Understanding gene expression variability in its biological context using theoretical and experimental analyses of single cells

Kempe, H.

Citation for published version (APA):
Chapter 2

The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise

In collaboration with:
Anne Schwabe, Fred Cremazy
Pernette J. Verschure and Frank J. Bruggeman

Adapted from: Kempe et al., Molecular Biology of the Cell, 2015.
2.1 Abstract

Transcriptional stochasticity can be measured by counting the number of mRNA molecules per cell. Cell-to-cell variability is best captured in terms of concentrations rather than molecule counts, because reaction rates depend on concentrations. We combined single-molecule mRNA counting with single-cell volume measurements to quantify the statistics of both transcript numbers and concentrations in human cells. We compare three cell clones that differ only in the genomic integration site of an identical constitutively expressed reporter gene. The transcript numbers per cell varied proportional with cell volume in all three clones, indicating concentration homeostasis. We find that the cell-to-cell variability in the mRNA concentration is almost exclusively due to cell-to-cell variation in gene expression activity, whereas the cell-to-cell variation in mRNA number is larger, due to a significant contribution of cell-volume variability. We conclude that the precise relationship between transcript numbers and cell volume sets the biological stochasticity of living cells. This study highlights the importance of the quantitative measurement of transcript concentrations in studies of cell-to-cell variability in biology.

2.2 Introduction

Spontaneous fluctuations in the activities of molecular processes cause heterogeneity in the molecular composition of isogenic cells [23, 33, 49]. Cell-to-cell variability, often referred to as ‘noise’, has been observed in mRNA [27, 29, 50] and protein levels [23, 33, 49, 51], in the timing of molecular processes [52, 53], and in cellular growth rates [54]. The causes of molecular noise involve molecules occurring at low numbers per cell, such as transcription factors or mRNAs, that tend to show large, spontaneous deviations relative to their mean number within the cell population [13]. These deviations (fluctuations) can be caused by cell division [55], transcription bursting [56], or transient imbalances between molecular synthesis and degradation rates that occur spontaneously through thermal noise (‘intrinsic’ noise) or due to fluctuations in the number of regulators (‘extrinsic’ noise) [13, 22–24]. Extrinsic noise indicates that fluctuations can propagate through the entire molecular network of a cell [25]. As a result, the molecular composition of cells can be highly variable and cause heterogeneity in differentiation decisions [57], stress response magnitudes [58], and the survival prospects of cells after drug exposure [59]. Two highly informative reviews covering gene expression noise and its consequences are [60, 61].

Early studies on stochastic gene expression relied on fluorescent proteins to assess protein noise by either taking snapshots [23, 33] or by real-time fluorescence imaging [49, 62]. More recently, single-molecule mRNA counting has been introduced [27, 29, 63] as a method for absolute quantification of mRNA numbers. The advantage of single-transcript counting with single-molecule RNA FISH (smFISH) [29, 50, 63, 64] is that it does not require genetic engineering. Specific
DNA-probes tagged with fluorescent dyes are used to visualize individual mRNA molecules within fixed cells (Figure 2.1A).

SmFISH has great potential in cell biology to assess the role of stochasticity in cellular behaviour, such as differentiation and drug responses. Cell-to-cell variability (noise) is best captured by concentrations of molecules because reaction rates depend on concentrations. Single-molecule counting accompanied by cell-volume measurements provides a reliable way to assess biological noise of single cells. Such data has so far not been reported for transcripts.

In this study, we determined single-cell mRNA concentrations by quantifying the volume of single cells and their mRNA numbers using confocal microscopy. We studied three human cell line variants that express the same constitutively-expressed gene from a different genomic location to identify gene-location dependent effects. In order to attain robust statistics of the volume dependency of the mRNA number statistics, we studied nearly a 1000 single cells of each clone. We find that transcript number noise overestimates biological noise (in concentration units) by a factor that depends on the cell volume variability and the correlation of transcript numbers with cell volume.

### 2.3 Results

**Single-cell transcript data indicates gene-location dependent mRNA expression**

We analyzed three clones derived from the same human cell line (HEK293) (described in [65]). Each clone has a single random insertion of the same GFP reporter-gene controlled by a constitutive phosphoglycerate kinase (PGK) promoter [65]. We determined the statistics of the GFP mRNA levels in single cells with smFISH (Figure 2.1A). The probe set contained 35 probes of 17 to 18 nucleotides coupled with a fluorescent label (Section 2.6.6). Images of single cells were obtained with confocal microscopy on smFISH treated cells counterstained with DAPI. Individual cells were recorded as 52 z-stack images (300 nm/slice; Figure 2.1D and Figure 2.6). Lamina staining confirmed that the DAPI staining correctly identifies the nuclear envelope (Figure 2.5). Based on the co-localization with the DAPI signal, we assigned each mRNA molecule to be either nuclear or cytoplasmic. An overview of the transcript statistics is given in Figure 2.1B and the mRNA distributions are shown in Figure 2.1E-G.

For cell clone I, the number of mRNA molecules expressed per cell ($m$) was on average 45.7 mRNA transcripts, obtained from a dataset containing 838 cells. The coefficient of variation indicates that the standard deviation is about 45% of the mean. Approximately 72% of the cells have mRNA numbers that deviate less than one standard deviation from the mean mRNA number. The symmetry of the mRNA number distribution is indicated by the 13.8% and 13.9%
of the cells deviating more than one standard deviation of the mean on the low and high tail of the distribution respectively.

Colocalization of a mRNA spot with the DAPI signal enabled us to calculate the mRNA number

**Figure 2.1: Statistics of single cell mRNA numbers.** (A) Schematic overview of the smFISH method applied to our reporter gene mRNA. Colocalization of the mRNA molecules with the DAPI counter staining identified spots as nuclear mRNA (mN), others are cytoplasmic mRNA (mC). (B) Statistics of the mRNA molecules in the cell (m), nucleus (mN) and cytoplasm (mC) for the three different clones (color coded). Notation: μ=mean, σ=standard deviation, cv= coefficient of variation, ∆ρ= correlation between mC and mN, and *p < 0.001 (H₀ : ρ = 0). (C) For a specific cell volume (V), the mean mRNA number is calculated from the data. This conditional mean (⟨m|V⟩) shows a linear scaling with respect to volume indicating homeostasis in mRNA concentration. The gray histogram in the background shows the total number of cells per volume bin for all three clones (bin size = 100 µm³). Higher counts indicate higher reliability of the corresponding determination of (⟨m|V⟩). A least-squares linear fit is shown for all three clones. The explained fraction of the variance in ⟨m|V⟩ with this fit is 0.80, 0.77 and 0.84 for clone I, II and III respectively. Figure S9 shows the single cell relation between cell volume and mRNA number. The conditional variances of the data are given in Figure 2.14 (D) Representative confocal images of a cell, with Z1 to Z12 corresponding to subsequent optical sections (z-slices) of the cell. The mRNA molecules are shown in red and the DAPI stained nucleus is shown in blue. Additional illustrative images are given in Figure 2.6. (E-G) Scatterplots of mC and mN for the three different clones. Marginal histograms show the distribution of mC (top) and mN (right). The measured number of cells is given by n.
in the nucleus of these cells. The mean number of mRNA per nucleus ($m_N$) was 9.7 with a standard deviation of 5.4. The cytoplasmic mRNA number ($m_C$) follows directly from $m - m_N$ and is 36.0 mRNA transcripts. The number of mRNA transcripts appears to be lower in the nucleus than in the cytoplasm. Given that the nucleus and cytoplasm are proximately equal in size (see section 2.3.1), this data indicates that the lifetime of the mRNA in the cytoplasm is higher than the residence time in the nucleus. Compared to the coefficient of variation of the mRNA in the nucleus, which is 56%, the cytoplasmic mRNA numbers are less noisy with a standard deviation of about 48% of the mean. The higher noise in nuclear mRNA numbers is mostly explained by the higher intrinsic noise contribution ($1/\mu_N$) in the nucleus. The fact that $(\sigma/\mu)^2$ exceeds $1/\mu$ for nuclear mRNA indicates that part of the gene expression noise derives from extrinsic sources [13]. The correlation coefficient ($\rho$) between the nuclear and cytoplasmic mRNA numbers per cell indicates a moderate correlation ($\rho = 0.57$) between the nuclear and cytoplasmic mRNA numbers.

The same analysis for the other two clones yields very similar results. The combined data of the three clones allows for a comparison of the different genomic integration sites. The mean expression level differs between the three clones (ANOVA; $p < 0.0001$, Section 2.6.2) on average by 20% and maximally by 40%. These numbers correspond with the protein expression data of the clones [65] (Figure 2.12). The cell-to-cell variability in mRNA numbers per cell, measured as the coefficient of variation, is significantly different (ANOVA; $p < 0.03$, Section 2.6.2) between the three clones. Since the three clones differ only in their genomic location of the reporter gene, these differences demonstrate an influence of gene location on expression stochasticity [66–68]. Additionally, an increase in the mean expression level does not necessarily cause an increase in the standard deviation when comparing the different clones. This indicates that the mean expression level and the variability can vary independently, which was previously observed for protein expression data [66].

### 2.3.1 Volume statistics of single cells

In order to assess the volume dependency of the mRNA numbers per cell, we need to measure the volume of each cell. The same confocal z-stack images used for smFISH were used to determine the whole-cell, cytoplasmic and nuclear volumes of the cells by tracing the contours of these compartments (Figure 2.2A). This allows us to obtain mRNA number, volume, and concentration data for each cell.

The measured volume distributions for the whole cell (Figure 2.9) as well as the cytoplasmic and nuclear volume distributions (Figure 2.2C) indicates that there are more smaller than average than larger than average cells. Similarly shaped distributions have previously been reported for stationary growing cell populations [69]. These positively skewed distributions are likely due to the formation of two (smaller) daughter cells from each (large) mother cell. The cell volume
distributions can be well approximated by theoretical cell volume distributions derived from balanced, exponential growth of the cells (Figure 2.11) [70].

The obtained volume measurements are summarised in Figure 2.2B. The mean cell volume \( V \) for clone I is 1800 \( \mu m^3 \), with on average a larger nucleus \( (979 \mu m^3) \) than cytoplasm \( (822 \mu m^3) \). For the volume distribution of clone I, 70% of the cells have a volume deviation maximally one standard deviation of the mean. The correlation coefficients between the nuclear and the cytoplasmic volume indicate a weak but significant, positive correlation, this phenomenon may well be the result of concurrent growth during cell maturation.

Since the clones are isogenic, except for the integration site of the construct, the measured volume statistics are expected to be similar. Figure 2.2B and Figure 2.9 confirm this expectation for the three measured volume distributions.

![Figure 2.2: Statistics of single cell volumes.](image)

(A) Overview of the determination of the cell volumes. The background intensity was used to track the contour of the cell and the DAPI signal provides the nuclear contour. The three dimensional cell image was reconstructed by combining the contours of subsequent z-slices. (B) Statistics of the volumes of the cell \( (V) \), nucleus \( (V_N) \) and cytoplasm \( (V_C) \) for the three different clones (color coded). Notation: \( \mu = \text{mean} \), \( \sigma = \text{standard deviation} \), \( cv = \text{coefficient of variation} \), \( \Delta \sigma = \text{the fraction of samples between } \mu - \sigma \text{ and } \mu + \sigma \), \( \rho = \text{correlation between } V_C \text{ and } V_N \), and \( *p < 0.001 \ (H_0: \rho = 0) \). (C-E) Scatterplots of \( V_C \) and \( V_N \) for the three different clones. Marginal histograms show the distribution of \( V_C \) (top) and \( V_N \) (right). Figure 2.9 gives the distributions of \( V \). The measured number of cells is given by \( n \).

### 2.3.2 mRNA concentration statistics of single cells indicate mRNA concentration homeostasis

Next, we combined the mRNA number and volume data of each cell to determine the statistics of cellular, cytoplasmic, and nuclear mRNA concentrations (Figure 2.3A). The concentrations
is defined as the total number of molecules in a compartment divided by the volume of that compartment. Figure 2.3B shows that the mean mRNA concentration differs between the three clones, indicating the dependence of expression levels on the gene location. Scaling of the mean and standard deviation are independent. We observe higher mRNA concentrations in the cytoplasm than in the nucleus. Similar conclusions were drawn from the mRNA numbers.

The comparison of the mRNA concentration and mRNA numbers in terms of the coefficient of variation indicates that the concentration displays smaller cell-to-cell variability for all clones. This indicates that mRNA number noise overestimates the functional noise of the cells. The concentration variability is smaller due to the correlation of the cell volume with mRNA numbers per cell. This is indicated in Figure 2.1C, showing a linear dependency of the mean mRNA number (at a specific volume) with volume. The dependency of the mean mRNA numbers on volume is proportional, such that a doubling in cell volume is accompanied by a doubling in the mRNA numbers. The cellular mRNA concentration conditional on the cell volume is therefore constant, indicating homeostasis of the mRNA concentration over the cell cycle (Figure 2.3C).

2.3.3 The volume scaling of the mRNA concentration statistics explains the concentration variability

To address the origins of the differences between the mRNA concentration (c) and the mRNA number (m) noise, we apply the law of total variance (Section 2.6.3 and Figure 2.10). This relation decomposes the mRNA number and mRNA concentration noise, each in a term that captures the volume-induced noise and another that quantifies gene-expression noise (here x denotes either m or c),

\[
\frac{\text{total noise } (\sigma^2_x)}{\langle x \rangle^2} = \frac{\text{Volume-induced noise } (\sigma^2_{V \mid x})}{\langle x \rangle^2} + \frac{\text{Gene-expression noise } (\sigma^2_x)}{\langle x \rangle^2}
\]

Figure 2.4B makes the same decomposition for the experimental data. It shows that the gene-expression noise term accounts for about 70% of the mRNA number noise and for over 95% of the mRNA concentration noise. This indicates that mRNA number noise has a large contribution that derives from the scaling of the number of transcripts per cell with the cell volume.

The volume-induced noise contribution to the mRNA number noise can be estimated using the experimentally-observed homeostasis relation (Figure 2.1C): \(\langle m \mid V \rangle = \alpha V\), with \(\alpha\) as a positive constant that equals the mRNA concentration (c). As a consequence, we obtain \(\langle m \rangle = cV\), \(\text{var}(\langle m \mid V \rangle) = c^2 \text{var}(V)\) and \(\text{var}(\langle m \mid V \rangle) / \langle m \rangle^2 = \text{var}(V) / \langle V \rangle^2\). This means that the noise in cell volume \((\text{var}(V) / \langle V \rangle^2)\) equals the volume-induced noise contribution to mRNA number noise. For an idealized model of cell growth, where cells divide at fixed intervals and into exactly equal halves, the noise in cell volume can be calculated to be approximately 0.04
Due to mRNA concentration homeostasis with cell volume, the volume-dependent term of the

(Separate section of the document)

(Section 2.6.3.3), which provides a lower bound for the volume-dependent noise in the mRNA numbers. The overview of the variances of the experimental data in Figure 2.4B shows that the volume-dependent noise is indeed close to the theoretical estimate of 0.04 and is almost equal to the noise directly calculated from the experimental volume distributions ($\eta_0^2$). The volume-dependent term explains 29.5%, 28.4%, and 30% of the mRNA number noise in clone I, II, and III, respectively.

Due to mRNA concentration homeostasis with cell volume, the volume-dependent term of the
concentration noise equals zero; i.e. \( \text{var}(\langle c | V \rangle)/\langle c \rangle^2 = 0 \), which is also indicated by the experimental data (Figure 2.3C and \( \eta_\text{c}^2 \) in Figure 2.4B). Thus, the mRNA concentration noise is almost entirely determined by gene-expression noise.

The experimental data indicate that the gene-expression contributions to mRNA concentration and mRNA number noise are similar in absolute values (Figure 2.4B). Since volume-derived noise in mRNA concentration is close to zero, the mRNA concentration noise is approximately equal to the gene-expression derived noise in mRNA numbers. At mRNA concentration homeostasis, the exact relationship between the gene-expression induced mRNA number and concentration noise is given by (Section 2.6.3.5):

\[
\frac{\langle \text{var}(c | V) \rangle}{\langle c \rangle^2} = \langle V \rangle^2 \left( \frac{1}{V^2} \right) \frac{\text{var}(m | V)}{\langle m \rangle^2} + \frac{\sigma^2(1/V^2, \text{var}(m | V))}{\langle m \rangle^2 / \langle V \rangle^2}
\]

(2.2)

with \( \sigma^2(x, y) \) denoting the covariance between \( x \) and \( y \). Since we observe that \( \text{var}(m | V) \) increases with volume (Figure 2.4B and Figure 14), the covariance will be negative. Therefore, \( \langle V \rangle^2 \left( \frac{1}{V^2} \right) \) is an upper bound for the relative deviation between the conditional noise in mRNA concentration and mRNA numbers in case of concentration homeostasis. This upper bound is reached when the covariance equals zero. From the volume probability distribution, \( \langle V \rangle^2 \left( \frac{1}{V^2} \right) \) is estimated to be 1.12 (Section 2.6.3.5). The deviation between \( \frac{\langle \text{var}(c | V) \rangle}{\langle c \rangle^2} \) and \( \frac{\text{var}(m | V)}{\langle m \rangle^2} \) requires the calculation of the covariance from the volume scaling relation of the mRNA number variance conditional on volume, i.e. from \( \text{var}(m | V) = \beta V^7 \). The experimental data (Figure 2.14) indicates that this scaling is maximally quadratic with volume. The volume growth model predicts that in case of a linear dependence \( \frac{\langle \text{var}(c | V) \rangle}{\langle c \rangle^2} \) is 4% higher than \( \frac{\text{var}(m | V)}{\langle m \rangle^2} \), in case of a quadratic dependence it is 4% lower and only when \( \text{var}(m | V) = \beta V^0 \) the maximal deviation of 12% is achieved (Section 2.6.3.6). As can be seen from Figure 2.4B the relative difference between the two conditional noise terms indeed is close to the \( \pm 4\% \) region as predicted by theory in combination with the experimentally-observed volume scaling.

As a result of these relations, we can conclude that for our data the difference between mRNA concentration noise \( \frac{\text{var}(c)}{\langle c \rangle^2} \) and mRNA number noise \( \frac{\text{var}(m)}{\langle m \rangle^2} \) is dominated by the contribution of the volume dependent noise term which is zero for concentrations but equals values between 0.04 and 0.06 for mRNA numbers. Under conditions of mRNA concentration homeostasis this term is expected to be independent of the average expression level. The relative difference between mRNA number and concentration noise then depends on the magnitude of the volume independent noise contribution. For the three clones we investigated, this amounted to a 36%, 33%, and 45% difference between mRNA concentration and mRNA number noise. Thus, biologically-relevant mRNA concentration noise differs greatly from mRNA number noise, indicating the importance of the combined measurement of mRNA numbers and volumes of single cells.
2.4 Discussion

Single-molecule RNA FISH is a powerful method to assess cell-to-cell variation in gene expression. It does not require genetic engineering and it is gives the exact numbers of mRNA molecules per cell. In this study, we combined smFISH with cell-volume measurements to obtain insight into the cell-to-cell variation of the mRNA concentration per cell. We studied three clones that only differed in the location of an identical reporter construct controlled by a constitutive PGK-promoter. The differences in the mRNA statistics of these clones indicates gene-location dependency, which presumably results from the different genomic contexts at the integration site. We found that mRNA number noise overestimates functional cell-to-cell variation. Noise measured in the mRNA concentration circumvents this problem by taking into account the correlation of mRNA numbers with cell volume. One candidate source of this correlation is cell growth.

We found that the mean mRNA number conditional on volume, \( \langle m|V \rangle \), scaled linearly with volume, i.e. \( \langle m|V \rangle = cV \), which indicates a constant mRNA concentration \( c \) as function of cell volume (homeostasis). This we interpret as a constant mRNA concentration while the cell volume grows. In addition, we found that the mRNA number variance conditional on the volume, \( \text{var}(m|V) \), displayed a stronger than linear scaling with volume, i.e. \( \text{var}(m|V) = \beta V^\gamma \) with \( 1 \leq \gamma \leq 2 \). The latter scaling explained the difference between the volume independent mRNA concentration and mRNA number noise, i.e. between \( \langle \text{var}(c|V) \rangle / \langle c \rangle^2 \) and \( \langle \text{var}(m|V) \rangle / \langle m \rangle^2 \), which maximally amounts to a relative deviation of \( \pm 4\% \) according to theory and in agreement with the experimental data (Figure 2.4). Taken together, these findings allow for a simple estimation of mRNA concentration noise based on mRNA number noise under conditions of mRNA concentration homeostasis: \( \langle \delta^2 c \rangle / \langle c \rangle^2 \approx \langle \delta^2 m \rangle / \langle m \rangle^2 - 0.04 \). We emphasise that the observed homeostasis of the mRNA concentration as function of the cellular volume may not apply to all genes.

Our results indicate that constitutive gene expression is not completely understood at the level of a single cell. For homeostasis to occur during volume growth of the cell requires that \( \langle m|V \rangle = cV = k_s V; \) with \( k_s \) and \( k_d \) as a zero-order and first-order rate constant for mRNA synthesis and degradation, respectively. (The half life of the mRNA is about 8 hours [71], i.e. much shorter than the generation time of about 24 hours.) In other words, either the transcription rate or the degradation rate of mRNA are volume dependent or both such that the net effect leads to the proportionality of \( \langle m|V \rangle \) with volume. This suggests a coupling between the net rate of increase in the transcript numbers and the cell volume. It is not clear how this results from the combined influences of mRNA decay, replication, and cell volume dynamics. Coupling between the rate of transcription and cell growth was previously shown by Zhurinsky and colleagues [72]. Their data indicates a global mechanism that determines the rate of transcription of most genes and the cellular growth rate.
Figure 2.4: The theoretical and experimental relation between mRNA concentration and mRNA number noise and their dependency on volume. a) Overview of theoretical relations used to analyse in the data. The mRNA number noise ($\eta_m^2$) and mRNA concentration noise ($\eta_c^2$) can be decomposed into two terms; the volume-dependent noise ($\eta_v^2$) and gene-expression noise ($\eta_g^2$) (see equation 2.1). When there is homeostasis of mRNA concentration, the relation between $\eta_g^2_m$ and $\eta_g^2_c$ depends on the scaling of the variance in the conditional mRNA numbers. Under these conditions the volume dependent noise in mRNA numbers ($\eta_v^2_m$) equals the noise in the volume distribution ($\eta_v^2$). b) Experimental data of the relations shown in the theory section above. The different colours give the corresponding measures for the three different clones. The pie charts show how the total mRNA number and mRNA concentration noise is decomposed. mRNA number noise is higher than concentration noise, mainly due to its volume contribution. The scaling of $\text{var}(m|V)$ with volume for Clone III is shown. The observed scaling is linear or quadratic, resulting in a deviation of $\pm4\%$ between $\eta_m^2$ and $\eta_c^2$. The scaling of $\text{var}(m|V)$ with volume for Clone I and II, as well as the scaling of $\text{var}(c|V)$ can be found in Figure 2.14.

We observed the volume scalings of $\langle m|V \rangle$ and $\text{var}(m|V)$ with the same construct expressed from different genomic locations and therefore it is likely not a genome location dependent phenomenon. This is likely also not a property of our reporter construct, as several other
Chapter 2. Single-cell mRNA concentration noise

studies reported similar effects [49, 62, 73, 74]). Cookson et al. [73] found in yeast a peaked dependency of a green fluorescent protein expressed from a constitutive promoter as function of cell volume, which is indicative of constant synthesis and an accelerating growth of cell volume as function of the cell cycle. Similar data was reported for several human proteins [49].

The three clones investigated show a 36%, 33%, and 45% difference between mRNA concentration and mRNA number noise. We conclude that functional mRNA noise differs greatly from mRNA number noise, indicating the importance of measuring mRNA noise in terms of concentrations. For inducible or cell-cycle dependent promoters the assessment of mRNA concentration noise is even more relevant. For such systems, a nonlinear relation between volume growth and mRNA synthesis is expected. This would introduce less predictable relations between the mRNA number and mRNA concentration noise than reported in this study, which is limited to a constitutive promoter.

Real-time monitoring of mRNA numbers (for instance by using MS2-labeling [75]) and volume growth of single cells for a set of (classical) constitutive or inducible promoters could provide more information about the origins and effects of the volume scaling relations of the mean and variance of the transcript numbers. A downside is that such studies would require the tracking of several hundreds of cell divisions to obtain robust statistics on the volume dependencies of mRNA-number statistics. To attain robust statistics in our experiments, we used a confocal microscopy setup and had to study roughly a 1000 cells per clone.

Single-molecule methods are a great addition to biology. They are quantitative and exact; as they give insight into the actual molecular composition of single cells. In addition, they directly relate to stochastic theory and model predictions giving deep insights into how cells exploit the inherent stochasticity of molecular processes to diversify isogenic populations. In this study, we have shown that in order to profit fully from the exactness of single-molecule methods in cell biology, these methods should ideally be combined with single-cell volume measurements. This finding is highly relevant, since concentration noise captures biological noise.

2.5 Materials and Methods

Cell clones and cell culture

Experiments were performed on human embryonic kidney cells (HEK293) with a single integration of a phosphoglycerate kinase (PGK) driven GFP gene construct, obtained from Gierman et al. [65]. We analyzed three different clones with the integration at different genomic locations: Clone I (HG19:chr1:225684028, within the ENAH gene), Clone II (HG19:chr1:150379508, within the RPRD2 gene) and Clone III (HG19:chr1:150664232, within the GOLPH3L gene). The cells were cultured in DMEM (Gibco®, 31965023) supplemented with 10% (v/v) fetal
calf serum (Gibco®, 16140) and 100 units/ml Penicillin-Streptomycin (Gibco®, 15140). Incubation was at 37°C in a humidified 5% CO₂ atmosphere. Before any experiments the cells were grown for at least 2 weeks after thawing to achieve steady state cell growth and steady state expression-statistics of the integrated construct.

2.5.1 Single Molecule RNA FISH

Samples were treated according to the Protocol for Adherent Mammalian Cell Lines for the Custom Stellaris FISH probes. Cells were cultured for 3 hours in Lab-Tek™ Chambered Coverglasses (Lab-Tek 155380) before fixation. EtOH permeabilization was done overnight at 4°C. For hybridization we used 125 nM probe in the hybridization buffer and incubated overnight at 37°C. Imaging was done without using anti-fade. The cells where counterstained with 5 ng/mL DAPI. The sequence of the probe targeting the eGFP insertion can be found in Section 2.6.6. The DNA probes were coupled to CAL Fluor® Red 590 fluorophores by the manufacturer (Biosearch Technologies, Inc.).

2.5.2 Image acquisition

Samples were imaged using a Nikon Ti-E scanning laser confocal inverted microscope (A1) with 60x oil objective in tandem with Nikon NIS-Elements imaging software. Excitation was by 561.5 nm diode-pumped solid state and 402.1 nm diode lasers. Detection was via 595-50 nm and 450-50 nm filters, respectively. Optical sections were captured at 0.300 µm intervals and a resolution of 256 by 256 pixels and zoom factor of 6.8, resulting in a voxel size of 0.0047 µm³ (0.1243 µm by 0.1243 µm by 0.3 µm). Four times averaging was used to reduce photon and camera noise.

2.5.3 Image analysis

Image analysis software was adapted from Raj et. al. [63]. Images are filtered with a semi three-dimensional Laplacian of Gaussian (LoG) filter which removes noise and enhances the signal to noise ratio (filter width=1.5). The number of mRNA spots was found by applying a threshold for which the number of mRNA was least sensitive to changes in this threshold. The threshold was determined by using a window function calculating the average spot numbers over 7 constitutive thresholds divided by the sum of the standard deviation of these spot counts and a constant α (=10) [76]. To measure the cell volume, the following operations were performed on all individual z-slices: 1) Median filtering (20 by 20 pixels), 2) Image thresholding (graythresh, Otsu’s method [77]), 3) Fill image regions and holes and 4) Morphological closing with a disk (radius = 4 pixels). From these processed images of the DAPI channel the nuclear volume can be obtained (Section 2.6.1). The cell volume is defined by the presence of signal from either
the nucleus (DAPI), the cytoplasm (red-channel) or both. The cytoplasmic size is given by the difference between nuclear and cell volume. The obtained pixel size was multiplied by the voxel size of $0.0047\mu m^3$ to provide the cell size in $\mu m^3$.

## 2.6 Supplemental Information

### 2.6.1 Identifying the nucleus with DAPI

The nuclear counterstain 4’-6-Diamidino-2-phenylindole (DAPI) is used to identify the localization of the mRNA’s as either cytoplasmic or nuclear and to measure the volume of the nucleus. To check how well the DAPI staining confines the nucleus, we compare DAPI staining with a mCherry tagged nuclear lamina protein (LMNB). Cells were transiently transfected with Lipofectamin 2000, the ratio of lipofectamin to DNA was 2.5, according to the manufacturer instructions. Images were acquired using 402.1 nm and 591.5 nm lasers for excitation and detection was via a 450-50 nm band pass and 605 nm long pass filters, respectively (Figure 2.5).

![Figure 2.5: Lamina staining confirms that DAPI reliably distinguishes the nuclear boundary. Overall the DAPI staining (Blue signal) coincides with the lamin staining (Red signal), as can be seen by the profile above the microscopy images. Images are one single z-slice. The profiles show that at the lamin (red signal) the dapi staining (blue line) increases or decreases rapidly.](image)

### 2.6.2 Experimental variation and statistics

Each single imaging session spanned several hours. Figure 2.7 shows that there is no systematic increase or decrease in spot count during the course of imaging. The two smallest samples in cell clone I (Figure 2.7; red lines) show the strongest decrease over time. For each cell clone multiple experiments were performed to obtain reliable amounts of mRNA numbers and cell volume measurements. The data collected on different days are shown in Figure 2.8. There is no indication that the individual samples of Cell clone II ($p = 0.52$) and Cell clone III ($p = 0.16$) are
Figure 2.6: Imaging examples of the single cell analyses (A) Identified mRNA spots (yellow circles) projected on a LoG-filtered image merged over the z-axis. The probe targeting the mRNA molecules is shown in red. (B) Raw image from within the cell. In addition to the mRNA probe, the computationally identified cell border is shown in white. Supplemental Movie 1 shows the whole cell raw data. (C) 3D reconstruction of the whole cell. The nuclear boundary is shown in blue, the cell boundary in green and the mRNA molecules in red. See also Supplemental Movie 2 and 3. (D) Raw images merged over the z-axis with the identified spots annotated by green circles.

not drawn from the same distribution. There is higher variation present between the individual samples of Cell clone I \((p=3 \times 10^{-7})\), p-values are obtained by Kruskal-Wallis tests. This larger variation in Cell clone I experiments might be due to the longer cell culturing period during sample acquisition and is negatively influenced by the two smallest samples, of which the mRNA count decreased over time. To test whether the different clones have different mRNA number characteristics we performed an ANOVA and compared the statistics of the three clones based on the data of the individual experiments, the mean \((p < 0.0001)\), the coefficient of variation \((p < 0.031)\), the noise \((p < 0.05)\), and the Fano factor \((p < 0.04)\). For the mRNA concentration data similar results are obtained; the mean \((p < 0.0001)\), the coefficient of variation \((p < 0.013)\), the noise \((p < 0.016)\). The volume distributions (Figure 2.9) of the different clones are expected to be the same since the cells differ only in the integration site of the construct. We tested wether the sample means of the different clones originate from the same distribution with
ANOVA. \((V; p = 0.65, V_N; p = 0.12, V_C; p = 0.24)\). For all additional analysis we pooled the individual experiments, as already shown in Figure 2.9.

**Figure 2.7:** Confirmation of the absence of scan time bias in the mRNA count. Overall we observe no influence of the acquisition time on the identified number of mRNA spots in the images. Although the two smallest samples (marked in red) show a decrease in the number of identified mRNA molecules over time, all samples were considered when analysing the dataset.

### 2.6.3 The law of total variance for mRNA numbers and concentrations of mRNA

#### 2.6.3.1 Law of total variance explained

The law of total variance can be derived from the definition of the variance, \(\text{var}(m) = \langle m^2 \rangle - \langle m \rangle^2\), and the law of total expectation, \(\langle m \rangle = \langle \langle m | V \rangle \rangle\). In which \(\langle x \rangle\) indicates the average (mean value) of \(x\) and \(x|y\) represents the conditional value of \(x\) given \(y\).

\[
\text{var}(m) = \langle m^2 \rangle - \langle m \rangle^2 \\
= \langle \langle m^2 | V \rangle \rangle - \langle \langle m | V \rangle \rangle^2 \\
= \langle \text{var}(m|V) + \langle m | V \rangle^2 \rangle - \langle \langle m | V \rangle \rangle^2 \\
= \langle \text{var}(m|V) + \langle m | V \rangle^2 \rangle - \langle \langle m | V \rangle \rangle^2 \\
= \langle \text{var}(m|V) \rangle + \text{var}(\langle m | V \rangle) \\
\quad (2.3)
\]

We can do the same for the concentration of mRNA \((c)\) and arrive at:

\[
\text{var}(c) = \langle \text{var}(c|V) \rangle + \text{var}(\langle c | V \rangle) \\
\quad (2.4)
\]

In this equation, \(\langle \text{var}(c|V) \rangle\) is the mean of the variance at fixed \(V\), corresponding to for instance transcription noise. The second term, \(\text{var}(\langle c | V \rangle)\), is the variance of the conditional means and represents the variance caused by the change in \(V\) (Figure 2.10).
Figure 2.8: Box-whisker plots of the individual experiments indicate almost no
day to day variation between single experiments. To accurately measure the mRNA
probability distribution of a single cell clone we need multiple experiments. The three box-
whisker plots show the mRNA counts of the single experiments. There is no indication that
the individual samples of Cell clone II \((p = 0.52)\) and Cell clone III \((p = 0.16)\) are not drawn from
the same distribution. Higher variation is present between the individual samples of Cell clone
I \((p=3 \times 10^{-7})\), p-values are obtained by Kruskal-Wallis tests, which assumes that the data has
a symmetric distribution. The last box shows all single experiments of the specific cell clone
combined.

Figure 2.9: Volume pooled measurements provide similar volume distributions for
the three clones. Histograms showing the volume distribution of the cells, of the cytoplasm
and of the nucleus. The three different clones are indicated by the different colours (red=cell
clone I, blue=cell clone II, and green=cell clone III). The bin-size is 100 \(\mu m^3\).

2.6.3.2 Deriving \(\frac{\text{var}(\langle m | V \rangle)}{(m)^2} = \frac{\text{var}(V)}{(V)^2}\) in case of mRNA concentration homeostasis

Homeostasis requires that the average mRNA number per cell at a given volume, i.e. \(\langle m | V \rangle\)
scales linearly with the volume, i.e. \(\langle m | V \rangle = \alpha V\). As a consequence, we obtain that \(\langle m \rangle = \langle \langle m | V \rangle \rangle_V = \alpha \langle V \rangle\) and \(\text{var}(\langle m | V \rangle) = \text{var}(\alpha V) = \alpha^2 \text{var}(V)\). This means that in case of perfect
homeostasis: \(\frac{\text{var}(\langle m | V \rangle)}{(m)^2} = \frac{\text{var}(V)}{(V)^2}\). In the next section we will show that \(\frac{\text{var}(V)}{(V)^2} \approx 0.04\).

2.6.3.3 Estimation of a lower bound for \(\frac{\text{var}(\langle m | V \rangle)}{(m)^2}\) assuming steady-state exponential
growth of the cells

Here we estimate the noise in the volume distribution for an idealized model of cell division. We
assume that cells divide at fixed intervals \(T\) and divide into exactly equal halves. Furthermore we
assume that the volume is a deterministic function of cell age (denoted by \(a\)): \(V = V(a) = V_0 e^{\mu a}\)
with \(\mu\) as the specific (exponential) growth rate. The daughter cell volume equals \(V(0) = V_0\)
and the mother volume at division equals \(V(T) = 2V_0\). Hence, \(0 \leq a \leq T\) with \(T = \ln 2/\mu\). At
balanced growth, the distribution of cell ages for this model is described by a so-called ideal age
Figure 2.10: Visualization of the law of total variance. For all samples there are two measurements, in this case called volume (\(V\)) (x-axis) and the number of mRNA molecules (\(m\)) (y-axis). The marginal distributions show the probability distributions of both variables, \(V\) (top) and \(m\) (right). The samples within each group \(V_i\) have a distribution of \(m\) of their own \((m|V_i)\), shown in orange, with their means indicated in green. The variance in \(m\) is the sum of the variance over the conditional means, \(\text{var}(m|V)\), and the mean over the conditional variances, \(\langle\text{var}(m|V)\rangle\).

The probability distribution [78] equal to

\[
u(a) = \mu 2^{1-a\mu/\ln(2)} \quad a \in [0, \ln(2)/\mu]
\]

The probability distribution of volumes, \(g(V)\), can be derived from the age distribution. Making age \((a)\) a function of volume \((V)\); \(a = \ln(V/V_0)/\mu\), and using the change of variable technique we get:

\[
g(V) = 2V_0/V^2 \quad V \in [V_0, 2V_0]
\]

The noise in this distribution is given by:

\[
\frac{\text{var}(V)}{\langle V \rangle^2} = \frac{2}{\ln(4)^2} - 1 \approx 0.04
\]

This ideal age distribution is shown in Figure 2.11. The assumption of deterministic interdivision...
times leads to the unrealistic discontinuous definition of the distribution function. A better fit to the experimentally observed distributions is obtained when the volumes at which cells divide (and those of newborn cells) are allowed to have some variation around the mean value. With scaled, symmetric beta distributions for the cell volume at cell birth and division, each with $CV = 0.1$ a much better fit to the experimental data is obtained (see Figure 2.11). The distribution of volumes is obtained from the equations deduced by Collins and Richmond [70].

![Figure 2.11: Correspondence of the theoretical and experimental volume distribution](image)

As explained in the main text we found from the experimental data that mRNA concentration and mRNA number noise differ. In this section we will derive a relation between $\frac{\langle \text{var}(c|V) \rangle}{\langle c \rangle^2}$ and $\frac{\langle \text{var}(m|V) \rangle}{\langle m \rangle^2}$.

The concentration is defined as $c = \frac{m}{V}$, which leads to the following additional relations,

$$ var(c|V) = var\left(\frac{m}{V}\right) = \frac{1}{V^2} var(m|V) $$ (2.8)

$$ \langle \text{var}(c|V) \rangle = \left\langle \frac{1}{V^2} var(m|V) \right\rangle $$

$$ = \left\langle \frac{1}{V^2} \right\rangle \langle var(m|V) \rangle $$

$$ + \sigma^2(1/V^2, var(m|V)) $$

$$ \text{covariance between the squared inverse volume and the conditional mRNA number variance} $$ (2.10)

$$ \frac{\langle \text{var}(c|V) \rangle}{\langle c \rangle^2} = \frac{\left\langle \frac{1}{V^2} \right\rangle \langle var(m|V) \rangle + \sigma^2(1/V^2, var(m|V))}{\langle m/V \rangle^2} $$ (2.11)

The relation used to go from the second to the third equation is $\langle \alpha \beta \rangle = \langle \alpha \rangle \langle \beta \rangle + \sigma^2(\alpha, \beta)$ with $\sigma^2(\alpha, \beta)$ as the covariance between the random variables $\alpha$ and $\beta$. 

2.6.3.4 General relation between $\langle \text{var}(c|V) \rangle / \langle c \rangle^2$ and $\langle \text{var}(m|V) \rangle / \langle m \rangle^2$
2.6.3.5 Simplification of the relation between $\frac{\langle \text{var}(c|V) \rangle}{\langle c \rangle^2}$ and $\frac{\langle \text{var}(m|V) \rangle}{\langle m \rangle^2}$ in case of homeostasis

In case of homeostasis we can simplify equation 2.11. Homeostasis means that $\langle m|V \rangle$ scales proportional with $V$ and, as a consequence, that the concentration is independent of volume and fixed: $\langle c|V \rangle = \alpha$. Thus, $\langle m|V \rangle = \alpha V$ and $\langle c|V \rangle = \langle m|V \rangle/V$.

First we will show that under these homeostasis conditions $\langle m/V \rangle^2 = \langle m \rangle^2/\langle V \rangle^2$. The definition of homeostasis implies,

$$\langle c|V \rangle = \frac{\langle m|V \rangle}{V} = \alpha$$  \hspace{1cm} (2.12)

Averaging this equation over volume leads to the relation $\langle m \rangle = \alpha \langle V \rangle$. Averaging $\langle c|V \rangle$ over the whole volume distribution gives:

$$\langle \langle c|V \rangle \rangle_V = \langle c \rangle = \frac{\langle m \rangle}{\langle V \rangle} = \langle \alpha \rangle = \alpha$$  \hspace{1cm} (2.13)

Hence, in case of homeostasis $\langle \frac{m}{V} \rangle = \frac{\langle m \rangle}{\langle V \rangle}$. Substituting the relation $\langle m/V \rangle^2 = \langle m \rangle^2/\langle V \rangle^2$ into equation 2.11 gives:

$$\frac{\langle \text{var}(c|V) \rangle}{\langle c \rangle^2} = \langle V \rangle^2 \frac{\langle \text{var}(m|V) \rangle}{\langle m \rangle^2} + \frac{\sigma^2(1/V^2, \text{var}(m|V))}{\langle m \rangle^2/\langle V \rangle^2}$$  \hspace{1cm} (2.14)

Equation 2.14 was used in the main text.

2.6.3.6 Simplifying the relation between $\frac{\langle \text{var}(c|V) \rangle}{\langle c \rangle^2}$ and $\frac{\langle \text{var}(m|V) \rangle}{\langle m \rangle^2}$ from volume scaling relations

To estimate the magnitude of the covariance term in equation 2.14 we approximate $\text{var}(m|V)$ as a polynomial in $V$ as this scaling is also observed in our data (Figure 2.14).

$$\text{var}(m|V) = a_0 + a_1 V + a_2 V^2$$  \hspace{1cm} (2.15)

Averaging this equation over the volume gives,

$$\langle \text{var}(m|V) \rangle = \langle a_0 + a_1 V + a_2 V^2 \rangle = a_0 + a_1 \langle V \rangle + a_2 \langle V^2 \rangle$$  \hspace{1cm} (2.16)
Using the volume distribution, $g(V)$ we can calculate the covariance between $\text{var}(m|V)$ and $1/V^2$:

$$
\sigma^2\left(\frac{1}{V^2}, \text{var}(m|V)\right) = \int_{V_0}^{2V_0} g(V) \left( \frac{1}{V^2} - \left\langle \frac{1}{V^2} \right\rangle \right) \left( a_0 + a_1 V + a_2 V^2 - (a_0 + a_1 V + a_2 V^2) \right) dV \\
= a_1 \left( \left\langle \frac{1}{V} \right\rangle - \left\langle V \right\rangle \left\langle \frac{1}{V^2} \right\rangle \right) + a_2 \left( 1 - \left\langle V^2 \right\rangle \left\langle \frac{1}{V^2} \right\rangle \right). \tag{2.17}
$$

Combining equations 2.14, 2.16 and 2.17 yields:

$$
\frac{\left\langle \text{var}(c|V) \right\rangle}{\left\langle c \right\rangle^2} = \frac{\left\langle V \right\rangle^2 \left\langle \frac{1}{V^2} \right\rangle}{\left\langle c \right\rangle^2} \cdot \frac{a_0}{\left\langle m \right\rangle^2} + \frac{\left\langle V \right\rangle \left\langle \frac{1}{V} \right\rangle}{\left\langle m \right\rangle^2} \cdot \frac{a_1}{\left\langle m \right\rangle^2} + \frac{\left\langle V \right\rangle^2}{\left\langle V^2 \right\rangle} \cdot \frac{a_2}{\left\langle m \right\rangle^2} \tag{2.18}
$$

where the approximated values are calculated using the volume distribution (eq. 2.6):

$$
\left\langle V \right\rangle^2 \left\langle \frac{1}{V^2} \right\rangle = \left( \int_{V_0}^{2V_0} V g(V) dV \right)^2 \int_{V_0}^{2V_0} \frac{1}{V^2} g(V) dV = \frac{7 \ln(4)^2}{12} \approx 1.12 \tag{2.19}
$$

$$
\left\langle V \right\rangle \left\langle \frac{1}{V} \right\rangle = \int_{V_0}^{2V_0} V g(V) dV \int_{V_0}^{2V_0} \frac{1}{V} g(V) dV = \frac{3 \ln(4)}{4} \approx 1.04 \tag{2.20}
$$

$$
\frac{\left\langle V \right\rangle^2}{\left\langle V^2 \right\rangle} = \frac{\left( \int_{V_0}^{2V_0} V g(V) dV \right)^2}{\int_{V_0}^{2V_0} V^2 g(V) dV} = \frac{\left( \ln(4) \right)^2}{2} \approx 0.96 \tag{2.21}
$$

With the experimental volume distribution (calculated based on the pooled volume data from all three clones) these values become:

$$
\left\langle V \right\rangle^2 \left\langle \frac{1}{V^2} \right\rangle = 1.18 \tag{2.22}
$$

$$
\left\langle V \right\rangle \left\langle \frac{1}{V} \right\rangle = 1.06 \tag{2.23}
$$

$$
\frac{\left\langle V \right\rangle^2}{\left\langle V^2 \right\rangle} = 0.95 \tag{2.24}
$$

On basis of the theory we can distinguish three regimes:

1. No scaling of $\text{var}(m|V)$ with volume: $\text{var}(m|V) = a_0$ then $\frac{\left\langle \text{var}(c|V) \right\rangle}{\left\langle c \right\rangle^2} = 1.12 \frac{\text{var}(m|V)}{\left\langle m \right\rangle^2}$; a 12% larger mRNA concentration noise than mRNA number noise.

2. Linear scaling of $\text{var}(m|V)$ with volume: $\text{var}(m|V) = a_1 V$ then $\frac{\left\langle \text{var}(c|V) \right\rangle}{\left\langle c \right\rangle^2} = 1.04 \frac{\text{var}(m|V)}{\left\langle m \right\rangle^2}$; a 4% larger mRNA concentration noise than mRNA number noise.

3. Quadratic scaling of $\text{var}(m|V)$ with volume: $\text{var}(m|V) = a_2 V^2$ then $\frac{\left\langle \text{var}(c|V) \right\rangle}{\left\langle c \right\rangle^2} = 0.96 \frac{\text{var}(m|V)}{\left\langle m \right\rangle^2}$; a 4% smaller mRNA concentration noise than mRNA number noise.
Our data do not distinguish between a linear and a quadratic dependence of the conditional mRNA number variance on volume, both regimes give decent fits (Figure 2.14). Hence, a discrepancy between $\langle \text{var}(c|V) \rangle / \langle c \rangle^2$ and $\langle \text{var}(m|V) \rangle / \langle m \rangle^2$ of ±4% is expected on the basis of the theoretical analysis. This agrees quite well with the difference in the experimental values of $\langle \text{var}(c|V) \rangle / \langle c \rangle^2$ and $\langle \text{var}(m|V) \rangle / \langle m \rangle^2$ (Table 1, main text); i.e. for cell clone I, II and III we find respectively 0%, 2% and 4.8% difference, which agrees very well with the theoretical estimates for this discrepancy to lie between ±4%. However, the noise in mRNA number and concentration show a larger discrepancy, i.e. of 36% ($(0.204 - 0.149)/0.149 \times 100\%$), 33% ($(0.208 - 0.156)/0.156 \times 100\%$), and 45% ($(0.148-0.102)/0.102 \times 100\%$) for cell clone I, II and III, respectively. These differences are due to the scaling of the mRNA number with cell volume due to the steady-state growth of the cells.

2.6.4 Average mRNA numbers correlate with protein expression

The GFP-protein expression levels of the integrated construct are compared to the GFP-mRNA number expression. The protein expression levels of the clones originate from Gierman et. al. [65]. As expected higher mean mRNA numbers correspond to higher protein levels (Figure 2.12).

![Figure 2.12: Apparent correlation between the mRNA and protein expression levels. Cell clone I is shown in red, Cell clone II in blue and Cell clone III in green.](image)

2.6.5 Volume dependencies within the data

The single cell data of the cell volume and the mRNA numbers is shown in Figure 2.13. This shows a positive correlation of the mRNA numbers with cell volume as shown by the proportionality of the cell volume with the conditional mean of the mRNA number ($\langle |m|V \rangle \rangle$. The conditional variances of $m$ and $c$ as a function of volume are given in Figure 2.14. The decomposition of averages and variances as function of volume for the nucleus and cytoplasm of the mRNA numbers (Figure 2.15) show similar proportionality as observed for the whole cell.
data. Homeostasis of mRNA concentrations is also observed at the nuclear (Figure 2.16A) and cytoplasmic level (Figure 2.16C).

Figure 2.13: Correlation of the mRNA numbers with cell volume for single cells. In addition to the binned data of the conditional mean \( \langle m|V \rangle \) shown in the main article this figure shows the degree of correlation between cell volume and mRNA number based on the single cell data. The three different clones are annotated by the different colors: Cell clone I is shown in red, Cell clone II in blue and Cell clone III in green.
Figure 2.14: Conditional variances of mRNA numbers and concentrations. A, C and E give the variance in mRNA number conditioned on the cell volume ($\text{var}(m|V)$). B, D and F give the variance in the conditioned concentration ($\text{var}(c|V)$). The three different clones are annotated by the different colors: Cell clone I is shown in red, Cell clone II in blue and Cell clone III in green. The gray plots in the background shows the number of data points per volume bin. The lower the count in one bin, the lower the reliability of the measurement of the data within the bin. The mRNA number data is fitted with a linear (solid, $a + bV$) and polynomial (dashed, $a + bV + cV^2$ with $c > 0$) fit averaging for the number of data points in each volume bin. For Cell clone II (C) the linear and polynomial fit are overlapping.
Figure 2.15: Conditional averages and variances in the mRNA number in the nucleus and cytoplasm. The three different clones are annotated by the different colors: Cell clone I is shown in red, Cell clone II in blue and Cell clone III in green. The gray plots in the background shows the number of data points per volume bin for either the nucleus (A and B) or the cytoplasm (C and D).

Figure 2.16: Conditional averages and variances in the mRNA concentration in the nucleus and cytoplasm. The three different clones are annotated by the different colors: Cell clone I is shown in red, Cell clone II in blue and Cell clone III in green. The gray plots in the background shows the number of data points per volume bin for either the nucleus (A and B) or the cytoplasm (C and D).
Figure 2.17: The whole-cell mRNA expression distributions in terms of mRNA number (upper row) and mRNA concentration (second row) of the three different clones (the different columns). The measured expression distribution is shown in gray. Fits of the data with a normal distribution (blue lines), lognormal distribution (red lines) and gamma distribution (black lines) are shown.
### 2.6.6 Probe sequence

The sequence of the mRNA FISH probe to detect the GFP reporter mRNA’s.

<table>
<thead>
<tr>
<th>sequence</th>
<th>position</th>
<th>%GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcgcccttgctcaccat</td>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>accaccccggtgaacag</td>
<td>22</td>
<td>65</td>
</tr>
<tr>
<td>ccagctgacccaggatg</td>
<td>42</td>
<td>65</td>
</tr>
<tr>
<td>gtggccttttaacgctg</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>tccgccgacagcgtgaa</td>
<td>82</td>
<td>65</td>
</tr>
<tr>
<td>tagggtggaatgccttc</td>
<td>103</td>
<td>65</td>
</tr>
<tr>
<td>acctcaggtgcagcttg</td>
<td>123</td>
<td>53</td>
</tr>
<tr>
<td>cttgccggtgtgcaga</td>
<td>143</td>
<td>65</td>
</tr>
<tr>
<td>agggattggcccagcaca</td>
<td>166</td>
<td>76</td>
</tr>
<tr>
<td>cgtaggtcaggggtgct</td>
<td>186</td>
<td>65</td>
</tr>
<tr>
<td>gggccgtgagcactgca</td>
<td>206</td>
<td>65</td>
</tr>
<tr>
<td>tgttcatgtggtcgcc</td>
<td>226</td>
<td>59</td>
</tr>
<tr>
<td>eggacttgagaagtcg</td>
<td>246</td>
<td>53</td>
</tr>
<tr>
<td>gaagtaccttcgctggca</td>
<td>266</td>
<td>65</td>
</tr>
<tr>
<td>agaagatgggcggctc</td>
<td>286</td>
<td>53</td>
</tr>
<tr>
<td>tgtggtgctcgctcc</td>
<td>306</td>
<td>59</td>
</tr>
<tr>
<td>aacttcactggtggccg</td>
<td>328</td>
<td>65</td>
</tr>
<tr>
<td>ttcacagggttcgtgcc</td>
<td>349</td>
<td>65</td>
</tr>
<tr>
<td>tgcctctgctgctgatg</td>
<td>369</td>
<td>59</td>
</tr>
<tr>
<td>ggcgtctctcttgagtt</td>
<td>389</td>
<td>59</td>
</tr>
<tr>
<td>agcttggtgcctccgat</td>
<td>409</td>
<td>59</td>
</tr>
<tr>
<td>ggttgtgtatggtgac</td>
<td>429</td>
<td>47</td>
</tr>
<tr>
<td>eggccatgataagcag</td>
<td>450</td>
<td>53</td>
</tr>
<tr>
<td>gatgcgctttctctgct</td>
<td>470</td>
<td>53</td>
</tr>
<tr>
<td>gggcatctctcaagttc</td>
<td>492</td>
<td>53</td>
</tr>
<tr>
<td>ggtgcgtctgctgcatg</td>
<td>512</td>
<td>65</td>
</tr>
<tr>
<td>tagtggtgcggcagctg</td>
<td>532</td>
<td>65</td>
</tr>
<tr>
<td>egatggggcgtctcgctg</td>
<td>552</td>
<td>65</td>
</tr>
<tr>
<td>tgtcggccagcagcagc</td>
<td>579</td>
<td>71</td>
</tr>
<tr>
<td>tgtgcaggctctcgagtt</td>
<td>599</td>
<td>59</td>
</tr>
<tr>
<td>ggtcttttgtgcagggc</td>
<td>619</td>
<td>65</td>
</tr>
<tr>
<td>tgtgtcctcgcttgctg</td>
<td>639</td>
<td>59</td>
</tr>
<tr>
<td>cagaaactccagacgga</td>
<td>659</td>
<td>59</td>
</tr>
<tr>
<td>agagtgatccggcgggc</td>
<td>679</td>
<td>71</td>
</tr>
<tr>
<td>tgtacagctgctccatg</td>
<td>699</td>
<td>53</td>
</tr>
</tbody>
</table>