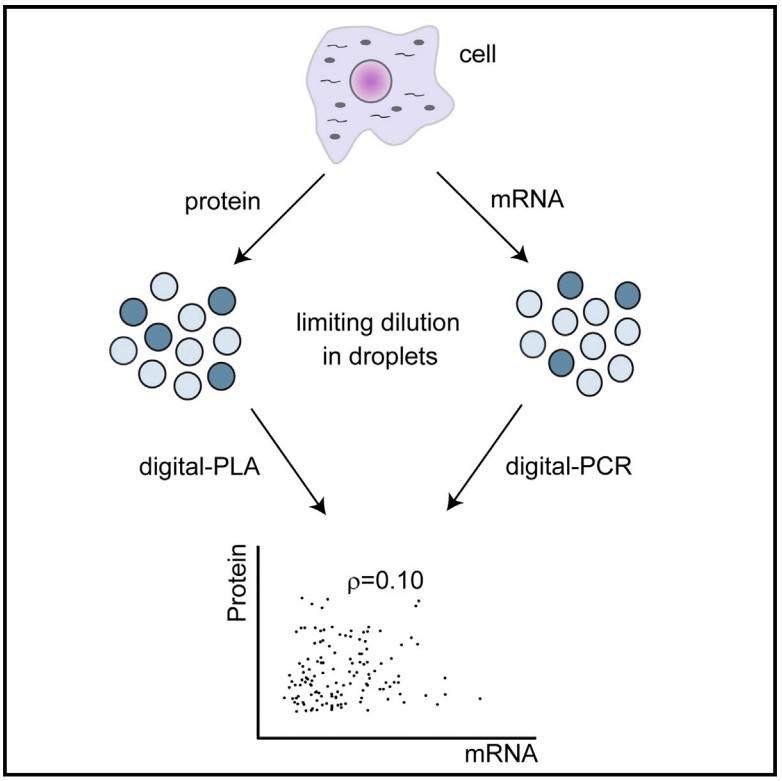
Technology

Molecular Cell

Digital Quantification of Proteins and mRNA in Single Mammalian Cells

Graphical Abstract



Authors

Cem Albayrak, Christian A. Jordi, Christoph Zechner, Jing Lin, Colette A. Bichsel, Mustafa Khammash, Savaş Tay

Correspondence

tays@uchicago.edu

In Brief

A single-protein counting method called digital PLA was developed, allowing ultrasensitive measurement of absolute protein and mRNA copy numbers from single mammalian cells. Joint single-cell mRNA/protein measurements were then used to construct and validate a stochastic two-state model for transcription and translation of the gene CD147.

Highlights

- Digital PLA protocol allows ultrasensitive protein quantification from single cells
- Combination with digital PCR allows joint single-cell protein and mRNA measurements
- Absolute mRNA and protein abundances measured jointly from single mammalian cells
- Joint mRNA-protein data was used to build a two-step model of gene expression

Albayrak et al., 2016, Molecular Cell *61*, 914–924
March 17, 2016
http://dx.doi.org/10.1016/j.molcel.2016.02.030

CellPress

CellPress

Molecular Cell
Technology

Digital Quantification of Proteins and mRNA in Single Mammalian Cells

Cem Albayrak,[1,3] Christian A. Jordi,[1,3] Christoph Zechner,[1] Jing Lin,[1] Colette A. Bichsel,[1,4] Mustafa Khammash,[1] and Savaş Tay[1,2,*]

[1]Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland
[2]Institute for Molecular Engineering, University of Chicago, Chicago, IL 60637, USA
[3]Co-first author
[4]Present address: Center for Biomedical Engineering, University of Bern, 3010 Bern, Switzerland
*Correspondence: tays@uchicago.edu
http://dx.doi.org/10.1016/j.molcel.2016.02.030

SUMMARY

Absolute quantification of macromolecules in single cells is critical for understanding and modeling biological systems that feature cellular heterogeneity. Here we show extremely sensitive and absolute quantification of both proteins and mRNA in single mammalian cells by a very practical workflow that combines proximity ligation assay (PLA) and digital PCR. This digital PLA method has femtomolar sensitivity, which enables the quantification of very small protein concentration changes over its entire 3-log dynamic range, a quality necessary for accounting for single-cell heterogeneity. We counted both endogenous (CD147) and exogenously expressed (GFP-p65) proteins from hundreds of single cells and determined the correlation between CD147 mRNA and the protein it encodes. Using our data, a stochastic two-state model of the central dogma was constructed and verified using joint mRNA/protein distributions, allowing us to estimate transcription burst sizes and extrinsic noise strength and calculate the transcription and translation rate constants in single mammalian cells.

INTRODUCTION

Single-cell heterogeneity features in diverse biological phenomena such as the emergence of antibiotic resistance (Allison et al., 2011), differential response to vaccines (Flatz et al., 2011) and signaling inputs (Kellogg and Tay, 2015), presence of cancer stem cells (Reya et al., 2001), variability in dynamics of immune signaling pathways (Tay et al., 2010), and virus infection (Snijder et al., 2009). Since proteins conduct the various actions that give rise to the observed phenotypes, their accurate quantification from single cells is necessary to understand the phenomena that arise from single-cell heterogeneity (Gordon et al., 2007; Sigal et al., 2006) and to build accurate mathematical models of complex biological systems (Karr et al., 2012). mRNA abundance is often a poor surrogate for protein quantity, since the two values are not necessarily correlated, as shown in prokaryotic cells (Taniguchi et al., 2010). In addition to determining presence and degree of correlation, joint mRNA and protein quantification from single cells would also enable researchers to build and accurately calibrate mathematical models of transcription and translation. If the mRNA and protein content of each cell can be measured, an mRNA-protein correlation term could be derived exactly for the heterogeneous population and used as an additional input to the model. A practical and sensitive method for absolute joint mRNA and protein quantification would thus be greatly desirable for quantitative systems biology, especially for mammalian cells where the correlations between the two species has remained largely unexplored at the single cell level due to technical difficulties resulting from counting individual molecules at high abundances.

Design

Methods developed for protein quantification had limited application to single cells due to significant shortcomings in sensitivity and practicality. PLA (Gullberg et al., 2004; Schallmeiner et al., 2007) and its variant, the proximity extension assay (Lundberg et al., 2011), can detect proteins at low concentrations in dilute protein solutions. These techniques, however, rely on qPCR readout that scales with exponential amplification and thus becomes inaccurate for quantifying differences smaller than fold changes (Whale et al., 2012). Microfluidic PLA (Blazek et al., 2013) and light-mediated cellular barcoding (Agasti et al., 2012) provide only relative protein quantification from single cells. Single-cell mass cytometry (CyTOF) enables multiplexed readout of protein and mRNA content but provides relative quantification as well (Bendall et al., 2011; Frei et al., 2016). Yet other methods such as the single-cell barcode chip and western blot chip either provide relative quantification or require extensive genetic manipulation, sophisticated device manufacture, and advanced microscopy (Huang et al., 2007; Hughes et al., 2014; Lu et al., 2015; Ma et al., 2011; Shi et al., 2012; Taniguchi et al., 2010). Digital quantification based on limiting dilution provides high sensitivity and the ability to quantify differences much smaller than fold changes (Fan et al., 2012). For example, digital ELISA (Kim et al., 2012; Rissin et al., 2010) achieved sub-femtomolar limit of detection (LOD) for proteins like prostate-specific antigen in human sera. Furthermore, PLA using rolling circle amplification reduced the LOD of the method compared to

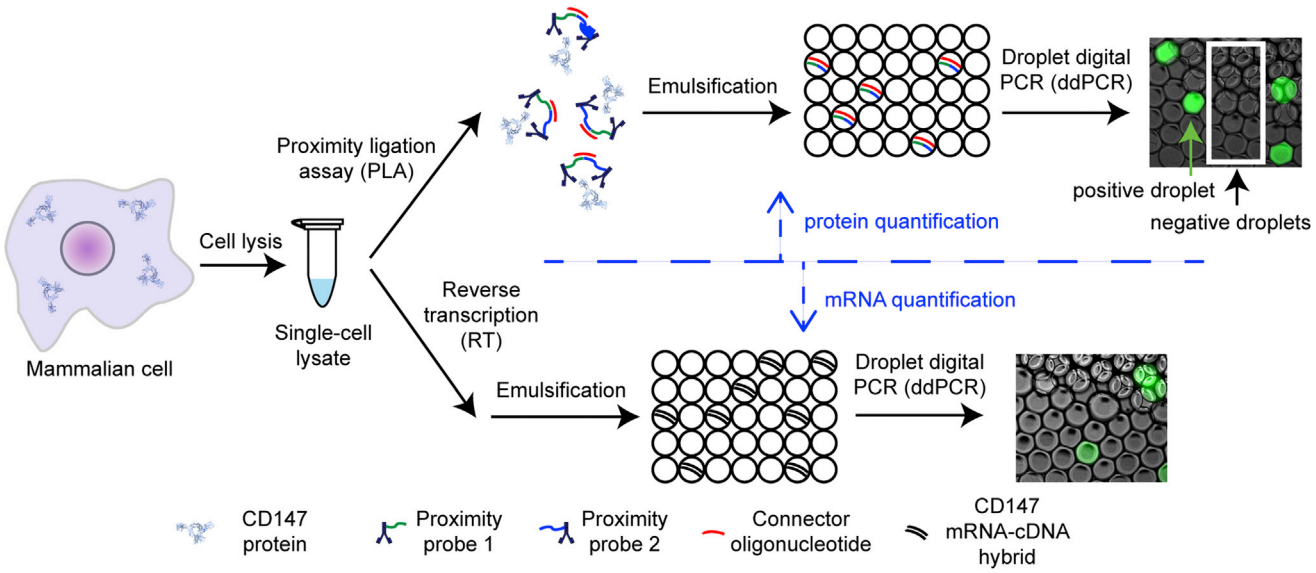


Figure 1. Digital PLA Protocol for Absolute Protein and mRNA Quantification from Single Cells

The single-cell lysate is split into two, and the mRNA and protein from individual mammalian cells are quantified using two-step RT-PCR and digital PLA, respectively. In digital PLA, proximity probes bind the target protein; the connector oligonucleotide hybridizes with the probes; a double-stranded DNA is formed following ligation; and the proteins are digested by proteases. The remaining dsDNA (and the cDNA in RT-ddPCR workflow) is distributed among 20,000 droplets at limiting dilution, each of which contains zero or one dsDNA molecule. These molecules are amplified by PCR to be detected. Pictures on the right shows fluorescent droplets that contain or lack DNA after ddPCR. CD147 structure was retrieved from RSCB PDB (PDB: 3B5H) (Yu et al., 2008).

qPCR readout (Ke et al., 2013). Besides CyTOF, two of the reported methods were used for joint protein and nucleic acid quantification from single cells, namely advanced fluorescence microscopy for bacteria and PLA in combination with RT-qPCR for mammalian cells. Single-molecule resolution was attained with the former, but the method required substantial genetic manipulation and is applicable only to dilute proteins in bacteria (Taniguchi et al., 2010), and the latter two only provided relative quantification and had low sensitivity due to qPCR comprising the terminal step in the workflow (Ståhlberg et al., 2012).

Here we describe absolute and simultaneous quantification of both proteins and mRNAs from single mammalian cells by combining PLA with digital PCR in a very practical, sensitive, and robust protocol called digital PLA. We first use droplet digital PCR (ddPCR) to achieve digital PLA and later combine it with two-step RT-ddPCR to generate a single workflow for joint mRNA and protein quantification (Figure 1). The use of ddPCR rather than qPCR enables absolute quantification of proteins independent of amplification efficiency, and with much greater precision (Hindson et al., 2013) (Figures S1A and S1B).

We quantified three proteins using digital PLA; namely the endogenous transmembrane proteins CD147 (also known as extracellular matrix metalloproteinase inducer [EMMPRIN]) and intercellular adhesion molecule-1 (ICAM-1), and exogenously expressed GFP. Femtomolar LODs were achieved for all three proteins, and digital PLA has a linear dynamic range over two to three orders of magnitude. Digital PLA and digital PCR were then used to quantify CD147 mRNA and protein copy numbers in the same single cell. CD147 mRNA and protein data from single cells, and especially their co-variance, were subsequently used to develop and evaluate different models of the central dogma. A two-state model of gene expression is shown to fit well to our experimental data (Peccoud and Ycart, 1995; Blake et al., 2003; Paulsson 2005). Using this model, we calculated the transcription burst sizes, transcription and translation rate constants, and extrinsic noise strength in single mammalian cells.

Our method for absolute mRNA-protein quantification from single cells can be readily adapted by other researchers; since the entire assay reagents and equipment are commercially available, complex microfluidic device fabrication is not needed, and no genetic manipulation of the cells is required. A single researcher can quantify proteins from 100 samples (single cells or clinical samples) in a day. Absolute mRNA and protein measurements from individual mammalian cells will enable scientists to better understand cell-to-cell heterogeneity and evaluate and/or construct more accurate biological models of the central dogma in various systems.

RESULTS

Digital PLA Development

In order to quantify proteins from single cells, they are first isolated, lysed, and combined with oligonucleotide-bound antibodies (proximity probes, Figure 1) and the connector oligonucleotide. The antibodies bind to their epitopes, the oligonucleotides bound to the antibodies come in close proximity, and the connector oligonucleotide hybridizes to them, thereby forming a complex of target protein, proximity probes, and connector DNA. The three-oligonucleotide complex becomes a double-stranded DNA (dsDNA) after ligation, completing the conversion

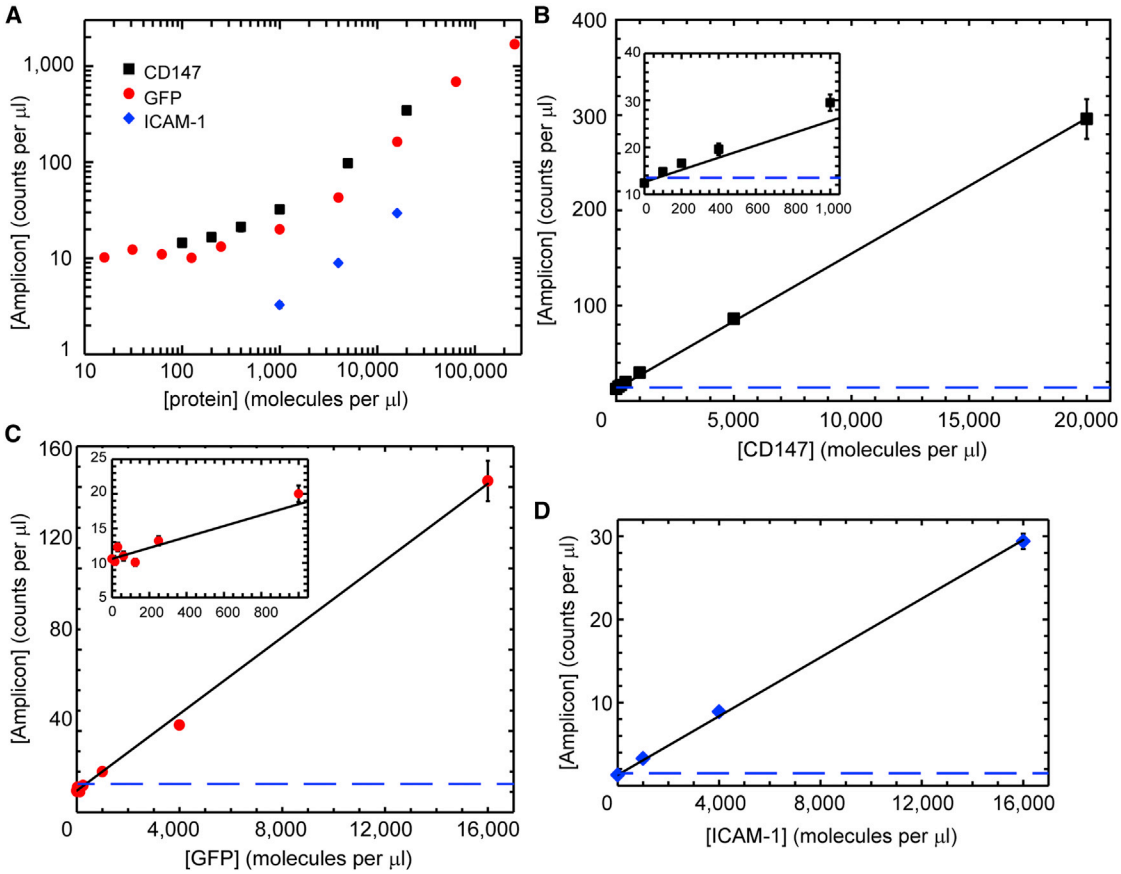


Figure 2. Dynamic Range and Sensitivity of Digital PLA for Protein Quantification

(A–D) Assay calibration curves are plotted either together in logarithmic scale (A) or individually in linear scale ([B]–[D]). Digital PLA readout is given in number of double-stranded DNA amplicons per µl and subsequently converted to absolute protein amounts using these calibration curves. Solid lines indicate linear regression for each of the three proteins: (B) CD147, $y = 0.0142x + 12.4$, $R^2 = 0.9997$; (C) GFP, $y = 0.0094x + 10.6$, $R^2 = 0.9982$; and (D) ICAM-1, $y = 0.00177x + 1.30$, $R^2 = 0.9993$. Insets in (B) and (C) given to clearly show sub-femtomolar assay detection limits (Assay LOD, blue dashed lines); exact LOD values are given in Table 1. Error bars show combined error from both Poisson noise and technical error from replicates (n = 15 or 29 for CD147, 3 or 6 for GFP, and 12 for ICAM-1). Additional details concerning digital PLA development can be found in Figure S1.

of the targeted protein to dsDNA as in standard PLA (Gullberg et al., 2004). After a brief proteolysis step, the dsDNA is quantified by ddPCR (Hindson et al., 2011; Pinheiro et al., 2012). Using a commercially available microfluidic device, the PLA solution (which now only contains the dsDNA) is emulsified to create ~20,000 nanoliter-sized droplets under limiting dilution, each of which contains either zero or one dsDNA molecule. Single dsDNA molecules in the droplets are then amplified by PCR and counted by measuring the resultant fluorescence from hydrolysis probes using the commercially available droplet reader. These absolute DNA counts are then converted to absolute protein numbers using a calibration curve. The method shows good reproducibility as confirmed by making triplicates of eight PLA samples—seven of which did not show any significant variation other than fundamental digital PCR error (Figure S1E).

We evaluated the digital PLA performance on pure protein standards. Dilution series of human CD147, murine ICAM-1, and GFP spanning three to four orders of magnitude were analyzed by digital PLA (note that GFP is used here as the target protein and not as a fluorescent tag for protein detection). The calibration curves show that digital PLA has a linear dynamic range across three orders of magnitude (Figure 2). Femtomolar LOD (defined as signal at 3 SDs above background) was achieved for all three proteins, CD147 giving the lowest LOD at 86 proteins per µl of ddPCR solution (corresponding to 16.2 fM in the sample solution) and the most sensitive calibration curve (i.e., 1/slope = 70 proteins per count) (Table 1). Specificity of the CD147 and GFP probes was confirmed by analyzing antigen-free lysate. No positive signal was generated when the human CD147 probes were used to analyze mouse 3T3 bulk lysate (Figure S3A). Mouse cells transfected with a plasmid expressing human CD147 recovered the PLA signal (Figure S3B). Background signal from one-cell equivalent lysate of GFP-negative cells was comparable to that from buffer (Figure S3C). In terms of absolute protein amounts (~20,000 proteins per cell, Table 1), the LOD of our method is comparable to those of other single-cell methods such as the barcode and western blot chips (Hughes et al., 2014; Shi et al., 2012). A major advantage of

Table 1. Digital PLA Specifications for Analyzed Proteins

Protein	Assay LOD[a] (fM)	Sensitivity[b] (proteins per DNA)	Sample LOD[c] (fM)
CD147	0.14	70	16.2
GFP[d]	0.18	95	20.1
ICAM-1	0.34	565	38.3

[a]Assay limit of detection (LOD), gives the detection limit in the assay conditions (mol protein per unit volume of ddPCR solution). LOD is defined as the protein concentration at which the signal is 3 SDs above background.
[b]Sensitivity, equal to the inverse of the slope of the calibration curve, is an inverted metric: the lower the number of proteins per counted DNA molecule, the higher the assay sensitivity.
[c]Sample LOD, given in moles of protein per unit volume of the original sample. The 113-fold dilution in the digital PLA workflow results in the difference between the sample and assay LODs.
[d]LOD and sensitivity calculated using data from a single day ("Curve 1," Figure S1F).

digital PLA over analog (i.e., qPCR-based) methods is the ability to measure very small protein concentration differences over its entire dynamic range.

Direct and Digital Quantification of Protein Copy Numbers in Single Cells

GFP-p65 (a subunit of the NF-κB protein family) fusion proteins were quantified from single mouse macrophages (RAW264.7) using digital PLA; individual macrophages were sorted by FACS into wells containing 2 μl of standard lysis buffer (Figure S2B). GFP-p65 counts from 45 single cells lie in the linear range of the assay and show significant variation, ranging from 32,400 ± 2,500 to 188,000 ± 11,000 proteins (Figures 3C and S3D). This large heterogeneity can only be observed on the single cell level and is masked in bulk measurements (Figure 3D). Error due to Poisson noise was small (6%–8% CV); we could thus determine single-cell protein content within several thousand molecules. In addition to the exogenously expressed GFP-p65 proteins, we also quantified endogenous CD147 proteins from individual human embryonic kidney (HEK) cells, as described in the following sections.

Direct and Digital Quantification of mRNA Copy Numbers in Single Cells

We developed a robust protocol for counting mRNAs in single cells so that they can be quantified simultaneously with proteins from the same single cell. Different combinations of RT-PCR protocols, primers, hydrolysis probes, and PCR cycle numbers were tested for mRNA quantification from bulk-extracted total RNA (see Experimental Procedures for detailed description of assay optimization). The combination of a two-step protocol with Probe 2 and 40 amplification cycles yielded the highest number of CD147 mRNAs (101 ± 1) per cell (Figures S4A and S4B) and easily distinguishable positive and negative ddPCR droplets (Figure S4C). Other permutations either failed to produce sufficient cDNA or suffered from high background fluorescence of empty ddPCR droplets (Figure S4D).

After a suitable PCR protocol for mRNA detection had been found, the remaining challenge was to find a cell lysis strategy compatible with both digital PLA and RT-ddPCR. After screening various conditions, particularly regarding the lysis buffer concentration (see Supplemental Experimental Procedures), we found that 0.1X TM lysis buffer did not interfere with droplet formation during ddPCR preparation (Figure S4E), could lyse up to 100 mammalian cells (Figure S4F), did not inhibit RT (Figures S4G and S4H), and was compatible with PLA (Figure S5A). Digital PLA was even more sensitive in the new buffer matrix (Figure S5B). Therefore, we successfully established a direct lysis protocol compatible with RT-ddPCR. We verified the performance of our RT-ddPCR workflow using T7-in-vitro transcribed CD147 mRNA and found that we recovered 88% of the theoretical signal (Figure S5C). The discrepancy is possibly due to RT biases (Ståhlberg et al., 2004). We corrected our entire single cell RT-ddPCR reads by the observed difference and introduced a LOD for mRNA quantification at 52.5 copies per cell (87aM).

Joint Protein and mRNA Measurements from Single Cells

CD147 proteins and mRNA were quantified from the same single cells using a joint two-step RT-ddPCR and digital PLA workflow in the new buffer matrix. Individual HEK293T cells were sorted into wells containing 3 μl of 0.1X TM buffer (Figure S2A). Single-cell lysates were then split in a 1:2 ratio; 1 μl was taken out for mRNA quantification, while CD147 proteins were quantified by digital PLA from the remaining 2 μl (Figures 3 and 4). In this configuration, the LOD for CD147 proteins was 29,300 molecules per cell. Analyzing both fractions of the sample with PLA revealed good agreement between the two half-lysates, which indicates that no further error, besides the fundamental ddPCR error was introduced (Figure S5D) in six out of nine cases. Of the 186 single cells sorted, 138 cells had measurable CD147 protein content above the LOD. Single-cell CD147 counts varied 18-fold, ranging from 35,900 ± 2,900 to 636,000 ± 36,000 proteins (Figures 3A). All single cell measurements lay within the calibrated range of the assay, whereas the 100 cell populations were slightly above (Figure S3E). The mean CD147 protein value (218,000 molecules per cell) from human cells we measured agrees well with estimates from single mouse fibroblasts (258,978 molecules per cell; Schwanhäusser et al., 2011). The range of CD147 amounts per cell narrowed as the number of cells per well increased from 1 to 100. Protein counts from 100 cells show an average of 67,200 ± 1,000 to 229,000 ± 3,000 CD147s per cell (Figure 3B).

Single-cell CD147 mRNA numbers ranged from 57 ± 18 to 2,651 ± 121 molecules per cell, with a population mean of 188.2 and a SD of 90 (Figure 4A). These mean mRNA values are comparable to 189 CD147 mRNAs per cell reported for single mouse fibroblasts (Schwanhäusser et al., 2011). Additionally, 100 cells were sorted and measured, serving as estimates for bulk average. 100-cell averages ranged from 106 ± 2 to 215 ± 3 copies per cell, exhibiting a narrower range on a per cell basis (Figure 4A). As a result, on average, approximately 1,000 proteins were translated from each CD147 mRNA in HEK cells.

Analysis of mRNA and Protein Correlation in Single Mammalian Cells

The unique ability to absolutely and sensitively quantify CD147 mRNA and protein content of the same single cell allowed us

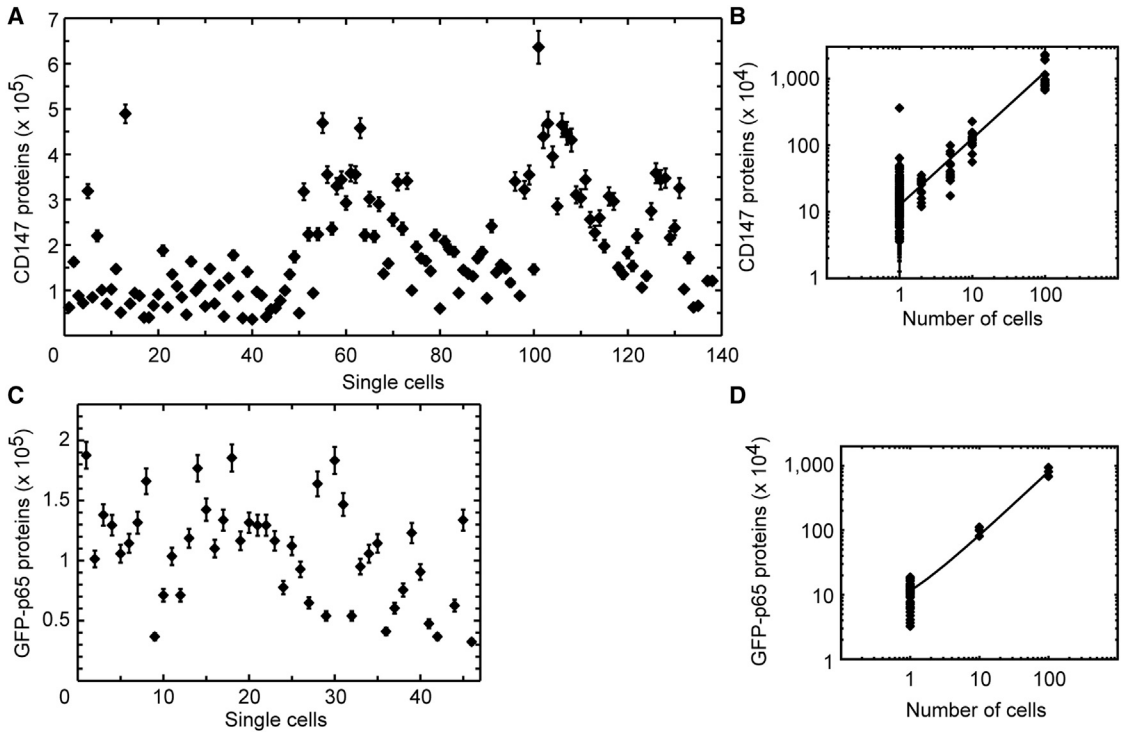


Figure 3. Absolute CD147 and GFP-p65 Protein Quantification from Single Mammalian Cells Using Digital PLA

(A) CD147 values from 138 single human embryonic kidney (HEK293T) cells display great single-cell heterogeneity.

(B) CD147 measurements from 1, 2, 5, 10, and 100 HEK293T cells. Solid line indicates linear regression (y = 125,000x, R^2 = 0.785).

(C) GFP-p65 measurements from 45 single mouse macrophages.

(D) GFP-p65 measurements from 1, 10, and 100 macrophages. Solid line indicates linear regression (y = 80,900x, R^2 = 0.980). Single-cell data included in (B) and (D) are the same as those plotted in (A) and (C), respectively. Error bars show Poisson error from 11,679 to 16,882 droplets. Additional information regarding the performed FACS sorts and used controls and calibration curves can be found in Figures S2 and S3.

to analyze the correlation of these molecules. Surprisingly, our high-sensitivity measurements showed that CD147 mRNA and protein are poorly correlated (Spearman R = 0.10, Figure. 4B), which contrasts with previous bulk observations that found a significantly higher degree of correlation (Pearson R = 0.41, 0.59, Schwanhäusser et al., 2011; Tian et al., 2004), similar to single-cell measurements from bacteria (Pearson R = 0.01, Taniguchi et al., 2010). We note that the mRNA/protein correlation would (incorrectly) appear to be higher if a method with insufficient sensitivity (i.e., qPCR/PLA) was used here, as most of the single-cell protein and mRNA variability lie within a 2-fold difference around the mean. These findings highlight the importance of high-resolution single-cell measurements compared to low-sensitivity or population-averaged (bulk) analysis in understanding protein transcription and translation.

Single-Cell mRNA/Protein Distributions Indicate a Two-State Stochastic Process for Protein Expression

Single-cell CD147 mRNA and protein copy numbers have very broad distributions and cannot be fit with a Poisson probability density function (pdf, Figure 5), suggesting that the intrinsic noise caused by random production and degradation of these macromolecules (i.e., a one-step birth-death process) is not enough to explain their variance (Raj and van Oudenaarden, 2008). A gamma pdf could be fit to the single-cell CD147 mRNA distributions ($p > 0.05$, Figure 5), indicating a two-state stochastic process for gene expression (Paulsson, 2005; Peccoud and Ycart, 1995) (Figure 6A). In this model, transcription occurs in random bursts of varying size and results in larger mRNA copy number variances. The mRNA distribution can take the form of a gamma distribution if the gene is mostly inactive and the RNA degradation during a burst is small (Raj et al., 2006). In this case, the rate and shape of the resulting gamma pdf is interpreted as the transcription burst rate and the average burst size of the gene. For our cell population, we calculated an average burst size of 40 ± 4 CD147 mRNAs, a gene activation rate of 3.78 ± 0.38 per day, and a mean copy number of 188 ± 28 CD147 mRNAs per cell; the latter of which agrees well with the mean of 188.2 CD147 mRNAs obtained from our single-cell measurements.

A Stochastic Model of Gene Expression Parameterized by Single Cell mRNA and Protein Measurements

Measurement of single-cell mRNA and protein abundances allowed us to construct and verify a stochastic model of CD17 gene expression in single-cells (Figure 6). We considered a two-state model, with random promoter switching, transcription, and translation, as well as extrinsic variability (Figure 6A; see Supplemental Experimental Procedures). Unlike fitting a pdf of

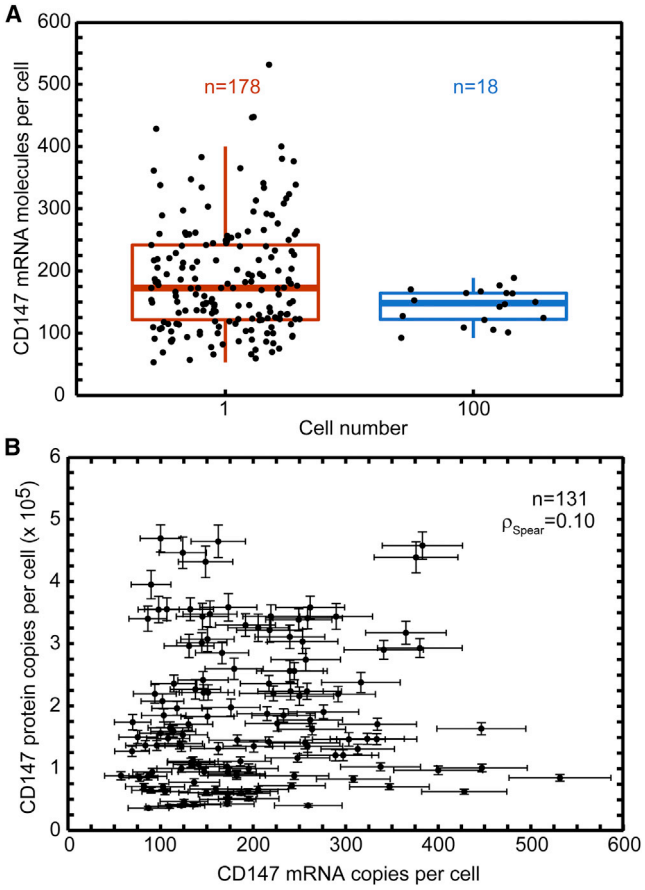


Figure 4. Absolute Quantification of CD147 mRNA and Protein Copy Numbers in Human Embryonic Kidney Cells

(A) CD147 quantification from single and 100 cells measured by digital PLA. Plotted copy numbers were normalized by the number of sorted cells. The number of replicates is indicated above each boxplot. Mean single-cell and population-averaged copy numbers show good agreement.

(B) CD147 mRNA and protein copy numbers from 131 single HEK293T cells. Error bars show Poisson error from >10,000 droplets. Supplemental information concerning RT-ddPCR development and technical controls can be found in Figures S4 and S5.

a presumed type, the stochastic two-state model is not based on any a priori assumptions about the regime in which gene activation and RNA degradation take place. However, exact inference of such models based on mRNA and protein distributions is computationally intractable. We therefore employed a moment-based analysis in which the single-cell abundances and their variability are summarized by only a few statistics such as means and variances (Zechner et al., 2012). We derived equations for first- and second-order moments of mRNA and protein (i.e., E[M(t)], E[P(t)], E[M(t)^2] and E[P(t)^2]), which were used as a basis for subsequent analyses. The combined mRNA and protein readout allowed us to use the correlation between mRNA and protein, E[M(t)P(t)], to verify the model. Since CD147 is expressed constitutively, we assumed stationarity of the system and fitted the analytically derived steady-state moments to the ones obtained experimentally. The model comprises seven parameters in total, which cannot be uniquely determined solely from steady-state measurements of mRNA and protein. We thus introduced a suitable reparameterization of the model in terms of an effective rate of transcription (Zenklusen et al., 2008), subsuming promoter switching and transcription into a single parameter. Furthermore, we incorporated prior knowledge based on the reported mRNA and protein half-lives (20.97 hr, 33.48 hr; Schwanhäusser et al., 2011). Although the used half-lives had been measured in a mouse cell line, we do not expect them to change dramatically in human cells. This assumption is strengthened by the fact that the mean CD147 protein and mRNA copy numbers we determined closely matched the values reported by Schwanhäusser and colleagues. A Markov-chain Monte Carlo scheme was then performed to infer the extrinsic variability and kinetic parameters from E[M(t)], E[P(t)], E[M(t)^2], and E[P(t)^2] (Figure 6B; Table S1). We used those parameters to computationally predict the mRNA-protein correlation E[M(t)P(t)] and checked its consistency with its experimentally measured counterpart (which was not used to fit the model parameters) to cross-validate our model. The data collected on different days were analyzed separately. We found that for four out of five experiments, the model predictions of mRNA-protein correlation agree very well with the experimental data (Figure 6C), indicating that for the considered CD147 expression system the proposed two-state model provides a suitable mathematical description.

DISCUSSION

We achieved digital protein quantification by combining PLA with digital-PCR. Furthermore, we were able to integrate RT-ddPCR into the same practical workflow, thus allowing us to measure absolute and ultrasensitive protein and mRNA copy numbers from single cells. The application of a modified digital PCR protocol to protein measurements resulted in high sensitivity, since digital PCR is better suited to detect small concentration differences than conventional qPCR. As a result, digital PLA enables the quantification of small changes in protein concentration (as few as 70 proteins/μl in the ddPCR solution and 8,000 proteins in single-cells) over the entire dynamic range of the assay (Figure 2). This high sensitivity is necessary for an accurate analysis and understanding of single-cell variability. The femtomolar detection limit and high sensitivity of digital PLA allowed us to spare a fraction of the single-cell lysate for mRNA quantification, where the mRNA molecules are counted using a two-step reverse transcription and droplet digital PCR protocol.

In addition to high sensitivity, digital PLA is very practical, as it does not require any genetic manipulation in the host cell or sophisticated microfluidic device manufacture, and all of the required hardware and consumables are commercially available. The method can thus be rapidly adapted by other laboratories and can be used to quantify macromolecules from primary cells. It can be generalized to any protein, provided that specific antibodies are available, and it can be used to quantify protein complexes. Indirect protein measurements like mass spectrometry can be accurately calibrated with digital PLA measurements. In addition to single-cell applications, digital PLA provides a sensitive method in detecting small changes in the concentrations of

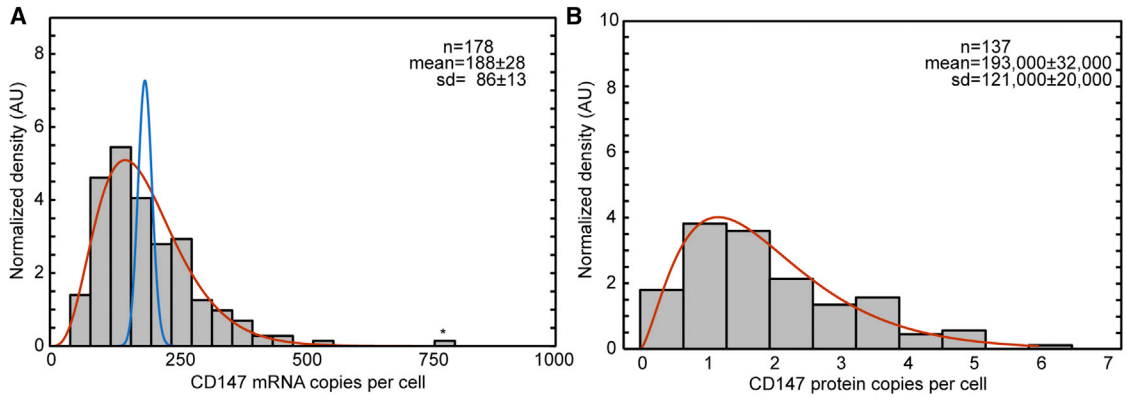


Figure 5. Fitting of Probability Density Functions to Single Cell Copy Number Distributions

(A and B) Histograms of single-cell CD147 mRNA (A) and protein (B) copy numbers.

(A) Poisson (blue) and gamma (red) probability density functions (pdf) were fit to single-cell CD147 mRNA data. The mRNA is binned in groups of 40 copies per cell. p value of the gamma fits generated using a χ^2 test is 0.65. Outliers indicated by an asterisk (*) were not included in the fits. The mRNA distributions fit the gamma pdf well.

(B) Gamma pdf (red) fit to the CD147 protein data (binned in groups of 65,000 copies per cell). Measurements below the LOD were omitted, p = 0.24.

clinically relevant proteins, which is critical for early diagnosis of disease.

Digital PLA complements existing methods such as in situ PLA (Blazek et al., 2013; Söderberg et al., 2006). These methods, while having single molecule sensitivity, require small copy numbers and even spatial distribution of the fluorescent macromolecules so that each diffraction-limited spot can be attributed to a single molecule (Golding et al., 2005; Taniguchi et al., 2010). In contrast, our method has a dynamic range across several orders of magnitude, is suitable for any native mRNA and protein (provided that the abundance of mRNA is >52 per cell and $\geq\sim10^4$ copies per cell for proteins), and does not require appending a bulky fluorescent protein to the target protein (which may alter the regulation and abundance of the latter).

While mRNA and protein levels in bacteria are not well correlated (Taniguchi et al., 2010), population-averaged (bulk) analyses estimated a higher degree of correlation in mammalian cells (Schwanhäusser et al., 2011; Tian et al., 2004). Our single-cell measurements, however, show that CD147 mRNA and protein copy numbers are actually poorly correlated and that mRNA abundance serves as a poor surrogate for protein abundance (Figure 4B). A possible explanation for this discrepancy from theoretical studies is that noisy post-transcriptional mechanisms regulating the translation and maturation of similar membrane-expressed proteins could destroy the correlation between mRNA and protein abundance in single cells (Caldwell et al., 2014). Another possible reason is that fluctuations are diminished when averaging mRNA and protein over multiple cells, resulting in higher correlation among the two (Vogel and Marcotte, 2012). Whether the observed lack of correlation is a general feature of mammalian gene expression, or if there are other genes where mRNA and protein levels better correlate, needs to be further investigated.

Our single-cell measurements revealed practical ''rule of thumb'' values for transcription and translation and allowed us to construct an accurate stochastic model for central dogma in mammalian cells. We found that approximately 1,000 proteins are created from a single mRNA molecule in a single mammalian cell. Cell-to-cell variability in protein abundance ranges from 5- to 18-fold (for GFP-p65 and CD147, respectively), with a mean of approximately 100,000 proteins per mammalian cell.

Finally, we used our single-molecule mRNA and protein dataset to parameterize a stochastic two-state model of CD147 transcription and translation. Our unique dataset enabled us to empirically determine the mRNA-protein correlation term E[M(t) P(t)] for individual mammalian cells and validate its computationally derived counterpart. The agreement between the two terms (Figure 6C) showed that the two-state model described well the stochastic transcription and translation of CD147. The paired mRNA-protein data (i.e., the correlation term) may also be used as an additional input (i.e., in addition to the means and variances of the single-cell mRNA and protein values) for modeling the transcription and translation of a given protein, or investigating fluctuations in dynamic systems in response to a stimulus. Joint absolute mRNA-protein quantification can thus be used to accurately model the central dogma for any gene on a single-cell level.

Limits

The LOD of the described assay currently prevents single cell quantification of very rare protein species. Digital PLA sensitivity is limited by the affinity of the proximity probes and the fact that proximity ligation is conducted in bulk. Furthermore, the target protein is initially diluted by two orders of magnitude, with an additional greater dilution occurring during cell sorting and lysis. The intracellular contents of a single mammalian cell (V ≈ 1–10 pl) are diluted 10^5-fold when the cell is lysed in a 3 μl solution. We hypothesize that digital PLA sensitivity can be further increased by using antibodies with higher affinity and by conducting assay in smaller volumes. Currently, a single researcher can analyze up to 100 samples with digital PLA in a day using standard pipette-and-dish techniques, and this number can be doubled if only

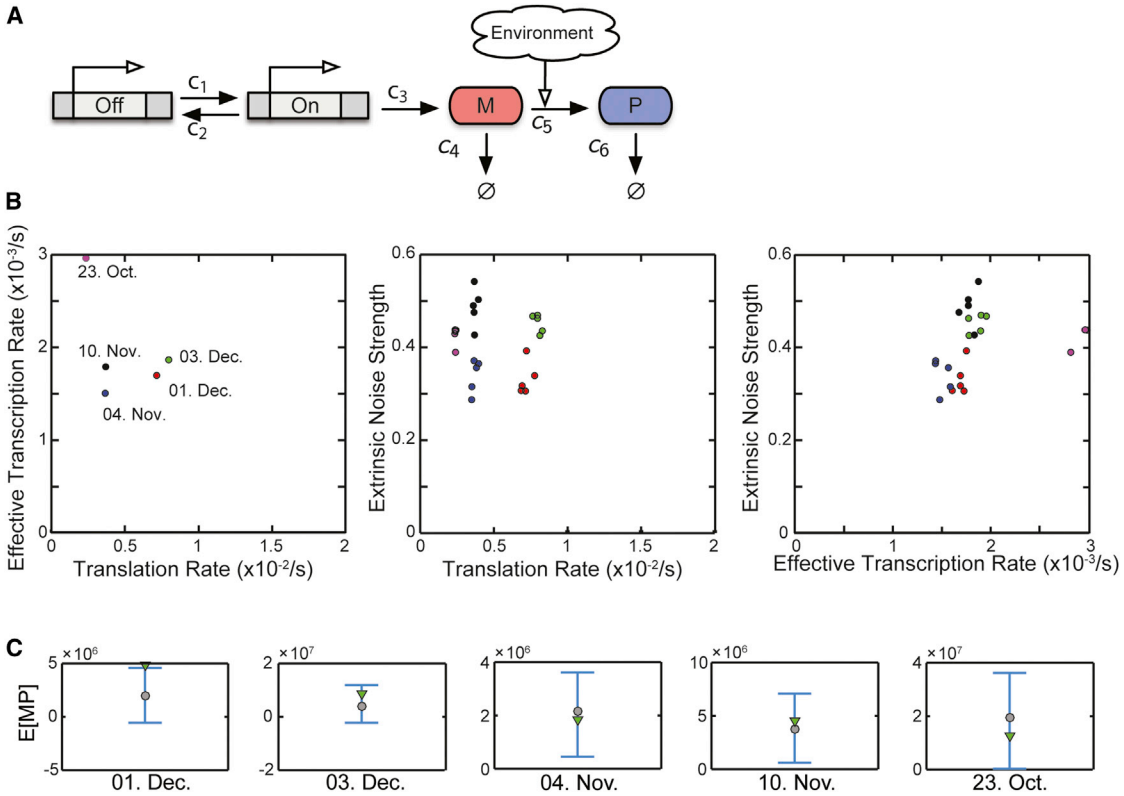


Figure 6. A Stochastic Model of CD147 Expression in Single Cells

(A) Schematic of the model. The gene stochastically switches between its inactive and active state, and mRNA can be transcribed only from the latter. The model comprises six kinetic parameters and an additional parameter describing the amount of extrinsic noise due to environment.

(B) Model parameters inferred from CD147 single-cell measurements shown in previous figures. The plot shows the inferred mean values for five independent repeats of the inference algorithm, indicating little variability across individual runs.

(C) Model validation. The inferred model was used to predict the correlation between mRNA and protein, i.e., E[M(t)P(t)]. The obtained results were compared against the experimentally measured correlation as revealed by our measurements. Markers indicate predicted (triangles) and empirical (circles) correlations, and the whiskers indicate 5% and 95% quantiles of the latter. In four out of five cases, the predictions agree well with the experimental data. Posterior estimates of the fitted model parameters can be found in Table S1.

protein or mRNA are quantified. While the use of such practical techniques provide familiarity and wider use by the community, integration of advanced microfluidic devices or pipetting robots would significantly increase the throughput. Another limitation of the PLA assay is its specific requirements on the used antibodies. Since two antibodies need to bind the antigen at the same time, and the exact position of an antibodies epitope is often unknown, it can be necessary to screen several antibody pairs until a working PLA probe set is found.

EXPERIMENTAL PROCEDURES

Construction of Calibration Curves for Digital PLA

For GFP and ICAM-1, a dilution series of each protein in 1X PBS (pH 7.4) was prepared; pure 1X PBS was also included in the workflow to account for background signal in the absence of protein. For CD147, the pure protein standard was diluted either in lysate dilution buffer (LDB) or 0.1X TM buffer instead of PBS; the diluent buffer was also included to account for background signal. 2 μl of each protein dilution was combined with 2 μl assay probe solution containing both proximity probes (at 400 pM each, thereby giving a final probe concentration of 200 pM) in 8-well PCR strips, and the mixture was incubated at 37°C for 1 hr. After proximity probe adsorption, the protein solutions were diluted 25-fold by adding 96 μl of ligation solution; the resultant solution was incubated at 37°C for 10 min. 2 μl of a diluted protease solution was added, and the mixture was incubated at 37°C for 10 min to terminate the ligation reaction; this step was followed by 15 min incubation at 95°C for inactivation of the protease. The proximity probes, ligation, and diluted protease solutions were prepared as per manufacturer's instructions.

For the (ddPCR) step, 9 μl of the resultant PLA solutions (or 1 μl in the case of ICAM-1) was combined with 10 μl of 2X ddPCR Supermix for Probes (Bio-Rad) and 1 μl of 20X Universal PCR Assay (Life Technologies) solution. For ICAM-1 quantification, 8 μl of deionized sterile water was added to bring the total volume to 20 μl. All of the solutions were mixed by pipetting up and down multiple times. The resultant 20 μl ddPCR solutions were transferred to DG8 cartridges, emulsified by the QX100 Droplet Generator (Bio-Rad), and the emulsions were placed in a Veriti thermal cycler (Life Technologies) for PCR. The temperature schedule for PCR was: 1X, 95°C for 10 min; 40X, 94°C for 30 s followed by 60°C for 1 min; 1X, 98°C for 10 min; and the ramp speed was 2.5°C/s. After, the emulsions were analyzed using the QX100 Droplet Reader and QuantaSoft software (Bio-Rad, v.1.3.2.0). Fluorescence from the emulsion droplets was quantified in the Absolute Quantification setting, and the signal threshold was manually set by applying to all wells the threshold value determined by automatic analysis of one of the most concentrated samples (Figures S1A and S1B). Once the threshold is set, QuantaSoft uses the following equation

to convert the numbers of positive and negative droplets (i.e., droplets with and without DNA) into concentration (Pinheiro et al., 2012):

$$C = \frac{-1,000}{V_d} \ln\left(1 - \frac{P}{R}\right),$$

where C denotes the DNA concentration in number of molecules per μl of ddPCR solution, V_d denotes the mean droplet volume in nl, P denotes the number of droplets with DNA (i.e., positive droplets), and R denotes the number of total (i.e., positive and negative) droplets. In the "merged" analysis setting, QuantaSoft gives both concentration and 95% confidence limits for a series of technical replicates. The SD of the DNA concentration can be calculated using this equation:

$$SD = \frac{CI_{max} - CI_{min}}{2 * 1.96},$$

where SD denotes standard deviation of the DNA calculation and CI_{max} and CI_{min} denote the upper and lower limits of the 95% confidence interval (given as "TotalConfMax" and "TotalConfMin" in QuantaSoft), respectively.

It has recently been reported that the used version of QuantaSoft software is using a wrong droplet volume (0.91 nl instead of 0.85 nl) to calculate concentrations (Corbisier et al., 2015), introducing a 10% error in the DNA concentration. By making calibration curves, however, we already correct our measurements for such errors. Therefore, the reported protein concentrations are not affected by the error in droplet volume. The same is true for the mRNA measurements, where the measured recovery rate accounts for this error as well.

The assay LOD of 107 GFPs per μl ddPCR solution corresponds to 24,200 GFP-p65s per cell due to the 113-fold dilution in the digital PLA workflow. Individual RAW macrophages were sorted into and lysed in a 2 μl solution, which was then diluted 51-fold prior to ddPCR. 9 μl of the PLA solution was added to the ddPCR mixture, whose final volume was 20 μl. The successive dilutions can be summarized as:

$$\frac{\text{Total Protein}}{2\ \mu l} \times \frac{1}{51} \times \frac{9}{20} = [\text{Amplicon}].$$

In the case of joint mRNA/protein detection where only 2 μl of the initial 3 μl are used for PLA, an additional 1.5-fold dilution needs to be taken in account.

Different proximity probe concentrations were used in optimization experiments shown in Figures S1C and S1D. In addition, after the probe adsorption step, the protein solutions were diluted either 2- or 25-fold by adding 4 or 96 μl of ligation solution, respectively. Component concentrations in the ligation solutions were adjusted accordingly to obtain the standard ligase and buffer concentrations in the final mixture.

Protein Quantification from Mammalian Cells

2 μl cell lysate (instead of the diluted protein solution) was taken through the PLA workflow as explained above. For CD147 protein quantification from HEK cells, 2 μl of cell lysate (3 μl total) was processed in digital PLA, while the remaining 1 μl was used in CD147 mRNA quantification. Error in absolute single-cell protein abundances was calculated by multiplying the single-cell protein amount with the coefficient of variance (CV) of the measured DNA concentration after digital PLA. A separate calibration curve was constructed on the same day as single-cell CD147 abundances were quantified to account for (albeit small) day-to-day variation in assay sensitivity and background (Figure S1G). A representative example of a daily CD147 calibration curve, together with the measured single cell data can be found in Figure S3E. The GFP-p65 values from individual cells were calculated using a GFP calibration curve from a single day. GFP calibration curves were highly variable on different days (Figure S1F), due to different aggregation levels in protein standards (Krasowska et al., 2010) (M. Shannon and D. Ruff, personal communication). CD147 was thus chosen as the model protein for further method development.

CD147 mRNA Quantification from Human Cells

1 to 100 Tlr4-GFP HEK293T cells were sorted by FACS into 3 μl of 0.1X TM lysis buffer for CD147 mRNA and protein quantification. 1 μl lysate was taken through the two-step RT-PCR workflow, while the remaining 2 μl of cell lysate were used for protein quantification as explained above. Probe 2 end primers and hydrolysis probe (i.e., CD147 TaqMan Gene Expression Assay, Cat. No. Hs00174305_m1) were used, and 40 cycles of amplification were conducted during the ddPCR step (Figures S4A–S4C). Fluorescence from the emulsion droplets was quantified using the QuantaSoft software; the obtained DNA concentrations and SDs were converted to single-cell CD147 mRNA values, accounting for the mRNA recovery rate determined using a synthetic CD147 mRNA standard (87.6%, see Supplemental Experimental Procedures). When multiple cells were sorted into a single well, the mRNA counts and the error were divided by the number of cells sorted to report the values on a per cell basis.

Error Bars

Unless stated otherwise, data are presented as mean ± SD.

Supplemental Experimental Procedures

Detailed information about "Reagents," "Cell Culture," "Fluorescence-activated sorting of mammalian cells," "Bulk protein lysate preparation," "Transfection of 3T3 cells," "Bulk RNA extraction," "Optimization of CD147 mRNA quantification," "Generation of artificial RNA standard," "Fitting of probability density functions (pdf)," "Derivation of gene burst rate and burst size from the parameters of a gamma distribution," "Moment-based modeling of CD147 expression," and "Inference of model parameters" can be found in the Supplemental Experimental Procedures section.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, one table, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.02.030.

AUTHOR CONTRIBUTIONS

C.A., C.A.J., C.A.B., and S.T. designed the research. C.A., C.A.J., and J.L. performed the experiments and analyzed the data. C.Z. and M.K. built the two-state stochastic model and conducted the computational analysis. C.A., C.A.J., C.Z., M.K., and S.T. wrote the manuscript.

ACKNOWLEDGMENTS

HEK293T cells were a kind gift from Dr. Kensuke Miyake at the University of Tokyo. We thank Drs. Mark Shannon and David Ruff for helpful advice on PLA and Verena Jäggin and Telma Lopes for their assistance with FACS. We thank the ETH Quantitative Genomics Facility for the access to their ddPCR setup and technical help. We also thank Professors Emre Araci, Sven Panke, and Sai Reddy for their insightful comments. This work was supported by an ERC Starting Grant (SingleCellDynamics) to S.T. and by Schweizerischer Nationalfonds, NCCR Molecular Systems Engineering, and a SystemsX Grant (NeuroStemX).

Received: September 18, 2015
Revised: January 19, 2016
Accepted: February 25, 2016
Published: March 17, 2016

REFERENCES

Agasti, S.S., Liong, M., Peterson, V.M., Lee, H., and Weissleder, R. (2012). Photocleavable DNA barcode-antibody conjugates allow sensitive and multiplexed protein analysis in single cells. J. Am. Chem. Soc. *134*, 18499–18502.

Allison, K.R., Brynildsen, M.P., and Collins, J.J. (2011). Heterogeneous bacterial persisters and engineering approaches to eliminate them. Curr. Opin. Microbiol. *14*, 593–598.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

Blake, W.J., KAErn, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. Nature *422*, 633–637.

Blazek, M., Betz, C., Hall, M.N., Reth, M., Zengerle, R., and Meier, M. (2013). Proximity ligation assay for high-content profiling of cell signaling pathways on a microfluidic chip. Mol. Cell. Proteomics *12*, 3898–3907.

Caldwell, A.B., Cheng, Z., Vargas, J.D., Birnbaum, H.A., and Hoffmann, A. (2014). Network dynamics determine the autocrine and paracrine signaling functions of TNF. Genes Dev. *28*, 2120–2133.

Corbisier, P., Pinheiro, L., Mazoua, S., Kortekaas, A.-M., Chung, P.Y.J., Gerganova, T., Roebben, G., Emons, H., and Emslie, K. (2015). DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials. Anal. Bioanal. Chem. *407*, 1831–1840.

Fan, H.C., Gu, W., Wang, J., Blumenfeld, Y.J., El-Sayed, Y.Y., and Quake, S.R. (2012). Non-invasive prenatal measurement of the fetal genome. Nature *487*, 320–324.

Flatz, L., Roychoudhuri, R., Honda, M., Filali-Mouhim, A., Goulet, J.-P., Kettaf, N., Lin, M., Roederer, M., Haddad, E.K., Sékaly, R.P., and Nabel, G.J. (2011). Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines. Proc. Natl. Acad. Sci. USA *108*, 5724–5729.

Frei, A.P., Bava, F.-A., Zunder, E.R., Hsieh, E.W.Y., Chen, S.-Y., Nolan, G.P., and Gherardini, P.F. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat. Methods *13*, 269–275.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025–1036.

Gordon, A., Colman-Lerner, A., Chin, T.E., Benjamin, K.R., Yu, R.C., and Brent, R. (2007). Single-cell quantification of molecules and rates using open-source microscope-based cytometry. Nat. Methods *4*, 175–181.

Gullberg, M., Gústafsdóttir, S.M., Schallmeiner, E., Jarvius, J., Bjarnegård, M., Betsholtz, C., Landegren, U., and Fredriksson, S. (2004). Cytokine detection by antibody-based proximity ligation. Proc. Natl. Acad. Sci. USA *101*, 8420–8424.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal. Chem. *83*, 8604–8610.

Hindson, C.M., Chevillet, J.R., Briggs, H.A., Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat. Methods *10*, 1003–1005.

Huang, B., Wu, H., Bhaya, D., Grossman, A., Granier, S., Kobilka, B.K., and Zare, R.N. (2007). Counting low-copy number proteins in a single cell. Science *315*, 81–84.

Hughes, A.J., Spelke, D.P., Xu, Z., Kang, C.-C., Schaffer, D.V., and Herr, A.E. (2014). Single-cell western blotting. Nat. Methods *11*, 749–755.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell *150*, 389–401.

Ke, R., Nong, R.Y., Fredriksson, S., Landegren, U., and Nilsson, M. (2013). Improving precision of proximity ligation assay by amplified single molecule detection. PLoS ONE *8*, e69813.

Kellogg, R.A., and Tay, S. (2015). Noise facilitates transcriptional control under dynamic inputs. Cell *160*, 381–392.

Kim, S.H., Iwai, S., Araki, S., Sakakihara, S., Iino, R., and Noji, H. (2012). Large-scale femtoliter droplet array for digital counting of single biomolecules. Lab Chip *12*, 4986–4991.

Krasowska, J., Olasek, M., Bzowska, A., Clark, P.L., and Wielgus-Kutrowska, B. (2010). The comparison of aggregation and folding of enhanced green fluorescent protein (EGFP) by spectroscopic studies. Spectroscopy (Springf.) *24*, 343–348.

Lu, Y., Xue, Q., Eisele, M.R., Sulistijo, E.S., Brower, K., Han, L., Amir, A.D., Pe'er, D., Miller-Jensen, K., and Fan, R. (2015). Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. Proc. Natl. Acad. Sci. USA *112*, E607–E615.

Lundberg, M., Eriksson, A., Tran, B., Assarsson, E., and Fredriksson, S. (2011). Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res. *39*, e102.

Ma, C., Fan, R., Ahmad, H., Shi, Q., Comin-Anduix, B., Chodon, T., Koya, R.C., Liu, C.-C., Kwong, G.A., Radu, C.G., et al. (2011). A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. Nat. Med. *17*, 738–743.

Paulsson, J. (2005). Models of stochastic gene expression. Phys. Life Rev. *2*, 157–175.

Peccoud, J., and Ycart, B. (1995). Markovian modeling of gene product synthesis. Theor. Popul. Biol. *48*, 222–234.

Pinheiro, L.B., Coleman, V.A., Hindson, C.M., Herrmann, J., Hindson, B.J., Bhat, S., and Emslie, K.R. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Anal. Chem. *84*, 1003–1011.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Rissin, D.M., Kan, C.W., Campbell, T.G., Howes, S.C., Fournier, D.R., Song, L., Piech, T., Patel, P.P., Chang, L., Rivnak, A.J., et al. (2010). Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. Nat. Biotechnol. *28*, 595–599.

Schallmeiner, E., Oksanen, E., Ericsson, O., Spångberg, L., Eriksson, S., Stenman, U.-H., Pettersson, K., and Landegren, U. (2007). Sensitive protein detection via triple-binder proximity ligation assays. Nat. Methods *4*, 135–137.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Shi, Q., Qin, L., Wei, W., Geng, F., Fan, R., Shin, Y.S., Guo, D., Hood, L., Mischel, P.S., and Heath, J.R. (2012). Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. Proc. Natl. Acad. Sci. USA *109*, 419–424.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. Nature *444*, 643–646.

Snijder, B., Sacher, R., Rämö, P., Damm, E.-M., Liberali, P., and Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. Nature *461*, 520–523.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat. Methods *3*, 995–1000.

Ståhlberg, A., Håkansson, J., Xian, X., Semb, H., and Kubista, M. (2004). Properties of the reverse transcription reaction in mRNA quantification. Clin. Chem. *50*, 509–515.

Ståhlberg, A., Thomsen, C., Ruff, D., and Åman, P. (2012). Quantitative PCR analysis of DNA, RNAs, and proteins in the same single cell. Clin. Chem. *58*, 1682–1691.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. Nature *466*, 267–271.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J., et al. (2004). Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Mol. Cell. Proteomics *3*, 960–969.

Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. *13*, 227–232.

Whale, A.S., Huggett, J.F., Cowen, S., Speirs, V., Shaw, J., Ellison, S., Foy, C.A., and Scott, D.J. (2012). Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Res. *40*, e82.

Yu, X.-L., Hu, T., Du, J.-M., Ding, J.-P., Yang, X.-M., Zhang, J., Yang, B., Shen, X., Zhang, Z., Zhong, W.-D., et al. (2008). Crystal structure of HAb18G/CD147: implications for immunoglobulin superfamily homophilic adhesion. J. Biol. Chem. *283*, 18056–18065.

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., and Koeppl, H. (2012). Moment-based inference predicts bimodality in transient gene expression. Proc. Natl. Acad. Sci. USA *109*, 8340–8345.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

Molecular Cell, Volume *61*

Supplemental Information

Digital Quantification of Proteins and mRNA in Single Mammalian Cells

Cem Albayrak, Christian A. Jordi, Christoph Zechner, Jing Lin, Colette A. Bichsel, Mustafa Khammash, and Savaş Tay

Supplemental Figures

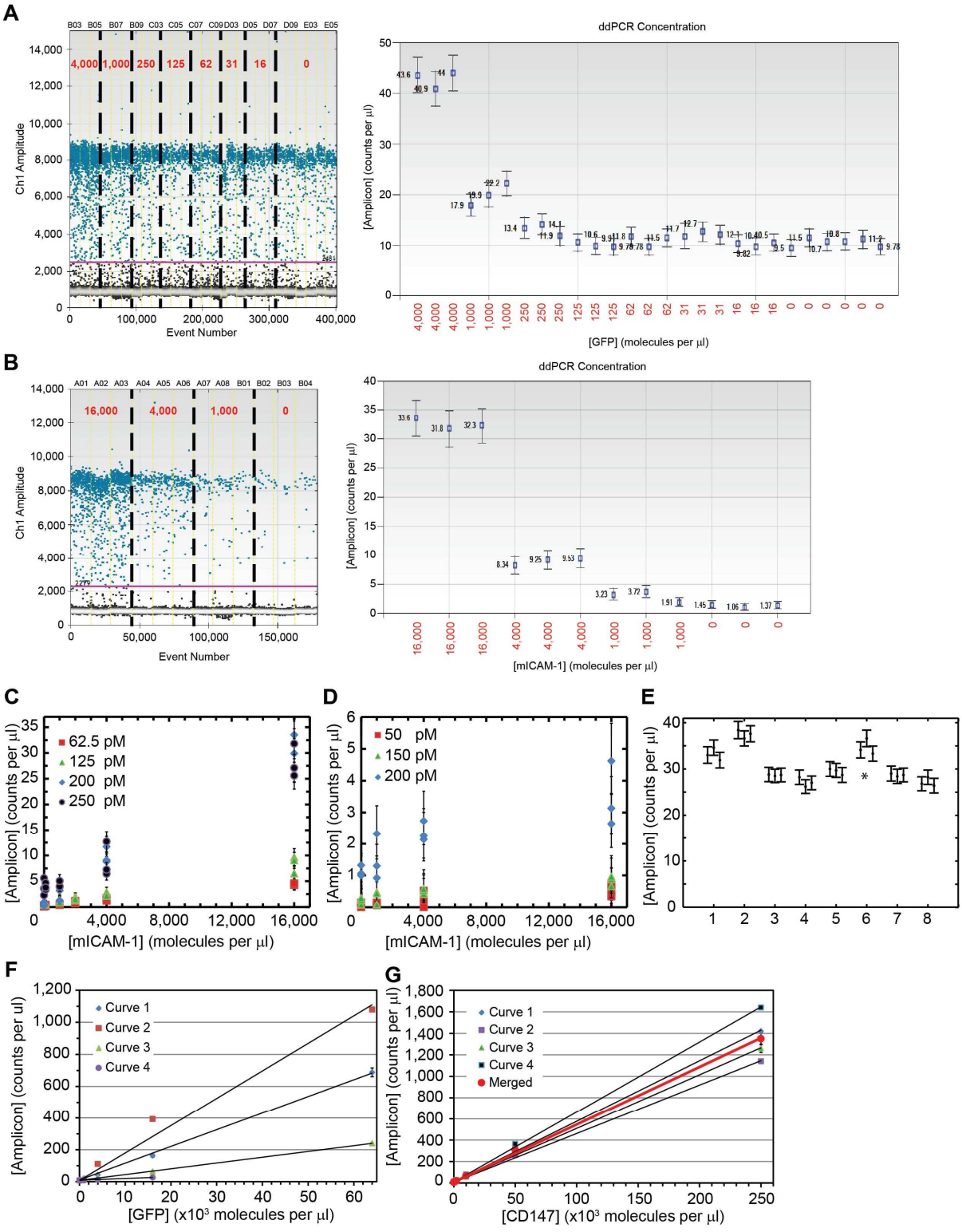


Figure S1, associated with Figure 2 One-dimensional (1D) droplet fluorescence and concentration plots from GFP **(A)** and ICAM-1 **(B)** dilution series. Horizontal magenta lines in 1D droplet plots indicate the fluorescence threshold values, above which a ddPCR droplet is considered positive. The red numbers indicate the number of proteins per microliter in each dilution of the protein standard. QuantaSoft software converts the number of positive droplets to DNA concentrations (right-hand side plots). Raw data shown in parts **(A)** and **(B)** were used in Figures 2A, 2C and 2D. Error bars, Poisson error from 10,500 to 16,470 droplets. **(C-D)** Optimization of the digital PLA protocol. Serial dilutions of pure ICAM-1 were used to evaluate assay performance at different proximity probe

concentrations with either the standard 25-fold **(c)** or 2-fold **(D)** dilution during the ligation step. 200 pM of proximity probes provided good sensitivity without high background signal **(C)**. digital PLA with only a 2-fold dilution during the ligation step yielded poor signal regardless of proximity probe concentration. Error bars, combined error from both Poisson noise and technical error from replicates (n=3). **(E)** ddPCR replicates of eight different PLA samples. Error bars originate from ddPCR error. (*) used to label samples containing error unexplained by ddPCR, as determined by QuantaSoft software. **(F)** Day-to-day variability in GFP calibration curves. Pure GFP standard was serially diluted in PBS, and different dilutions were measured by digital PLA on four different days. Error bars, combined error from both Poisson noise and technical error from replicates (n=6 for buffer background of Curve 1, n=3 for nonzero GFP dilutions of Curve 1, n=2 for other data points). **(G)** CD147 calibration curves in Lysate Dilution Buffer (LDB) matrix. Pure CD147 standard was serially diluted in LDB, and different dilutions were measured by digital PLA on three different days (2 calibration curves were measured on one day). Error bars, combined error from both Poisson noise and technical error from replicates (n=6 for LDB background, n=3 for nonzero CD147 dilutions). The merged calibration curve (red) was constructed by combining all available data (n=24 for LDB background, n=12 for nonzero CD147 dilutions).

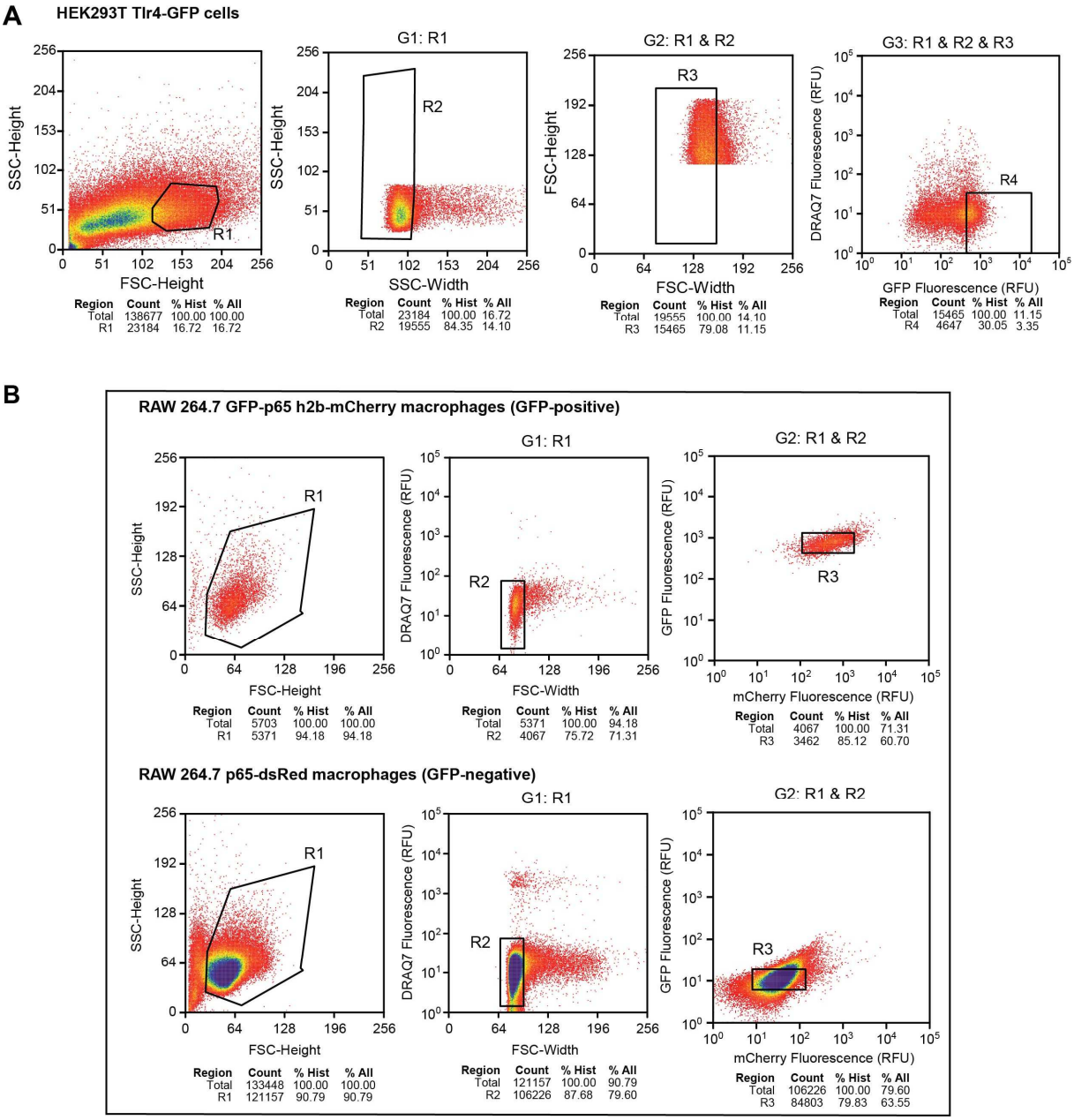


Figure S2, associated with Figure 3. Single cell protein measurements, FACS raw data. **(A)** Fluorescence-activated sorting of HEK293T mammalian cells expressing the Tlr4-GFP fusion protein. Four nested gates were used and are shown from left to right, in order of application. DRAQ7 nuclear dye was used to distinguish viable cells from non-viable ones. CD147 mRNA and protein of the cells within the R4 population were analyzed. **(B)** Fluorescence-activated sorting of RAW 264.7 macrophages with (top row) and without GFP expression (bottom row). The GFP-negative cells were included as negative control. Three nested gates were used during sorting and are shown form left to right, in order of application; the middle graphs show cells within gate R1, while the graphs on the right-hand side show cells within gates R1 and R2. The DRAQ7 nuclear dye was used to segregate viable cells from dead ones, where nuclei in non-viable cells are stained with DRAQ7 and thus display higher fluorescence.

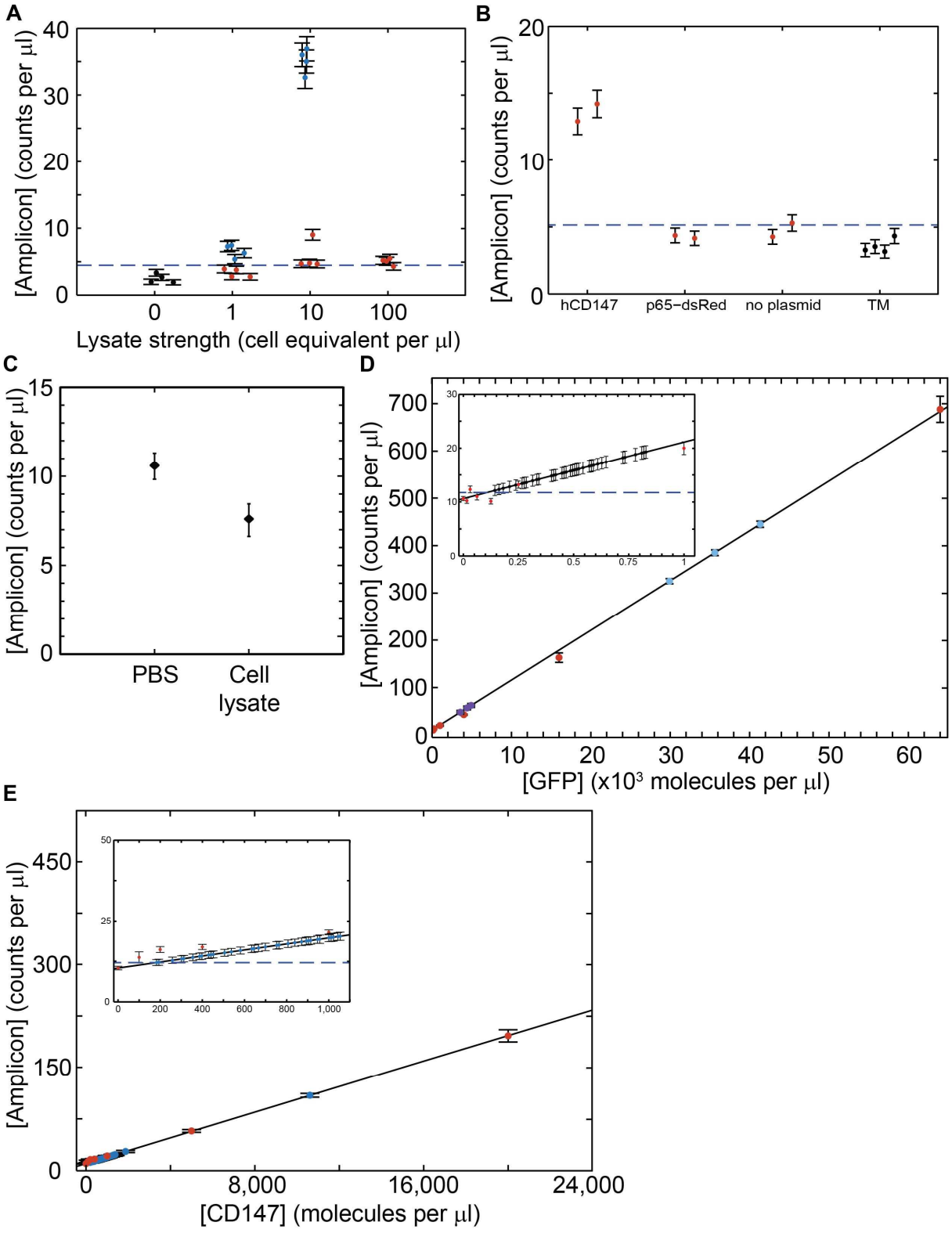


Figure S3, associated with Figure 3. Technical details of single cell digital PLA: **(A)** digital PLA signal from lysates of 3T3 cells (murine, red), HEK 293T cells (human, blue), or from, pure 0.1x TM buffer (black), using anti-hCD147 probes. The limit of detection is indicated by the dashed blue line. Error bars show ddPCR error of individual measurements **(B)** Digital PLA signal produced by anti-hCD147 probes used to analyze lysates of 3T3 cells. The 3T3 cells originated from mixed populations that had been either transfected with plasmids expressing hCD147 or p65-DsRed, or treated with plain transfection reagents. 0.1x TM buffer was used to estimate the limit of detection (dashed blue line). Error bars show the Poisson error of individual measurements originating from ddPCR. **(C)** Single-cell lysate matrix does not increase background signal. Two GFP-free solutions were taken through the digital PLA workflow to assess effect of buffer matrix on digital PLA signal, namely phosphate-

buffered saline (PBS) or lysate of GFP-negative macrophages diluted to 1-cell equivalent in PBS (Cell lysate). Error bars, combined error from both Poisson noise and technical error from replicates (n=6). **(D)** GFP calibration curve used to calibrate single cell measurements. Signal deriving from protein standards (red), single cells (black), 10 cells (purple), and 100 cells (blue), are shown. All samples lie in the linear range of the assay. **(E)** CD147 calibration curve used to calibrate measurements of '03. Dec.'. Shown are protein standards (red), and single cells (black). 100 cell populations (not shown) lie outside the calibrated range around 23,000/μL – 50,000/μL. In both calibration curves, the error bars of single cell measurements originate from ddPCR error only. The error bars of protein standards consist of combined ddPCR and technical error from three replicates.

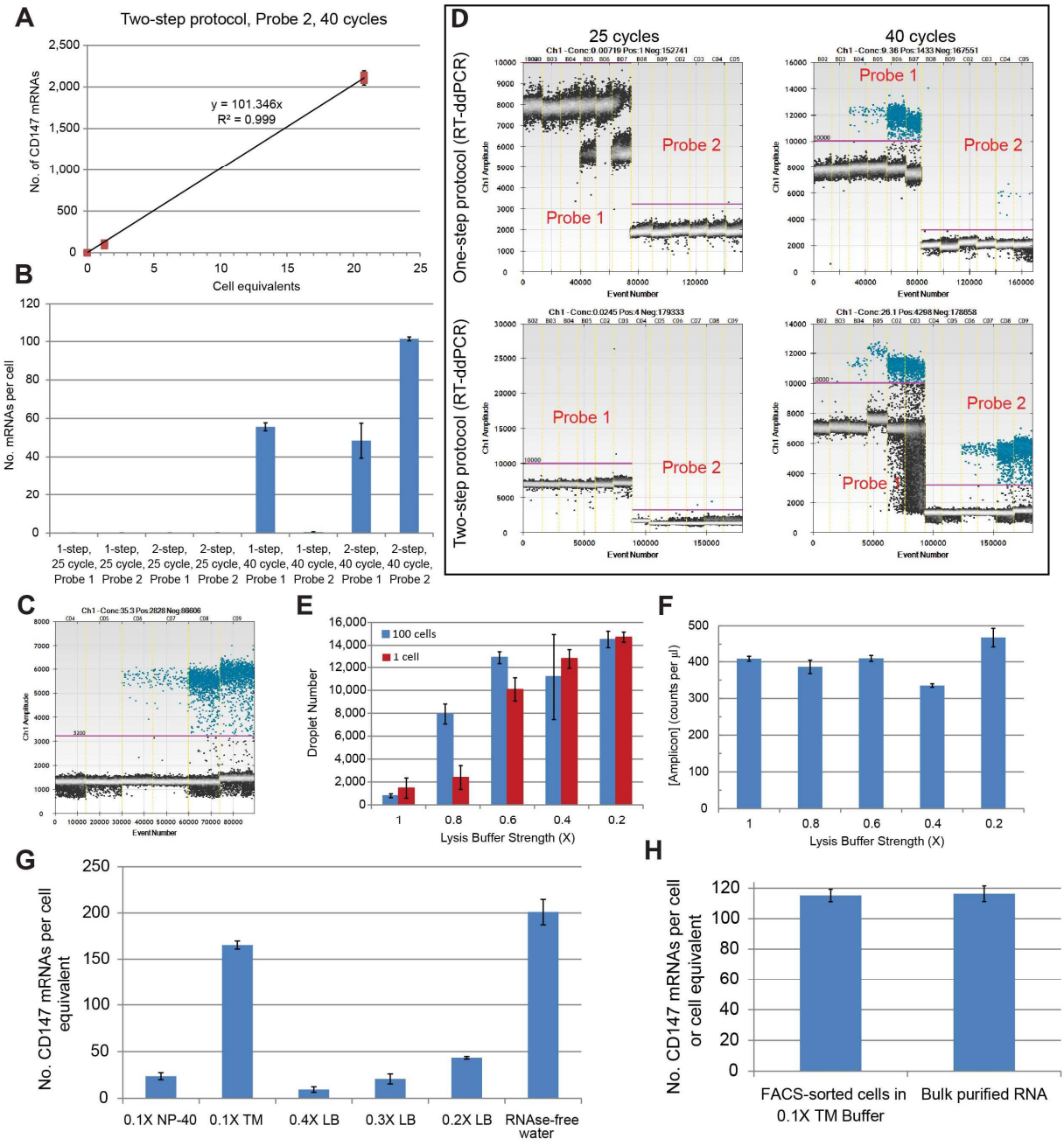


Figure S4, associated with Figure 4. Optimization of CD147 mRNA quantification. Total RNA was bulk-purified from HEK293T cell lysates, and permutations of RT-PCR protocols (1-step vs. 2-step), primers and hydrolysis probes (Probes 1-3), and PCR cycle numbers (25 vs. 40) were tested. **(A)** CD147 mRNA in different dilutions of bulk HEK lysate (in units of cell equivalents) was quantified using the combination of a two-step (reverse transcription followed by ddPCR, RT-ddPCR) protocol, Probe 2, and 40 amplification cycles in ddPCR. Error bars show ddPCR error of individual measurements **(B)** Slopes of the CD147 mRNA dilution curves (such as those in **A**) are equal to the number of CD147 mRNAs per cell equivalent. Probe 3 (i.e. custom-designed primers and probe targeting the CD147 mRNA) failed to yield any ddPCR signal, and was thus omitted from the plot. Error bars, standard error of the slope from the standard linear regression fit. **(C)** One-dimensional (1D) droplet fluorescence plot for the optimal workflow for CD147 mRNA quantification (i.e. 2-step, 40 cycle, Probe 2 combination). Horizontal magenta line (Fluorescence = 3200) indicates the threshold that separates ddPCR droplets with CD147 amplicon from those without. **(D)** 1D droplet fluorescence plots from tested permutations of RT-PCR protocols, primers and hydrolysis probes and PCR cycle numbers. Horizontal magenta lines indicate the fluorescence threshold values that separate ddPCR droplets with amplicon from those without. Droplets on the right-hand side of the bottom right plot were shown in **(c)**. Evaluation of different standard Lysis Buffer (LB) dilutions for **(E)** ddPCR compatibility and **(F)** adequate lysis. **(E)** 1 or 100 HEK293T cells were sorted into indicated dilutions of the standard LB, and the number of analyzed droplets (out of 20,000 total) after RT-ddPCR (i.e. the 2-step protocol) are plotted. Error bars, standard deviation from technical replicates (n=4). **(F)** 100 HEK293T cells were

sorted into indicated LB dilutions, and the digital PLA signal was measured. Error bars, combined error from both Poisson noise and technical error from replicates (n=4). **(G)** Evaluation of different lysis buffer formulations for two-step RT-ddPCR compatibility. Total RNA was bulk-extracted from HEK293T cells, diluted in the indicated buffers (i.e. 0.1X NP-40 lysis buffer, 0.1X TM lysis buffer, 0.2 to 0.4X standard lysis buffer) or RNAse-free water, and the CD147 mRNA quantified using the two-step RT-ddPCR protocol. Error bars, combined error from both Poisson noise and technical error from replicates (n=3 for different lysis buffers, n=6 for RNAse-free water). **(H)** FACS sorting and lysis buffer components in 0.1X TM buffer do not interfere with CD147 mRNA quantification using the two-step RT-ddPCR protocol. 100 cells were sorted by FACS into 0.1X TM lysis buffer, and the CD147 mRNA was quantified using the 2-step RT-ddPCR protocol. This measurement is compared to CD147 mRNA quantified from bulk-lysed HEK293T cells. Error bars, combined error from both Poisson noise and technical error from replicates (n=4).

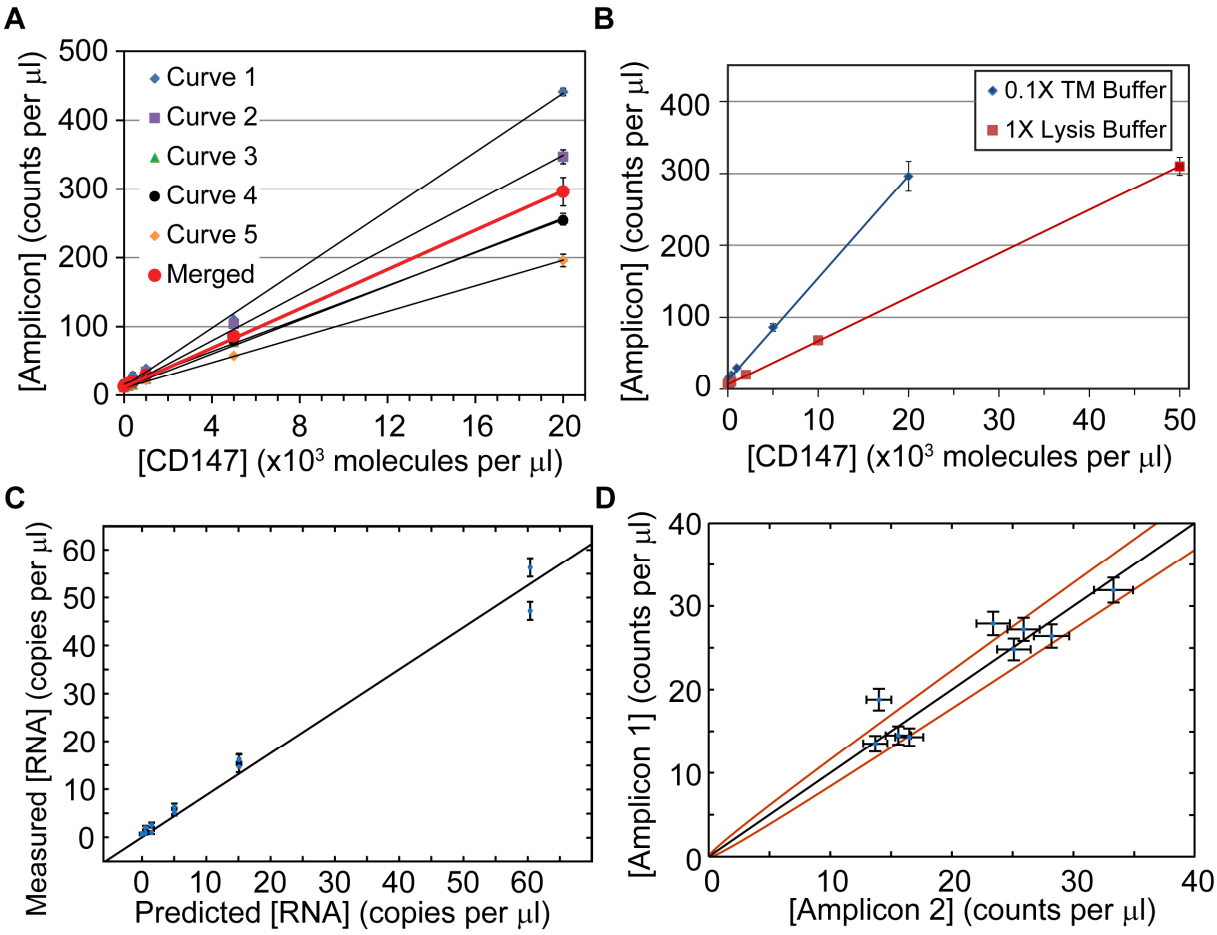


Figure S5, associated with Figure 4. Characterization of joint digital PLA RT-ddPCR workflow. **(A)** CD147 calibration curves in 0.1X TM lysis buffer matrix. Pure CD147 protein was serially diluted in 0.1X TM lysis buffer, and the various dilutions were measured by digital PLA on five different days. Error bars, combined error from both Poisson noise and technical error from replicates (n=6 for 0.1X TM buffer background, n=3 for nonzero CD147 dilutions). The merged calibration curve (red) was constructed by combining all available data (n=30 for 0.1X TM buffer background, n=15 for nonzero CD147 dilutions). **(B)** Head-to-head comparison of merged calibration curves in two different buffer matrices. The merged calibration curve for lysate dilution buffer is also shown in Figure S1G. The merged curve for 0.1x TM buffer is also shown in Figure S4A. Error bars are consist of Poisson error and technical error from replicates (n=24-30 for buffer background, and n=12-15 for nonzero CD147 dilutions). **(C)** RT-ddPCR calibration curve made using *in-vitro* transcribed CD147 mRNA. Absorbance intensity (λ=260nm) was used to predict the RNA concentration, which was compared to RT-ddPCR measurements. Error bars originate from ddPCR error of individual measurements. Solid line indicates linear regression with formula: $y=0.876x$, $R^2=0.988$. **(D)** Comparison of CD147 digital PLA signal originating from two equally split fractions of HEK 293T single cell lysate. The black curve indicates perfect agreement between both half-lysates. Red curves indicate the 95% confidence interval of ddPCR measurement error, as described by Dube et al., 2008. Error bars indicate ddPCR error of individual measurements.

Table S.1, associated with Figure 6. Experimental parameters of the filter cascade. Entries denote mean posterior estimates and their standard deviations.

Exp	c_T (1/s)	c_4 (1/s)	c_5 (1/s)	c_6 (1/s)	σ_Z
01. Dec.	1.536e-03 (6.380e-04)	9.367e-06 (3.860e-06)	8.150e-03 (3.523e-03)	6.066e-06 (2.598e-06)	8.989e-02 (4.633e-02)
03. Dec.	1.617e-03 (7.427e-04)	9.254e-06 (4.105e-06)	8.897e-03 (4.157e-03)	5.913e-06 (2.622e-06)	2.011e-01 (7.372e-02)
04. Nov.	1.235e-03 (5.485e-04)	9.535e-06 (4.141e-06)	4.508e-03 (2.015e-03)	5.534e-06 (2.414e-06)	1.175e-01 (4.340e-02)
10. Nov.	1.365e-03 (6.383e-04)	8.780e-06 (4.043e-06)	4.530e-03 (1.967e-03)	6.156e-06 (2.554e-06)	1.979e-01 (1.089e-01)
23. Oct.	2.034e-03 (9.816e-04)	9.362e-06 (4.229e-06)	3.154e-03 (1.511e-03)	5.885e-06 (2.542e-06)	2.185e-01 (1.743e-01)

Supplemental Experimental Procedures

Reagents

All of the reagents were purchased from Sigma (St. Louis, MO) unless indicated otherwise. Low-adsorption 8-well PCR strips (Catalog No. AB-0266) and 1.5 ml microfuge tubes (Catalog No. AM12450) were purchased from Thermo Fisher Scientific and Ambion, respectively; these low-adsorption containers were used for all of the experiments. Proximity ligation assay (i.e. TaqMan protein assay) and droplet digital PCR reagents and supplies were purchased from Life Technologies and Bio-Rad, respectively.

Pure murine intracellular adhesion molecule-1 (ICAM-1, Catalog No. 796-IC), biotinylated anti-murine ICAM-1 polyclonal goat antibody (Catalog No. BAF796), human cluster of differentiation 147 (CD147, Catalog No. 972-EMN), and biotinylated anti-human CD147 polyclonal goat antibody (Catalog No. BAF972) were purchased from R&D Systems. Pure green fluorescent protein (GFP, Catalog No. AB84191) and biotinylated anti-GFP polyclonal goat antibody (Catalog No. AB6658) were purchased from Abcam. Total Protein Extraction Kit, which contains the TM buffer and a 50X protease inhibitor, was purchased from Biochain.

AmpliTaq Gold® 360 Master Mix (Catalog No. 4398901), MEGAscript® T7 Transcription Kit (Catalog No. AM1333M), were purchased from ThermoFisher Scientific. DNA clean and concentrator kit 5 (Catalog No. D4003) was bought from Zymo Research. Custom DNA oligos were ordered from Microsynth (Switzerland).

Lipofectamine® 3000 Transfection Reagents were obtained from ThermoFisher Scientific (Catalog No. L3000-001). A plasmid expressing hCD147 isoform 2 under the control of the hEF1-HTLV promotor (pUNO1-hCD147b) was ordered from InvivoGen.

Lysis solution preparation

Full-strength (1X) standard lysis buffer (LB) used to lyse murine RAW 264.7 macrophages (Wall et al., 2009) was prepared by combining 100 parts of 2X Cell Lysis Reagent (Life Technologies) with 1 part Protease Inhibitor Cocktail I (Merck Millipore) and 1 part Phosphatase Inhibitor Cocktail II (Merck Millipore), and further diluting this mixture 1:1 with Cell Resuspension Buffer (CRB, Life Technologies).

Three different lysis solutions were tested for CD147 mRNA and protein quantification from HEK293T cells, namely the standard lysis buffer (LB), the NP-40 Cell Lysis Buffer (Invitrogen), and the TM lysis buffer. Different dilutions of the standard LB (0.2 to 1X, in 0.2X increments) were formulated by appropriately reducing the amount of 2X Cell Lysis Reagent; Lysate Dilution Buffer (Applied Biosystems) was added to complete the volume. Concentrations of protease and phosphatase inhibitors were not changed in dilute LB formulations. 0.1X NP-40 lysis buffer contains 1 mM phenylmethylsulfonyl fluoride (PMSF, PanReac AppliChem), Protease Inhibitor Cocktail I (at the same concentration as in standard LB), 10-fold diluted NP40 Cell Lysis Buffer, and CRB. 0.1X TM lysis buffer contains 10-fold diluted TM Buffer (Biochain), protease inhibitor cocktail (Biochain), and CRB. CRB was added to the NP-40 and TM buffers to complete the volume. Finally, all lysis buffers were supplemented with a ribonuclease inhibitor (RNAseOUT, Invitrogen) to prevent RNA degradation.

Cell culture

Murine GFP-p65 RAW 264.7 macrophages were transfected with the Addgene plasmid 20972 (Nam and Benezra, 2009) to insert the Histone cluster 1-mCherry fusion protein coding sequence. The transfected cells were grown for a week; single cells from this culture were sorted; and a clone with normal growth, morphology and stable mCherry expression was selected. Murine RAW 264.7 macrophages with stable genomic dsRed integration were used as negative control during fluorescence-activated sorting of the dual-fluorescent strain. These cells, henceforth referred to as "GFP-negative macrophages", were created by transfecting murine RAW 264.7 macrophages with the p65-dsRed plasmid (Nelson et al., 2002) as explained above.

Both macrophage cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Catalog No. D6429) supplemented with 8.8% v/v fetal bovine serum (FBS, Catalog No. F9665), 0.88X GlutaMAX (Life Technologies, Catalog No. 35050-038) and 17.7 mM HEPES (Life Technologies). They were grown in disposable polystyrene flasks with filter caps (Greiner Bio-One), incubated in a humidified incubator at 37°C with 5% carbon dioxide, and passaged every 2-3 days to a final density of 10^5 cells per ml. Macrophage cultures used for single-cell protein quantification did not exceed 15 passages.

Tlr4-GFP human embryonic kidney (HEK293T) cells (Kobayashi et al., 2006) were a kind gift from Prof. Kensuke Miyake. These cells were cultured in high-glucose DMEM (Gibco, Catalog No. 41965-039) supplemented with 10% v/v fetal calf serum (FCS, Gibco, Catalog No. 26010-066), 100 U/ml penicillin G, and 100 μg/ml

streptomycin sulfate (Gibco, Catalog No. 15070063). They were grown and incubated as detailed above, and passaged every 3-4 days to a final density of 10^5 cells per ml. HEK293T cultures used for single-cell protein quantification did not exceed 30 passages.

The used mouse 3T3 cell line expressing p65-DsRed and H2B-GFP were described previously (Tay et al., 2010, Lee et al., 2009), and were cultured in the same high-glucose DMEM as the HEK293T cells, supplemented with 10% fetal bovine serum (FBS, Sigma Aldrich, Catalog No. F9665), 1x GlutaMax™ Supplement (ThermoFisher Scientific, Catalog No., 35050-038), and 100U/ml penicillin-streptomycin (ThermoFisher Scientific, Catalog No., 15140-122). The cells were passaged every 2-3 days when they reached ~90% confluency.

Fluorescence-activated sorting of mammalian cells

After culturing RAW 264.7 macrophages with two fluorescent reporters as explained, they were dislodged with Versene, washed twice in PBS, and resuspended in 2 ml PBS supplemented with 0.2% w/v bovine serum albumin (BSA) and 5 mM ethylenediaminetetraacetic acid (EDTA). The cell suspension was filtered and DRAQ7 nuclear dye (BioStatus, Catalog No. DR71000) was added (at a final concentration of 600 nM) to the cell suspension to stain dead cells before sorting.

Single, ten or hundred dual-fluorescent viable macrophages were sorted into wells containing 2 μl lysis solution using a MoFlo XDP sorter (Beckman-Coulter). Several precautions were taken to ensure that the sorted cells would make direct contact with 2 μl of fluid in the well bottom. The 96-well plate containing 2 μl of lysis solution per well was placed on top of a deep-well 96-well plate to reduce the distance between the nozzle and the lysis solution. The angle with which the sorted stream is diverted from the vertical was also minimized to reduce the chance of sorted cells hitting the walls rather than the bottom of the well. The sheath fluid was flowed at 60 psi, and a 70 μm nozzle was used. Two lasers (488 nm blue and 561 nm yellow-green) along with 529/28, 613/20 and 785/60 nm bandpass filters were used to create three nested gates (forward/side scatter, DRAQ7/forward scatter, GFP/mCherry), identify dual-fluorescent macrophages, and separate them from the ones without GFP-p65 (Figure S2B). Flow cytometry data were analyzed using Summit Software (Beckman Coulter). Immediately after sorting, the 96-well plates were briefly centrifuged, sealed with clear adhesive sheets, covered with Parafilm, and stored at -80°C.

A similar protocol was followed for sorting 1 to 100 HEK293T cells in 2 or 3 μl lysis solutions. HEK293T cells were dislodged from the culture flask using Accutase instead of Versene. The lysis buffer volume in each well was increased from 2 to 3 μl when both CD147 mRNA and protein were to be quantified from sorted cells. The same lasers and filters were used to create four nested gates (side scatter-height/forward scatter-height, side scatter-height/side scatter-width, forward scatter-height/forward scatter-width, DRAQ7/GFP) for sorting.

Bulk protein lysate preparation

To prepare the "cell lysate" sample for the background comparison experiment (Figure S3C), GFP-negative macrophages were first pelleted and washed 3 times in PBS. After the final wash and centrifugation, the cells were resuspended in 1 ml of CRB and counted using a Neubauer-improved hemocytometer (Marienfeld Superior). This cell suspension was diluted 1:1 with a lysis mixture that consisted of 100 parts 2X Cell Lysis Reagent, 1 part Protease Inhibitor Cocktail I, and 1 part Phosphatase Inhibitor Cocktail II. The resultant solution was incubated on ice for 10 minutes. (Complete lysis was confirmed by diluting the lysate in trypan blue and observing only blue-stained nuclei in the hemocytometer, data not shown.) The lysate was then diluted to 1-cell equivalent in PBS.

To create bulk lysate of 3T3, and HEK 293T cells, the cell pellets were resuspended in fresh medium after passaging. The concentrations of the suspension was determined with a hemocytometer, and a known amount of cells was centrifuged at 250xg for 3min. After the medium had been removed, the pellet was resuspended in 0.1x TM buffer (identical formulation as for single cell measurements), thoroughly resuspended and frozen at -20ºC. Before PLA the lysate was diluted to 50 cell equivalents per μL using 0.1x TM buffer.

Transfection of 3T3 cells

3T3 cells were seeded in a 24 Well plate (Nunclon, Catalog No. 142475) at a density of 40'000 cells per well, in regular culture medium. Afterwards, the cells were rested for at least 4h before starting the transfection procedure. Lipofectamine 3000 was used for the transfection of plasmids using the following protocol: For each transfected well 25μL of Optim-MEM I reduced serum medium (ThermoFisher Scientific, Catalog No. 31985-062) were mixed well with 1.5μL of Lipofectamine 3000 reagent. In parallel, 25uL of Opti-MEM were combined with 1μg of plasmid DNA and 2μL of the P3000 reagent. Thereafter, both solutions were incubated for 5min at room temperature, and then added carefully to the attached cells. As negative controls, plain transfection reagents

without plasmid, and plain culture medium were used. The cells were harvested and lysed, using 0.1x TM buffer, 2.5 days after transfection

Bulk RNA extraction

After culturing HEK293T cells as explained, they were dislodged with Versene, sedimented at 300 g for 5 min, resuspended in 5 ml medium, and counted using a hemocytometer. Total RNA from 100,000 viable cells was extracted using the RNeasy Mini Kit (Qiagen, Catalog No. 74104) as per manufacturer's protocol.

Optimization of CD147 mRNA quantification

Three different combinations of end primers and probe were tested with two different RT-PCR protocols (1-step vs. 2-step protocol). Two of the three primer-probe combinations were purchased from Applied Biosystems (CD147 TaqMan Gene Expression Assays). Assays with catalog numbers Hs00936295_m1 and Hs00174305_m1 were designated Probe 1 and Probe 2, respectively; 6-carboxyfluorescein (FAM) and a minor groove binder (MGB) molecule are conjugated to the probes. The third combination (i.e. Probe 3) was custom-designed to target CD147 mRNA sequence; the sequences are as follows: forward primer: 5'- GAG TTC AAG GTG GAC TCC GAC G -3', reverse primer: 5'- GAT GTG TTC TGA CGA CTT CAC AGC C -3', probe: 5'- HEX-CAC GGC CAA CAT CCA GCT CCA CGG GC -BHQ-1 –3'. The 5' and 3' modifications indicated in the probe sequence are 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (HEX) and Black Hole Quencher-1 (BHQ-1), respectively.

For the 1-step protocol, 6 μl of RNA sample was added to a 20 μl solution containing 1X one-step RT-ddPCR supermix (Bio-Rad, diluted from a 2X stock), 1 mM manganese acetate, 900 nM of each of the end primers, and 250 nM of the hydrolysis probe. The mixture was emulsified as explained previously, and the emulsions were placed in a thermal cycler. The temperature schedule for RT-ddPCR was: 1X, 60°C for 30 min (for reverse transcription); 1X, 95°C for 5 min; 25 or 40X, 94°C for 30 s followed by 60°C for 1 min; 1X, 98°C for 10 min; and the ramp speed was 2.5°C/s. For the 2-step protocol, where reverse transcription (RT) is followed by ddPCR of the resultant cDNA, we used the GoScript Reverse Transcription System (Promega) and the standard ddPCR reagents, respectively. First, 1 μl of RNA sample was mixed with 1 μl of random primer solution, and 2.95 μl of nuclease-free water were added to bring the volume to 4.95 μl. This mixture was incubated at 70°C for 5 min, and then immediately chilled on ice for 5 min. After, it was supplemented with 1X GoScript reaction buffer, GoScript reverse transcriptase (diluted 20-fold from the stock solution), 1 U/μl ribonuclease inhibitor, 0.5 mM of each dNTP, and 3.1 mM magnesium chloride. The 9-μl RT solution was first incubated at 25°C for 5 min (for primer annealing), then at 42°C for 1 h (for reverse transcription), and finally at 70°C for 15 min (for reverse transcriptase inactivation). The resultant cDNA solution was then combined with 10 μl of 2X ddPCR Supermix for Probes (Bio-Rad), 900 nM of each of the end primers, and 250 nM of the hydrolysis probe; the final volume was 20 μl. This mixture was emulsified and the emulsions were placed in a thermal cycler. The temperature schedule was: 1X, 95°C for 10 min; 25 or 40X, 94°C for 30 s followed by 60°C for 1 min; 1X, 98°C for 10 min; and the ramp speed was 2.5°C/s. Fluorescence from the emulsion droplets was quantified using the QuantaSoft software as explained above. Different dilutions of the bulk-extracted total RNA were subjected to different permutations of primer-probe, RT-PCR protocols, and PCR amplification cycles. A calibration curve was constructed for each combination; and the number of mRNAs per cell equivalent was calculated from the slope of the calibration curve by linear regression. Standard error propagation was conducted to determine the error of the slope (Figures S4A-D).

We next tried CD147 mRNA quantification from FACS-sorted HEK293T cells using the optimal two-step protocol (RT-ddPCR). When 100 cells were sorted into full-strength (1X) digital PLA lysis buffer, ddPCR droplets were not properly formed. We hypothesized that the lysis buffer components interfered with droplet formation and tested different dilutions (0.2 to 1X) of the lysis buffer for RT-ddPCR compatibility and adequate lysis of 100 cells. 0.6X lysis buffer did not inhibit droplet formation (Figure S4E) while providing enough buffer strength to lyse 100 cells (Figure S4F); however, no CD147 mRNA was detected from 100 cells lysed in 0.6X buffer. Standard lysis buffer inhibited RT even at 0.6X strength; additional lysis buffers were thus evaluated for RT-ddPCR compatibility. All buffers except 0.1X TM buffer inhibited RT-ddPCR; only bulk RNA diluted in 0.1X TM buffer and RNAse-free water yielded high CD147 mRNA counts (Figure S4G). Digital PLA was also more sensitive in the new buffer matrix (1/slope = 70 proteins per count); assay LOD for CD147 was reduced from 200 to 86 proteins per μl ddPCR solution (16.2 fM in the sample solution, Table 1 and Figure S5A-B). When 100 cells were sorted by FACS into 0.1X TM lysis buffer, digital PCR analysis yielded CD147 mRNA numbers comparable to those obtained from bulk-purified total RNA (Figure S4H).

Production of full length CD147 mRNA

Total mRNA extracted from HEK293T cells was reverse transcribed using the GoScript Reverse transcription system, using the protocol outlined above.

Two PCR primers spanning a 1.5kbp fraction of the target CD147 mRNA isoform (Genebank: NM_198589.2) were designed, and a T7 promotor sequence (TAATACGACTCACTATAGG) was placed at the 5' end of the forward primer: CD147.full.rev: GGC GAT CTT TAT TGT GGC GG, CD147.full.fwd+T7: TAA TAC GAC TCA CTA TAG GGA GAG TAC ATG CGA GCG TGT GCG T. A 25 μL reaction comprising of 1x AmpliTaq Gold® 360 Master Mix, the two primers at an end concentration of 0.5 μM each, 2.5 μL 360 GC enhancer, less than 1 μg of the purified total cDNA, and nuclease free water, was used to PCR amplify the CD147 cDNA and simultaneously integrate the T7 promotor at the 5' end of the transcript. The used temperature program was 95 °C for 10min, followed by 30 cycles of 95 °C and 60 °C 130 sec, and a final incubation at 72 °C for 7 min.

Subsequently, the PCR product was purified using the DNA clean & concentrator 5 kit, and the identity of the sequence was confirmed using agarose gel electrophoresis (1% in TAE buffer) and Sanger sequencing (Microsynth, Switzerland). Afterwards, the Ambion MEGAscript® Kit was used for run-off transcription of the CD147 mRNA standard. Particularly, 1μg of DNA template was mixed with 2uL of ATP, UTP, CTP and GTP solution, 1x reaction buffer, 2 μL Enzyme mix in a total volume of 20 μL, adjusted with nuclease free water. This mixture was then incubated at 37°C for 4 h. Afterwards 1 μl of TURBO DNase was added to the solution, which was then well mixed and incubated another 15 min at 37°C. The produced RNA was purified using the Qiagen RNeasy Mini Kit, and eluted into nuclease free water. The RNA concentration of the eluate was determined using absorbance measurements at a wavelength of 260 nm, using a Nanodrop 2000c Spectrophotometer (Thermo Scientific). Finally, the RNA was snap frozen in liquid nitrogen and stored at -80°C until it was analyzed using the optimized RT-ddPCR protocol outlined above.

Fitting of probability density functions (pdf)

Protein and RNA data were analyzed using R (v3.1.0) (R Core Team, 2013). The function *fitdist* (Delignette-Muller and Dutang, 2015) was used to fit lognormal, gamma and Poisson pdfs to the empirical RNA distributions. Additionally, fits of the protein data were prepared using the *fitdistcens* function. Cells with copy numbers below the LOD were discarded. Additionally mRNA values above 600 and protein values above 600,000 were considered outliers, and not used for the fitting and derivation of mean/sd values. Goodness of fit was evaluated using a χ^2 test comparing the predicted and measured number of cells in each bin of the plotted histograms. Shape and rate parameters were used to calculate mean and standard deviations of the fitted gamma distributions.

Derivation of gene burst rate and burst size from the parameters of a gamma distribution

Analytical mRNA distributions from a two-state model of gene expression have previously been derived; the mRNA abundances fit a gamma distribution function, if the gene is mostly in the "off" state, and the expression bursts are shorter than the mRNA lifetime (Peccoud and Ycart, 1995; Raj et al., 2006). In this case, the shape parameter of the gamma distribution (α) corresponds to the ratio of the activation rate over the mRNA degradation rate (c_1/c_4), while the rate parameter (β) corresponds to the ratio of the gene inactivation rate to the transcription rate (c_2/c_3, i.e. the reciprocal of the burst size, that is, the average number of molecules produced in one activation):

$$P(m) = \frac{\left(\frac{c_2}{c_3}\right)^{\frac{c_1}{c_4}}}{\Gamma\left(\frac{c_1}{c_4}\right)} m^{\left(\frac{c_1}{c_4}-1\right)} e^{-\frac{c_2}{c_3}m} = \frac{\beta^{\alpha}}{\Gamma(\alpha)} m^{\alpha-1} e^{-\beta m} = Gamma(\alpha, \beta)$$

, where m stands for the mRNA copy number and Γ for the gamma function. Since the CD147 mRNA half-life is known (Schwanhäusser et al., 2011), we can infer the mRNA degradation rate (c_4) and calculate the gene activation rate (c_1) by multiplying it with the rate parameter (α).The mean mRNA numbers can be calculated by dividing the shape by the rate parameter of the gamma pdf.

Moment-based modeling of CD147 expression

We model CD147 transcription and translation through the reaction network

$$\begin{aligned} G_0 &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} G_1 \\ G_1 &\xrightarrow{c_3} G_1 + M \\ M &\xrightarrow{c_4} \emptyset \\ M &\xrightarrow{c_5} M + P \\ P &\xrightarrow{c_6} \emptyset, \end{aligned}$$

with G_0 and G_1 as the inactive and active gene, M as mRNA, P as protein and $c_1, \ldots, c_6$ as stochastic rate constants associated with the respective reactions. In the following we will denote by $X(t) = ([G_0], [G_1], [M], [P]) \equiv (X_1(t), \ldots, X_4(t)) \in \mathcal{X}$ the molecular state of the network at time t. Assuming that the reaction volume is well-mixed, the dynamics of such a reaction network can be described through a continuous-time Markov chain (CTMC) X with propensities $h_k(x, c_k) = c_k g_k(x)$ for $k = 1, \ldots, 6$ where g_k is a polynomial in the state x determined by the law of mass-action. With all rate constants known, the probability distribution over $X(t) = x$ satisfies a chemical master equation (CME) of the form

$$\dot{P}(x,t) = \sum_{k=1}^{6} h_k(x - \nu_k, c_k) P(x - \nu_k, t) - h_k(x, c_k) P(x, t), \tag{1}$$

with ν_k as the stoichiometric change vector associated with reaction k (i.e., the net change in x when reaction k fires). However, in many practical scenario, such equations agree only poorly with what is measured experimentally due to environmental – or *extrinsic* variability (Swain et al., 2002). Here we assume that the majority of such variability is due to differences in the ribosomal abundance from one cell to the next (Colman-Lerner et al., 2005). More technically, we assume that the propensity function h_5 depends on an environmental random variable $Z \sim p(z)$, i.e., $h_5(x, c_5, Z) = c_5 Z g_5(x)$. For convenience, we assume that $\mathbb{E}[Z] = 1$ and $\mathrm{Var}[Z] = \sigma_Z^2$. Each cell's dynamics can the be described by a conditional CME

$$\begin{aligned} \dot{P}(x_1, x_2, x_3, x_4, t \mid Z = z) = {} & c_1 P(x_1 + 1, x_2 - 1, x_3, x_4, t \mid Z = z) \\ & + c_2 P(x_1 - 1, x_2 + 1, x_3, x_4, t \mid Z = z) \\ & + c_3 P(x_1, x_2, x_3 - 1, x_4, t \mid Z = z) \\ & + c_4 (x_3 + 1) P(x_1, x_2, x_3 + 1, x_4, t \mid Z = z) \\ & + c_5 z x_3 P(x_1, x_2, x_3, x_4 - 1, t \mid Z = z) \\ & + c_6 (x_4 + 1) P(x_1, x_2, x_3, x_4 + 1, t \mid Z = z) \\ & - (c_1 + c_2 + c_3 + c_4 x_3 + c_5 z x_3 + c_6 x_4)\, P(x_1, x_2, x_3, x_4, t \mid Z = z). \end{aligned} \tag{2}$$

Assuming a negligible degree of technical noise, our single-cell measurements can be assume to be N independent samples from the stationary mRNA- and protein distribution of the an ensemble of differently parameterized CTMCs $X \mid (Z = z^m)$ for $m = 1, \ldots, N$. Therefore, in order to achieve compatibility with the data one needs to integrate (2) with respect to the extrinsic factors Z and solve for $P(x_3, x_4, t) = \int \sum_x P(x_1, x_2, x_3, x_4, t \mid Z = z) p(z) \mathrm{d}z$ for large t. However, this turns out to be infeasible for most systems of realistic size. We therefore apply a moment-based description of the model in which the dependency on extrinsic factors can be marginalized analytically (Zechner et al., 2012) and furthermore, only a few moments (e.g., mean and variance) of X need to be considered (as opposed to whole probability distributions). With $f(x, z) : \mathbb{R}^5 \mapsto \mathbb{R}$ as a polynomial in x and z the moments of the gene expression network are defined as

$$\mathbb{E}[f(x,z)] = \int \sum_{x \in \mathcal{X}} f(x,z) P(x, t \mid Z = z) p(z) \mathrm{d}z$$

and their temporal dynamics are obtained through differentiation, i.e.,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int \sum_{x \in \mathcal{X}} f(x,z) P(x, t \mid Z = z) p(z) \mathrm{d}z &= \int \sum_{x \in \mathcal{X}} f(x,z) \frac{\mathrm{d}}{\mathrm{d}t} P(x, t \mid Z = z) p(z) \mathrm{d}z \\ = \int \sum_{x \in \mathcal{X}} f(x,z) \left(\sum_{k=1}^{6} h_k(x - \nu_k, c_k) P(x - \nu_k, t \mid Z = z) - h_k(x, c_k) P(x, t \mid Z = z) \right) & p(z) \mathrm{d}z. \end{aligned} \tag{3}$$

In this work we consider moments of mRNA and protein up to order two, i.e., $\mathbb{E}[M(t)]$, $\mathbb{E}\left[M^2(t)\right]$, $\mathbb{E}[P(t)]$, $\mathbb{E}\left[P^2(t)\right]$ and $\mathbb{E}[M(t)P(t)]$. However, those moments will implicitly depend on moments and cross moments of the other species G_0, G_1 and the extrinsic variable Z. In general, this can cause the resulting dynamics to be non-closed, meaning that the k-th order moment depends on the $(k+1)$-th order moment and so forth. However, due to the linearity of the gene expression network, the moment dynamics can be shown to close at order four such that the desired moments can be computed without resorting to approximations. For the considered network we derived the exact moment equations using Matlab's symbolic toolbox and subsequently solved for their steady-state solution. The resulting expressions for the mRNA and protein moments are functions of the rate constants as well as the extrinsic noise strength σ_Z^2. Note that explicit expressions are not provided for the sake of compactness.

Inference of model parameters

The resulting moment equations can be readily used to infer the model parameters $c_1, \ldots, c_6$ and σ_Z^2. More specifically, one aims to find a set of parameters such that the analytical moments agree best with the empirical moments. The latter are computed as Monte Carlo averages using N independent data points $(M^{(i)}, P^{(i)})$, e.g., $N^{-1}\sum_{i=1}^{N} M^{(i)}P^{(i)}$ in case of the correlation between mRNA and protein. Here we employ the approach from (Zechner et al., 2012), where – based on the central limit theorem – the empirical moments are assumed to be normally distributed around the true value with some standard deviation, which we estimated using *bootstrapping*. We furthermore incorporated prior knowledge in terms of mRNA and protein half-lives that have been reported earlier (Schwannhäusser et al., 2011). In particular, we assumed Gamma-priors for both c_4 and c_6 with mean $9.18e-6s^{-1}$ and $5.75e-6s^{-1}$, respectively and a coefficient of variation of 0.447. All other priors were chosen to be practically non-informative (i.e., gamma distributions with parameters $\alpha = 0.01$ and $\beta = 0.01$). In conjunction with the a standard Metropolis-Hastings Markov Chain Monte Carlo scheme, the moment-based modeling approach allowed us to infer posterior distributions over the unknown parameters. We remark that one cannot absolutely quantify the gene switching rates c_1 and c_2 from only steady state measurements. We therefore reparameterized the model in terms of the *effective rate of transcription* (Zenklusen et al., 2008), summarizing promoter switching and transcription into a single parameters, i.e, $c_T = c_3 c_1/(c_1 + c_2)$. The whole inference procedure was repeated five times, each time returning equivalent results (up to sampling variance). The resulting parameter estimates can be found in Table S.1.

Supplemental References

Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., and Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. Nature 437, 699–706.

Delignette-Muller, M.L., and Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. e. J. Stat. Softw. 64, 1–34.

Dube, S., Qin, J., and Ramakrishnan, R. (2008). Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. PLoS ONE 3, e2876.

Kobayashi, M., Saitoh, S., Tanimura, N., Takahashi, K., Kawasaki, K., Nishijima, M., Fujimoto, Y., Fukase, K., Akashi-Takamura, S., and Miyake, K. (2006). Regulatory roles for MD-2 and TLR4 in ligand-induced receptor clustering. J. Immunol. 176, 6211–6218.

Lee, T.K., Denny, E.M., Sanghvi, J.C., Gaston, J.E., Maynard, N.D., Hughey, J.J., and Covert, M.W. (2009). A noisy paracrine signal determines the cellular NF-kappaB response to lipopolysaccharide. Sci Signal 2, ra65.

Nam, H., and Benezra, R. (2009). High levels of Id1 expression define B1 type adult neural stem cells. Cell Stem Cell 5, 515–526.

Nelson, G., Paraoan, L., Spiller, D.G., Wilde, G.J.C., Browne, M.A., Djali, P.K., Unitt, J.F., Sullivan, E., Floettmann, E., and White, M.R.H. (2002). Multi-parameter analysis of the kinetics of NF-kappaB signalling and transcription in single living cells. J. Cell Sci. 115, 1137–1148.

Wall, E.A., Zavzavadjian, J.R., Chang, M.S., Randhawa, B., Zhu, X., Hsueh, R.C., Liu, J., Driver, A., Bao, X.R., Sternweis, P.C., et al. (2009). Suppression of LPS-induced TNF-alpha production in macrophages by cAMP is mediated by PKA-AKAP95-p105. Sci. Signal. 2, ra28.