: GGAAGTAGTCTCCACGAGCG	1.073	91	0.985
	Tfrc	4	F: AGCAAAGTCTGGCGAGATGAA R: CCACATAACCCTCGGGAGAC	1.189	79	0.989
Parental	Gusb	1	F: AACAACACACTGACCCCTCA	-2.39	163	0.991

cells (N31)			R: ACCACAGATCGATGCAGTCC			
	Gusb	2	F: TGGCTGGGTGTGGTATGAAC R: GGTGACCTCCCTCATGTTCC	-1.69	289	0.988
	Gusb	3	F: GGTGGAACATGAGGGAGGTC R: AGGGTATGAGGGGTCAGTGT	-2.32	170	0.980
	Gusb	4	F: GGCCTCTAGATAGCCTTGAGC R: ACACGCACTCCATTTTAGGGA	-1.30	486	0.955
	Hprt	1	F: GTTGGGCTTACCTCACTGCT R: TAATCACGACGCTGGGACTG	-2.23	181	0.979
	Hprt	2	F: GATCAGTCAACGGGGGACAT R: GGTCCTTTTCACCAGCAAGC	-1.49	367	0.976
	Hprt	3	F: ACAGGCCAGACTTTGTTGGA R: ACTTGCGCTCATCTTAGGCT	-1.40	423	0.960
	Hprt	4	F: CAGTCCCAGCGTCGTGATTA R: TGGCCTCCCATCTCCTTCAT	-2.60	143	0.995
	Tfrc	1	F: AAACTGGCTGAAACGGAGGA R: AGATCCAGCCTCACGAGGAG	-3.16	107	0.987
	Tfrc	2	F: AAGAGCTGCTGCAGAAAAGC R: ACGGTCTGGTTCCTCATAACC	-3.02	114	0.999
	Tfrc	3	F: GTTCGTACAGCAGCGGAAGT R: GGAAGTAGTCTCCACGAGCG	-3.08	111	0.972
	Tfrc	4	F: AGCAAAGTCTGGCGAGATGAA R: CCACATAACCCTCGGGAGAC	-3.28	102	0.991

Supplementary Table 1. Assay performance characteristics on 4 different primer pairs for Gusb, Hprt and Tfrc evaluated in parental cell line and in fully reprogrammed iPS cells. PCR efficiency *E*, slope, and associated correlation coefficient R^2 are shown. The serial dilutions in fully reprogrammed iPS cells were twofold.The efficiency for twofold dilutions was calculated using the formula E=2^(1/slope)-1*100. The serial dilutions in parental cells were tenfold, and the formula for calculating efficiency was E=10^(1/slope)-1*100. In fully reprogrammed iPS cells, the primer pair closest to 100% efficiency was #1 for Gusb (E=93%), #4 for Hprt (E=99%), and #2 for Tfrc (E=94%). The efficiency

for the same sets of primers differed in parental cells, giving E=163% for Gusb, E=143% for Hprt, and E=114% for Tfrc. The best pair of primers for each gene (Gusb #1, Hprt #4 and Tfrc #2) was chosen for the main experiment.

Gene	Comprehensive Ranking		Delta Ct		geNorm		NormFinder		BestKeeper	
	Value	Rank	SD aver.	Rank	M value	Rank	Stability	Rank	SD	Rank
Atp5f1	1.57	1	0.49	1	0.333	5	0.099	1	0.077	1
Pgk1	1.68	2	0.50	2	0.168	1	0.251	2	0.247	2
Gapdh	3.44	3	0.55	5	0.168	1	0.417	7	0.322	4
Ppia	3.66	4	0.56	3	0.197	2	0.353	4	0.323	5
Gusb	4.53	5	0.59	4	0.300	4	0.311	3	0.345	7
Tbp	5.18	6	0.60	6	0.467	7	0.400	5	0.285	3
Tfrc	6.88	7	0.62	7	0.253	3	0.458	8	0.422	10
Ywhaz	7.20	8	0.64	8	0.415	6	0.414	6	0.367	8
Rps18	8.80	9	0.65	10	0.532	9	0.564	10	0.337	6
Hprt	9.00	10	0.70	9	0.501	8	0.464	9	0.395	9
Actb	11.00	11	0.74	11	0.560	10	0.631	11	0.429	11
B2m	12.00	12	1.03	12	0.638	11	0.961	12	0.879	12

Supplementary Table 2. Ranking of the candidate reference genes' stability during reprogramming according to five different evaluation methods in mouse embryonic fibroblasts (MEFs). Atp5f1, Pgk1 and Gapdh were ranked as the most stable candidate reference genes overall, while Hprt, Actb and B2m were designated as the least stable ones.

	wells 1-6	wells 7-12
A	0.801	0.794
B	0.838	0.823
C	0.797	0.814
D	0.882	0.770
E	0.778	0.795
F	0.763	0.779
G	0.808	0.726
H	0.839	0.776

Supplementary Table 3. Efficiency values obtained by the standard curve method for all 16 replicas of a dilution set. The efficiency values E calculated by the classical calibration curve method are shown for the corresponding wells on the 96-well plate (for pipetting layout see Fig. S2). The Cq data for E calculation was taken from Dataset 2.

Cycle	F0=0.007	F0=0.0035	F0=0.00175	F0=0.000875	F0=0.0004375	F0=0.0002188
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0.8	0	0	0	0	0
14	0.8	0.83	0	0	0	0
15	0.81	0.81	0.79	0	0	0

16	0.81	0.82	0.8	0	0	0
17	0.8	0.82	0.8	0.73	0.82	0
18	0.78	0.81	0.8	0.78	0.82	0
19	0.76	0.79	0.81	0.79	0.82	0.81
20	0.72	0.76	0.79	0.79	0.81	0.8
21	0.69	0.73	0.77	0.78	0.82	0.82
22	0.66	0.7	0.74	0.76	0.8	0.81
23	0.62	0.67	0.71	0.74	0.78	0.8
24	0.6	0.64	0.68	0.71	0.76	0.79
25	0.57	0.61	0.65	0.68	0.73	0.76
26	0.54	0.58	0.62	0.65	0.7	0.74
27	0.52	0.56	0.59	0.63	0.67	0.71
28	0.5	0.54	0.57	0.6	0.65	0.68
29	0.48	0.51	0.55	0.58	0.62	0.66
30	0.46	0.49	0.53	0.55	0.6	0.63
31	0.44	0.48	0.51	0.53	0.58	0.61
32	0.43	0.46	0.49	0.51	0.55	0.59

Supplementary Table 4. The efficiency values calculated with the formula for the mean efficiency (4) with varying F0. The values corresponding to the region with relatively constant efficiency (SD=0.01) are denoted in red.

Boundaries	SD	Max E	Min E	Max-Min difference	Average E
20-150	0.0124	0.8346	0.7779	0.0567	0.8012
30-150	0.0125	0.8339	0.7790	0.0549	0.8039
40-150	0.0124	0.8311	0.7806	0.0505	0.8005
50-150	0.0132	0.8404	0.7750	0.0655	0.8028
60-150	0.0145	0.8485	0.7776	0.0709	0.8044
40-120	**0.0113**	0.8234	0.7798	**0.0436**	0.8012
40-150	0.0124	0.8311	0.7806	0.0505	0.8005
40-180	0.0116	0.8191	0.7675	0.0516	0.7894
40-210	0.0137	0.8274	0.7690	0.0583	0.7926
40-240	0.0165	0.8296	0.7568	0.0728	0.7852

Supplementary Table 5. Standard deviations, maximal and minimal efficiency (E) values and their difference, as well as average efficiency for differently set boundaries are shown. The minimal standard deviation (SD written in bold) is derived when setting the lower boundary at 40 RFU, and the upper boundary at 120 RFU , which falls within the exponential region of the curve. The lowest difference between the maximal E value and the minimal E value (Max-Min difference, in bold) is also observed with the same boundaries (40 - 120 RFU). Note that the average efficiency value tends to decline when the upper boundary is increased in the curve (150, 180, 210, 240) which agrees well with the notion of progressively declining efficiency with the gradual reaction saturation at later cycles. Overall, this result shows that the optimal region for mean efficiency calculation lies within the exponential region of the curve, and that the standard deviation will rise if fluorescence readings from later cycles are included in the calculations.

Wells	**A1-A6**	**A7-A12**	**B1-B6**	**B7-B12**	**C1-C6**	**C7-C12**	**D1-D6**	**D7-D12**
N of data points	218	208	215	233	228	211	205	204
Math. expectation	0.836	0.804	0.822	0.817	0.819	0.821	0.880	0.820
SD	0.096	0.112	0.083	0.085	0.053	0.079	0.101	0.071
Chi-square value	33.727	21.701	12.491	14.503	11.556	15.515	10.004	22.780
Degrees of freedom	9	10	9	8	5	7	9	7
Expected chi-square	16.919	18.307	16.919	15.507	11.07	14.067	16.919	14.067
Does it fit?	No	No	Yes	No	No	No	Yes	No
Wells	**E1-E6**	**E7-E12**	**F1-F6**	**F7-F12**	**G1-G6**	**G7-G12**	**H1-H6**	**H7-H12**
N of data points	207	199	247	224	209	199	183	227
Math. expectation	0.812	0.807	0.796	0.804	0.830	0.783	0.840	0.798
SD	0.054	0.079	0.060	0.083	0.062	0.059	0.067	0.080
Chi-square value	7.049	29.687	13.195	5.821	20.549	33.988	19.465	36.350
Degrees of freedom	5	7	6	8	6	6	6	8
Expected chi-square	11.07	14.067	12.592	16.919	12.592	12.592	12.592	16.919
Does it fit?	Yes	No	No	Yes	No	No	No	No

Supplementary Table 6. The results of Chi-square test on all 16 identical six-sets from Dataset 1. Chi-square test was performed on the groups of pairwise

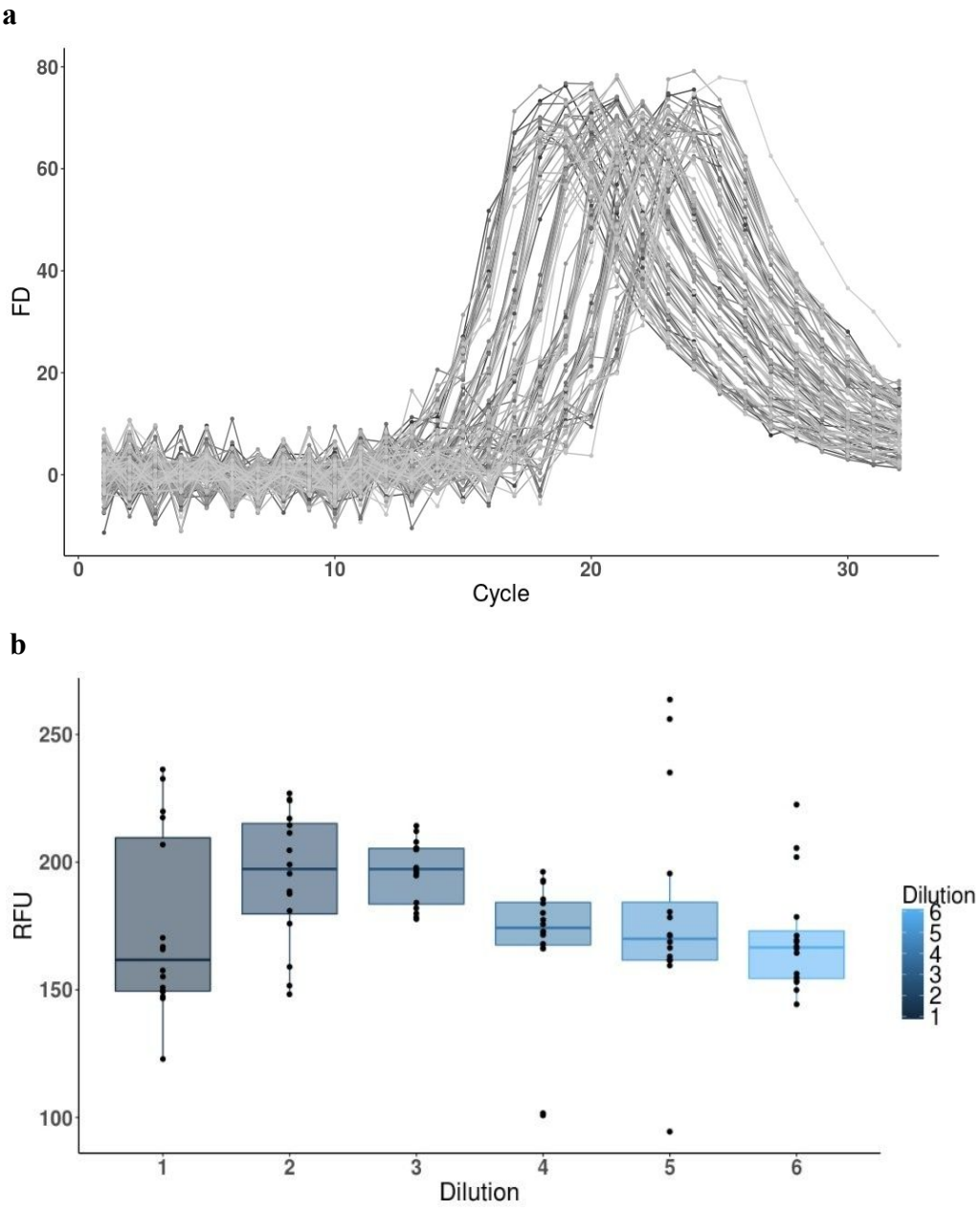
E measurements for each of the six-sets, as indicated. The number of data points (after the removal of statistically unreliable values whose frequency was less than 5, as per Chi-square criteria), the mathematical expectation of each group, standard deviation (SD) of each group, the chi-square value obtained for each group, the degrees of freedom for each group, and the expected Chi value based on the degrees of freedom, are shown. According to the Chi-square test principles, if the Chi-square value exceeds the expected Chi value, the distribution significantly deviates from normal, and parametric statistical instruments, such as quartile ranges or sigma, cannot be applied. Since all 16 six-sets were identical, and the number of non-normally distributed groups was significantly higher (12 out of 16), the analysis of outliers for these data should be performed using non-parametric tools.

Cycle	A1	A2	A3	A4	A5	A6
3	0.4	0.94	-0.09	1.3	0.31	0.33
4	0.13	0.27	0.21	0.53	0.08	0.77
5	0.1	0.17	0.27	0.24	0.33	0.43
6	-0.01	0.08	0.27	0.26	0.22	0.25
7	-0.05	0.02	0.17	0.23	0.14	0.12
8	0.01	-0.01	0.14	0.16	0.08	0.09
9	0.02	-0.02	0.1	0.12	0.08	0.09
10	0.03	0	0.04	0.1	0.04	0.06
11	0.02	0	0.04	0.08	0.01	0.06
12	0.02	0	0.04	0.07	0.02	0.04
13	**0.05**	0.01	0.04	0.05	0	0.02
14	0.08	**0.03**	0.05	0.04	0.01	0.01
15	0.13	0.05	0.05	0.05	0.01	0.01
16	0.2	0.09	**0.07**	0.05	0.01	0.01
17	0.3	0.15	0.09	0.05	0.02	0.01
18	0.41	0.23	0.13	**0.07**	0.03	0.01
19	0.51	0.33	0.2	0.1	**0.05**	0.02
20	0.6	0.42	0.28	0.14	0.08	**0.03**
21	0.67	0.51	0.36	0.2	0.13	0.06

22	0.71	0.57	0.43	0.27	0.19	0.1
23	0.72	0.61	0.49	0.33	0.25	0.15
24	0.72	0.64	0.53	0.39	0.32	0.2

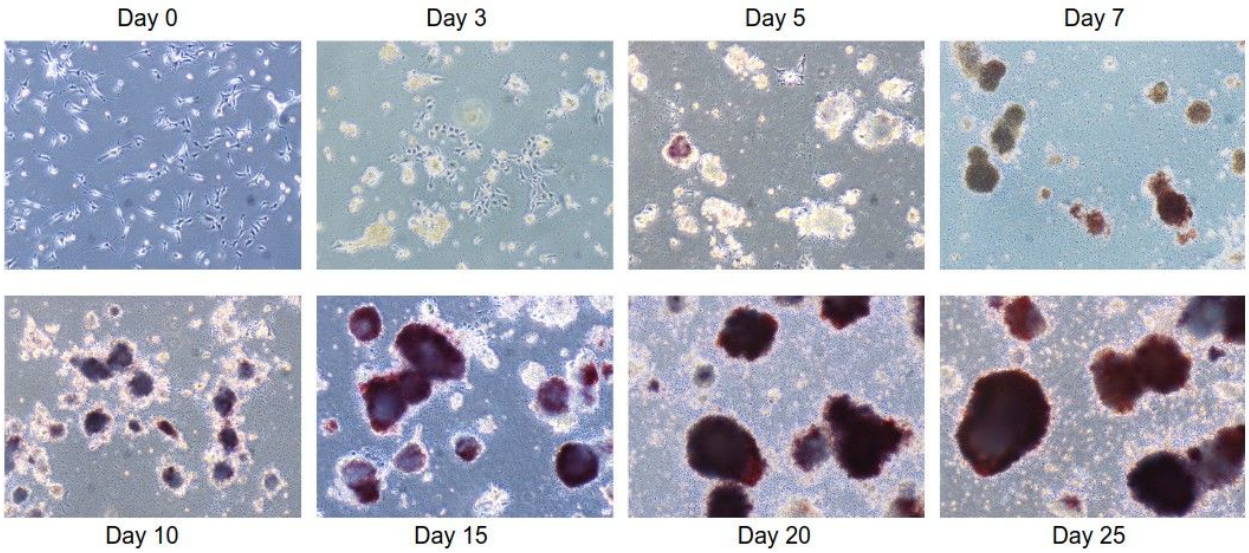
Supplementary Table 7. The 'first outliers' calculated by the formula from Tichopad et.al, 2003. The 'first outlier' values calculated for the wells A1 through A6 are denoted in red. The relatively constant values preceding them are denoted in blue. The fluorescence values from Dataset 1 for these wells in the corresponding cycles were A1=14.09 RFU, A2=16.7 RFU, A3=21.28 RFU, A4=27.19 RFU, A5=37.11 RFU, A6=28.2 RFU. Hence, the minimal fluorescence value was RFU=14.09, and the maximal fluorescence value was RFU=37.11. Thus, the tentative lower boundary of the exponential region can be set at approximately 10-40 RFU, depending on the actual curve.

Supplementary Figures



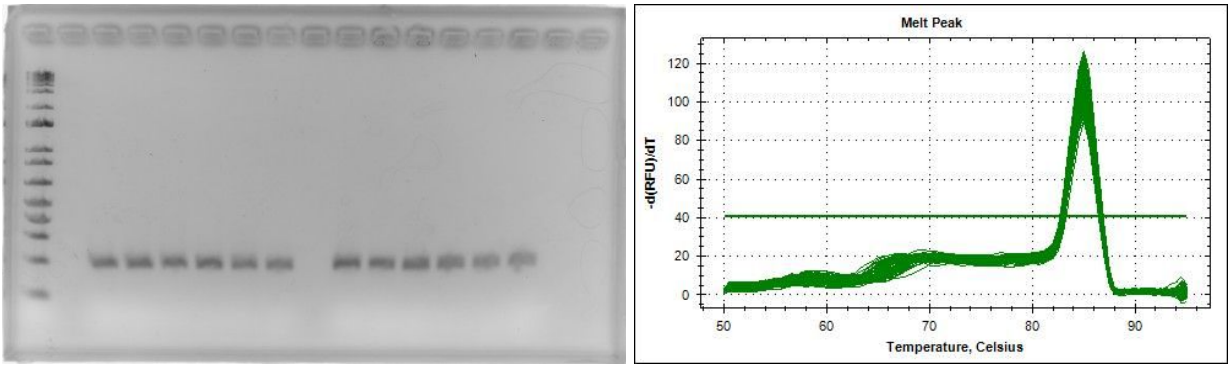
Supplementary Figure 1. The first derivative (FD) values and the corresponding fluorescence (RFU) values for 16 replicas of a 6-step serial dilution set taken from Dataset 1. (a) First derivative values of 96 amplification curves were plotted against the cycle at which they were obtained. As the dilution factor increases, the FD values are delayed and come at later cycles. The maximum

of the first derivative (FDM) for the most concentrated sample corresponds to cycle 18, while for the least concentrated sample this occurs at cycle 25. **(b)** The obtained FDM values were plotted against corresponding fluorescence units at the same cycle. Horizontal lines across the boxes denote the mean. Black dots indicate individual values, and the colors correspond to dilution (most concentrated sample - dark blue, least concentrated sample - light blue). The majority of FDM values roughly correspond to a RFU of 150-230.



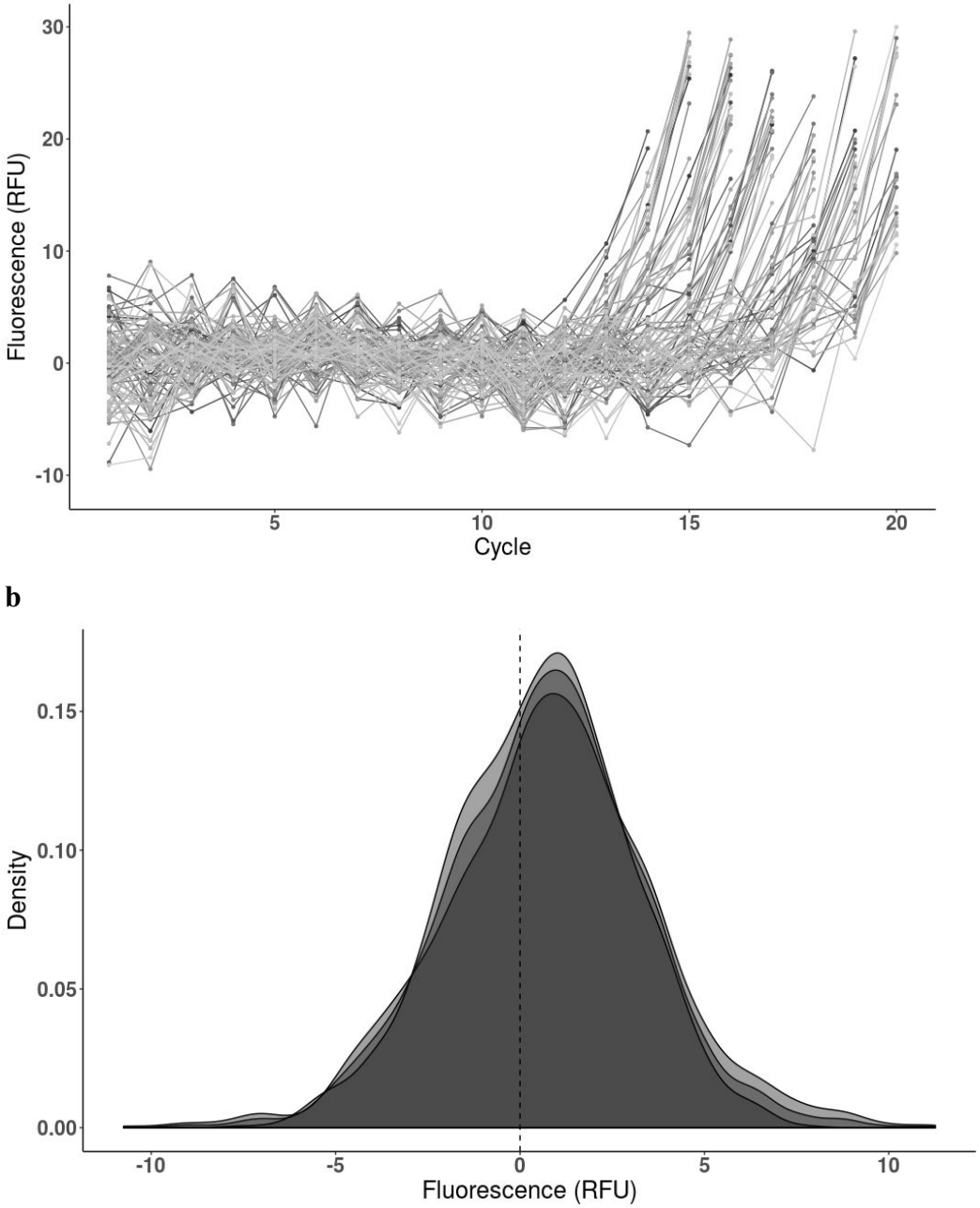
Supplementary Figure 2. The whole set of alkaline phosphatase staining throughout the iPS reprogramming course. Day 20 and Day 25 were excluded from mRNA analysis in Chapter 3 due to high apoptotic markers' expression (data not shown) which probably resulted from overgrowth of the colonies at this late stage, since the cells were not once split during the course of the experiment.

a **b**



Supplementary Figure 3. **Agarose gel of the PCR product and melting curve analysis.** **(a)** Agarose gel confirms the amplification of the expected product, showing a band at 194 bp. **(b)** Melting curve analysis shows no primer dimers and a single sharp peak, as expected.

a



Supplementary Figure 4. **Noise values and distribution in the beginning cycles of amplification. (a)** Amplification curves showing the beginning cycles for 96 qPCR reactions from Dataset 1. The noise is distributed close to zero and the noise phase appears to continue up to cycle 13. **(b)** Distribution of the noise across 2880 qPCR reactions taken from Dataset 1. The graph shows three groups of noise values: cycles 1-5, cycles 5-10 and cycles 1-10. All groups have nearly normal distribution with a non-zero mean, not shifting with increasing cycles, and the maximal data values reach approximately 10 fluorescence units.

List of the author's previous academic accomplishments

Publications

1. N-terminal dual lipidation-coupled molecular targeting into the primary cilium.
Kumeta M, Panina Y, Yamazaki H, Takeyasu K, Yoshimura SH.
Genes Cells. 2018 Jun 13. doi: 10.1111/gtc.12603. [Epub ahead of print]

2. Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process.
Panina Y, Germond A, Masui S, Watanabe TM.
Sci Rep. 2018 Jun 7;8(1):8716. doi: 10.1038/s41598-018-26707-8.

Presentations

1. 2014, November: "37th Annual Meeting of the Molecular Biology Society of Japan" (poster).
2. 2016, December: "39th Annual Meeting of the Molecular Biology Society of Japan" (poster).