

ARTICLE

Received 6 Feb 2014 | Accepted 21 Jul 2014 | Published 1 Sep 2014

DOI: 10.1038/ncomms5761

# Contribution of RNA polymerase concentration variation to protein expression noise

Sora Yang<sup>1</sup>, Seunghyeon Kim<sup>1</sup>, Yu Rim Lim<sup>2</sup>, Cheolhee Kim<sup>1</sup>, Hyeong Jeon An<sup>1</sup>,  
Ji-Hyun Kim<sup>2</sup>, Jaeyoung Sung<sup>2</sup> & Nam Ki Lee<sup>1,3</sup>

Cell-to-cell variation in gene expression, or noise, is a general phenomenon observed within cell populations. Transcription is known to be the key stage of gene expression where noise is generated, however, how variation in RNA polymerase (RNAP) concentration contributes to gene expression noise is unclear. Here, we quantitatively investigate how variations in absolute amounts of RNAP molecules affect noise in the expression of two fluorescent protein reporters driven by identical promoters. We find that intrinsic noise is independent of variation in RNAP concentrations, whereas extrinsic noise, which is variation in gene expression due to varying cellular environments, scales linearly with variation in RNAP abundance. Specifically, the propagation of RNAP abundance variation to expressed protein noise is inversely proportional to the concentration of RNAP, which suggests that the change in noise that results from RNAP fluctuations is determined by the fraction of promoters that is not occupied by RNAP.

<sup>1</sup>Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Korea. <sup>2</sup>Department of Chemistry and Institute of Innovative Functional Imaging, Chung-Ang University, Seoul 156-756, Korea. <sup>3</sup>School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang 790-784, Korea. Correspondence and requests for materials should be addressed to N.K.L. (email: nklee@postech.ac.kr) or to J.S. (email: jaeyoung@cau.ac.kr)

Even a population of genetically identical cells exhibits different levels of protein expression, which is termed cell-to-cell variation or noise<sup>1,2</sup>. Noise, an unavoidable phenomenon in biochemical reactions, often shows crucial functional roles in cells, such as resulting in state switching, differentiation and cell fate determination<sup>3–6</sup>. Typically noise is generated during the two central gene expression processes, that is, transcription by RNA polymerase (RNAP) and translation by the ribosome; these processes are associated with a small number of templates, such as genes and messenger RNAs (mRNAs)<sup>7,8</sup>. When protein expression levels are low (<10 copies per cell), intrinsic noise, originating from the innate stochasticity of biochemical reactions<sup>2</sup>, is the dominant source of total noise: intrinsic noise is inversely proportional to the protein expression level<sup>9</sup>. However, as the protein expression level increases, the total noise decreases and reaches a saturation floor, or lower limit, in both bacteria and yeast<sup>9–12</sup>. This limit is thought to originate mostly from extrinsic noise, that is, the variation in gene expression brought by varying cellular environments that are extrinsic but influence the rate of gene expression, which is a dominant source of total noise when the protein expression level is sufficiently high<sup>1,9,13–15</sup>.

Various factors, such as cellular differences due to the cell cycle<sup>16,17</sup> and variation in the concentration of essential cellular proteins, such as RNAP, ribosomal proteins, transcription factors and other metabolites, contribute to extrinsic noise<sup>15,18–20</sup>. Indeed, cell cycle differences contribute to extrinsic noise in bacteria, but a substantial amount of extrinsic noise remains following cell cycle correction<sup>16</sup>. Although extrinsic noise is believed to be the major factor determining total noise<sup>1</sup>, how noise sources other than the cell cycle generate and control extrinsic noise has not yet been properly addressed experimentally. For example, how the gene expression machinery proteins, such as RNAP and ribosomal proteins, generate or regulate extrinsic noise has not been investigated, because these proteins are essential factors: variation in their concentrations or distributions significantly alters gene expression profiles. Through the use of different promoter- or ribosome-binding site sequences to vary expression levels, the effects of transcriptional and translational efficiency on noise have been described<sup>7,21</sup>. However, these studies did not directly vary the concentration of RNAP.

Here, we develop a new gene expression system for measuring noise in RNAP expression together with the noise in the expression of downstream proteins. To control the RNAP expression level so that variations in its expression would have a minor effect on cell growth conditions, we use yellow-fluorescent-protein (YFP)-fused T7 RNAP in *Escherichia coli* and placed two fluorescent reporter genes, cyan-fluorescent protein (CFP) and mCherry, under the control of identical copies of the T7 RNAP promoter (Fig. 1a). Tagging T7 RNAP with YFP allows us to monitor the concentration of T7 RNAP in individual cells. The two fluorescent proteins, CFP and mCherry, are used to measure both intrinsic noise, that is, variations between CFP and mCherry expression levels within the same cell, and extrinsic noise, that is, variations in CFP or mCherry expression levels between cells (Fig. 1b)<sup>1</sup>. We observe that RNAP concentration variation indeed propagates to downstream protein expression noise, contributing only to extrinsic noise. The efficiency of the propagation of RNAP noise to downstream protein noise is inversely proportional to the concentration of RNAP. Remarkably, when RNAP noise is diminished, even without correcting for cell cycle heterogeneity and possible variations in the concentrations of other extrinsic noise factors, extrinsic noise is markedly reduced to 30% of the noise limit observed in *E. coli* proteins<sup>9</sup>. This result suggests that the transcriptional noise

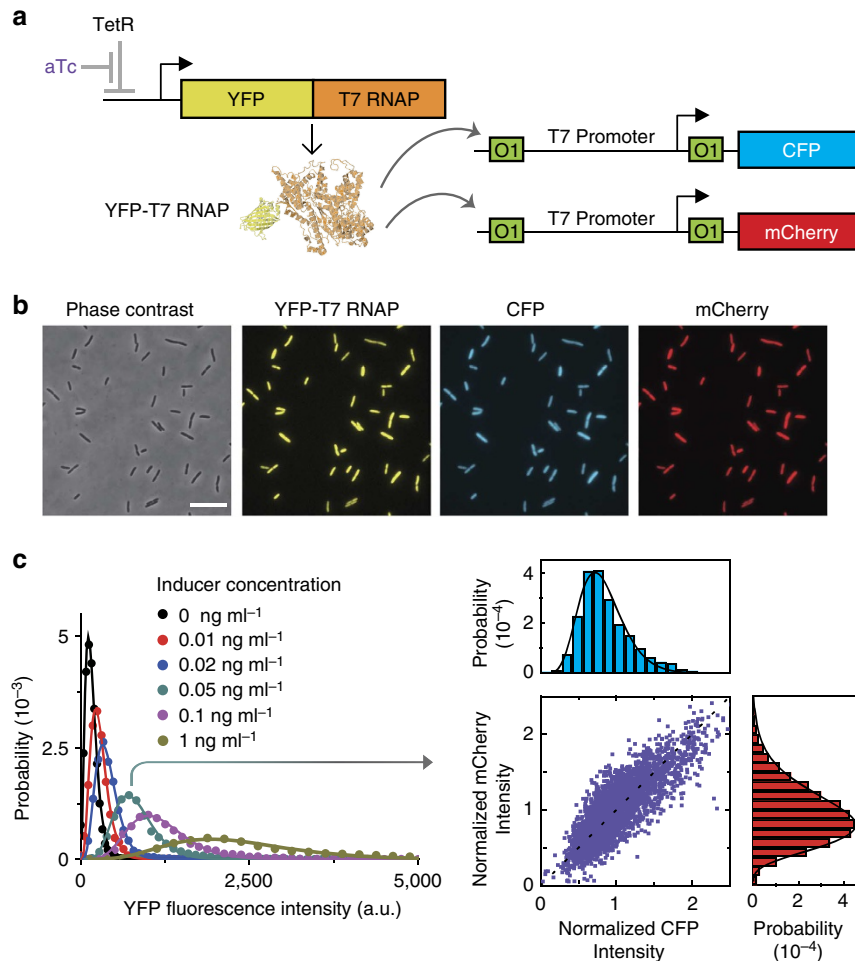
caused by RNAP is the major source of the total noise of highly expressed proteins in *E. coli*.

## Results

**CFP and mCherry expression via T7 RNAP in *E. coli*.** Intrinsic and extrinsic noise is typically measured based on the pairwise correlation between two fluorescent reporters<sup>1,18</sup>. Because RNAP is an essential protein, variation in RNAP concentration will alter cell growth conditions and metabolism significantly, thus investigation of RNAP-dependent effects on noise is challenging. To avoid such undesirable changes in growth conditions due to variation in RNAP concentrations, we used T7 RNAP instead of endogenous *E. coli* RNAP (Fig. 1a). We incorporated the genes encoding CFP and mCherry into the *E. coli* chromosome under the control of the T7 RNAP promoter (Fig. 1a): we replaced *galK* with *cfp* and *intC* with the *mCherry* gene, as reported by Elowitz *et al.*<sup>1</sup>, to maintain identical copy numbers of both genes during the cell cycles (Supplementary Fig. 1). T7 RNAP fused with YFP was expressed from a low-copy-number plasmid (pNL001), and its expression level was controlled by the concentration of the inducer anhydrotetracycline (aTc). After growing cells in M9 minimal medium with glucose, the relative concentrations of YFP-T7 RNAP, CFP and mCherry in individual cells were measured by fluorescence microscopy (Fig. 1b). The expression level of T7 RNAP was effectively controlled by the aTc concentration, and the observed T7 RNAP distributions fit gamma distributions well (Fig. 1c)<sup>22</sup>. The intrinsic and extrinsic noise of each protein's expression level was calculated based on pairwise correlations between the concentrations of CFP and mCherry (Fig. 1d)<sup>1</sup>. Figure 2a indicates that the expression levels of CFP and mCherry are dependent on T7 RNAP concentration: at high T7 RNAP concentrations, the expression levels of both proteins are close to saturation (Fig. 2a).

**Quantifying the noise of downstream proteins.** From the distributions of RNAP concentrations (Fig. 1c), we measured noise in T7 RNAP under each inducer concentration (Fig. 2b). The noise was calculated based on variance over the squared mean of the T7 RNAP concentration in a population ( $\eta^2 = (\sigma/\mu)^2$ ). The noise in T7 RNAP ( $\eta_{\text{RNAP}}^2$ ) converged to 0.2, which is relatively high compared with the noise of other proteins expressed from the *E. coli* genome<sup>9</sup>. Figure 2c shows the noise of downstream proteins at each inducer concentration. The intrinsic noise ( $\eta_{\text{int}}^2$ , green triangles) was inversely proportional to mCherry expression, which varied due to the different concentrations of T7 RNAP. This is consistent with previous work, in which protein expression level was modulated by different inducer concentrations<sup>1</sup>. The intrinsic noise was relatively low, contributing <20% of the total noise, and the majority of noise therefore came from extrinsic noise ( $\eta_{\text{ext}}^2$ , red circles) in this gene-expression system. It is to be noted that the extrinsic noise of the downstream proteins was consistently lower than the noise in RNAP itself, regardless of protein expression level (Fig. 2d). This observation indicates that RNAP noise is not simply added or amplified to downstream protein noise, but attenuated by an unidentified factor.

Thus, we further investigated how RNAP noise propagates to affect downstream protein levels<sup>23</sup>. To directly prove the propagation of RNAP noise to downstream proteins, one had to vary RNAP noise while a constant mean RNAP concentration was maintained. However, such conditions are not practically achievable using natural gene-expression systems due to the noise itself. We therefore investigated the mode of noise propagation by randomly generating cell distributions with pre-determined means and s.d. of the RNAP concentration from the total



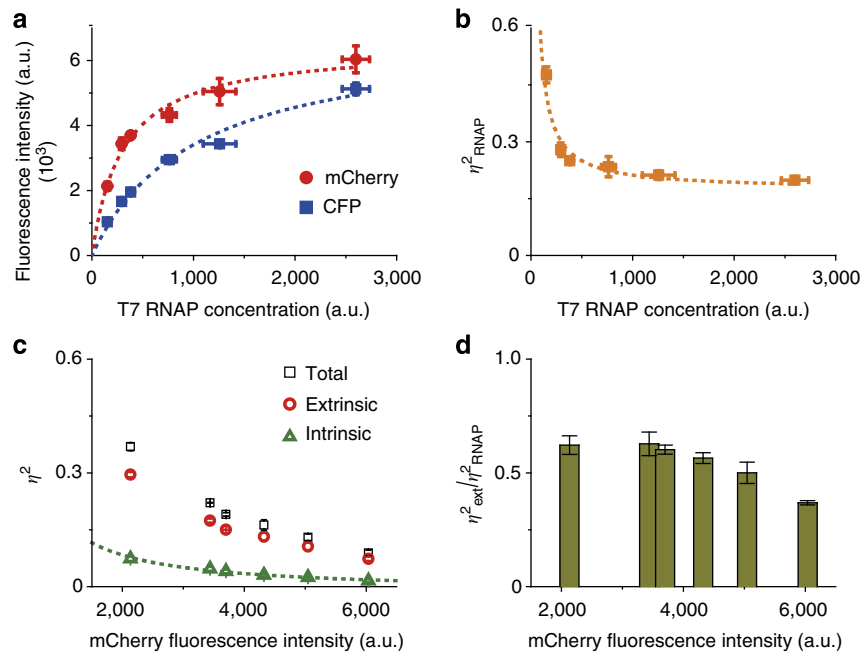
**Figure 1 | Genetic design for achieving T7 RNAP-driven expression of dual fluorescent proteins in *E. coli*.** (a) A schematic description of the T7 RNAP-driven gene expression system. YFP-fused T7 RNAP is expressed from a tetracycline promoter in the pNLO01 plasmid. The inducer anhydrotetracycline (aTc), which binds to the tet repressor, controls the expression level of YFP-fused-T7 RNAP. Then, the genes encoding CFP and mCherry, which are integrated into *E. coli* genomic DNA, are expressed by T7 RNAP. Two Lac operators (O1) are located upstream and downstream of the T7 promoter. To achieve full expression of CFP and mCherry, 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) was added. (b) Representative phase contrast, YFP, CFP and mCherry images of *E. coli* NLO10 carrying pNLO01. Scale bar, 10  $\mu\text{m}$ . (c) Steady-state distributions of YFP-T7 RNAP at different aTc concentrations. Each distribution fits a gamma distribution well (solid curves). (d) Correlation between the CFP and mCherry intensity in the NLO10 strain, in which T7 RNAP is induced under an aTc concentration of 0.05  $\text{ng ml}^{-1}$  (green curve in Fig. 1c). Each point represents a cell, for which fluorescence intensities are normalized by the average intensities of the cell population. The distributions of the CFP (top) and mCherry (right) intensities are described well by the gamma functions (solid curves).

collection of cells (Fig. 3a, left panel). We combined all of the cells generated under various aTc concentrations in Fig. 1c, and then randomly selected 500 cells from the total cell data ( $\sim 40,000$  cells) to obtain pre-determined means and s.d. of the RNAP concentration. To mimic the natural distribution of expression levels, we selected cells so that the resulting distribution of the T7 RNAP concentration in the subset of cells corresponded to a gamma distribution (Fig. 3a, middle panel)<sup>9,22</sup>. Cells with similar concentrations of T7 RNAP expressed similar levels of CFP and mCherry, regardless of aTc induction conditions (Supplementary Fig. 2). Using the selected subset of cells, we then calculated the RNAP noise and the intrinsic and extrinsic noises of downstream proteins (Fig. 3a, middle and right panels). We repeated this process with fixed average RNAP concentrations while varying RNAP noise (Fig. 3b–e).

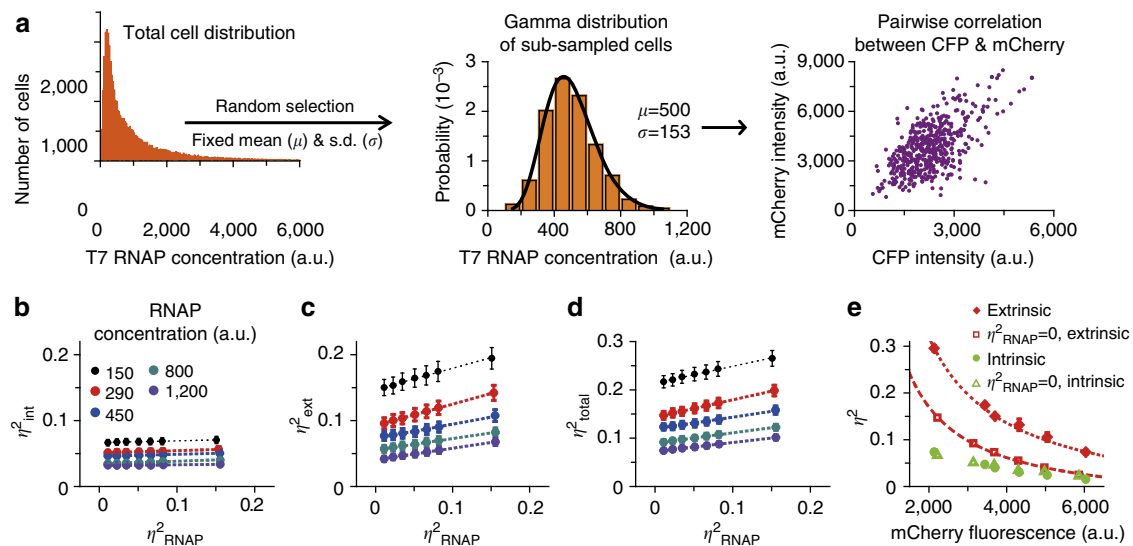
Figure 3b demonstrates that the intrinsic noise of downstream proteins was nearly invariant with respect to RNAP noise. However, extrinsic noise increased with an increase in RNAP noise (Fig. 3c), and as a consequence, the total noise increased

linearly (Fig. 3d). Figure 3e shows the contribution of RNAP noise to intrinsic and extrinsic noise: although removing RNAP noise had a negligible effect on intrinsic noise, extrinsic noise was significantly reduced. These results clearly demonstrate the long-lasting prediction that RNAP concentration variation, one of the global factors involved in gene expression, contributes to extrinsic, but not intrinsic noise<sup>1,2,15,16</sup>.

**Real-time analysis of the single-cell lineage.** Because we measured the concentration of RNAP at a given time point for the noise analysis, the history of the fluctuation in RNAP concentration was not fully considered in our sampling approach<sup>24,25</sup>. For example, although the RNAP concentration may be identical in two cells at the sampling point, the expression levels of the downstream proteins could differ in these two cells depending on the fluctuation history of the RNAP concentration of each cell. In this case, selecting a subset of cells to have a specific amount of RNAP noise does not guarantee that the noise



**Figure 2 | Quantification of T7 RNAP noise and the intrinsic and extrinsic noises of downstream proteins.** (a) Expression levels of CFP and mCherry by T7 RNAP at various aTc concentrations. (b) T7 RNAP noise obtained from the distribution of the YFP fluorescence intensity (Fig. 1c) as a function of the mean T7 RNAP concentration. The dotted line is fitted to the function  $a/R + b^{1.15}$ , where  $R$  is the T7 RNAP concentration;  $a = 38$ ; and  $b = 0.17$ . (c) The total (black open squares), extrinsic (red open circles) and intrinsic (green open triangles) noise of downstream protein expression, measured at each aTc concentration. Intrinsic noise is fitted to the function  $a/P + b$ , where  $P = \text{mCherry fluorescence intensity}$ ;  $a = 196$ ; and  $b = -0.01$  (dotted line). (d) The ratio of extrinsic noise to the RNAP noise, calculated at each aTc concentration. Error bars represent the mean  $\pm$  s.d. of three independent experiments.

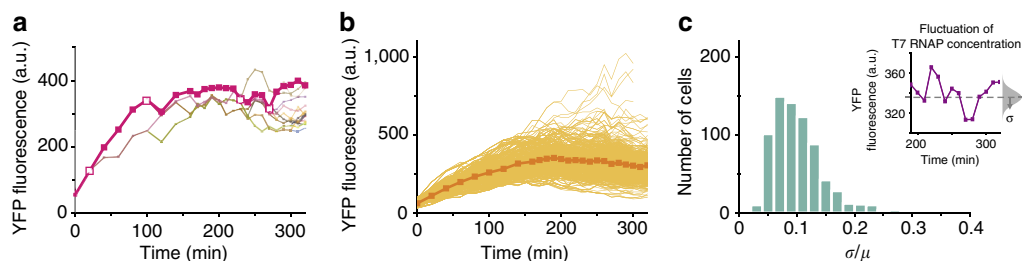


**Figure 3 | Analysis of RNAP noise-dependent intrinsic and extrinsic noise.** (a) Random generation of cell distributions with a pre-determined mean and s.d. of the RNAP concentration from the total collection of cells. Left, total cell distributions obtained by combining all cells from Fig. 1c ( $\sim 40,000$  cells). Middle, distribution of the subset of cells (500 cells) randomly selected from the total cell distribution to satisfy a given gamma distribution with a pre-determined mean ( $\mu$ ) and s.d. ( $\sigma$ ). Right, pairwise correlation between the CFP and mCherry intensities in the sub-sampled cells. (b-d) Intrinsic, extrinsic and total noise of the sub-sampled cells depending on RNAP noise. The dotted lines represent linear guidelines for clarity. Error bars indicate  $\pm$  s.d.,  $n = 1,000$ . (e) Contribution of RNAP noise to intrinsic and extrinsic noise. We compared the gene expression noises shown in Fig. 2c with those at an RNAP noise = 0. The intrinsic and extrinsic noises at an RNAP noise = 0 (open symbols) were obtained from the y axis intercepts resulting from linear fitting in b and c, respectively.

of the downstream proteins of the selected cells includes the effect of the pre-determined amount of RNAP noise that we chose during the selection step. However, for less marked fluctuations of RNAP concentration, the RNAP concentration at a given time maintains much of the correlation with the concentrations of

downstream proteins; thus, our sampling approach can reduce or increase the contribution of RNAP concentration noise on the extrinsic noise of downstream proteins.

We measured the real-time fluctuation in RNAP concentration by monitoring its concentration at the single-cell level using a gel



**Figure 4 | Real-time observation of fluctuation in T7 RNAP concentration.** (a) Time traces of YFP-T7 RNAP fluorescence intensity in single-cell lineages. The pink bold line follows a single lineage, and the other daughter cells have different light-coloured lines. Each open rectangle indicates cell division. (b) All time traces of T7 RNAP concentration (total 792 traces). The bold orange line represents the average intensity of YFP-T7 RNAP. (c) Analysis of the fluctuation in T7 RNAP concentration. After a steady-state condition was reached, we obtained the mean ( $\mu$ ) and s.d. ( $\sigma$ ) of T7 RNAP intensity from 2 h of each time trace (inset). The histogram shows the s.d. of T7 RNAP intensity relative to the mean T7 RNAP intensity in a single time trace. The average relative s.d. ( $\sigma/\mu$ ) of T7 RNAP intensity in the time traces is  $10 \pm 4\%$ .

pad (Supplementary Fig. 3; Supplementary Movie 1). The concentration of T7 RNAP increased with the addition of the inducer, aTc, and then reached a steady-state condition after  $\sim 2$  h of induction (Fig. 4a). The concentrations of T7 RNAP at the steady state presented considerable heterogeneity within the cell population (Fig. 4b), but the T7 RNAP concentrations of most cells remained at the steady-state level for 3 h after reaching the steady state. The s.d. of the normalized T7 RNAP concentration at the steady state was  $\sim 10\%$  for 2 h (Fig. 4c); that is, the T7 RNAP concentration fluctuated  $< 20\%$  ( $2\sigma$ ) with a 95% confidence interval. In line with this real-time measurement of T7 RNAP fluctuation, the correlation coefficient ( $r$ ) between the RNAP concentration and the concentrations of the downstream proteins of the cells used for the analysis in Fig. 3 was measured to be 0.88 for CFP and 0.74 for mCherry from the joint-probability distributions between the RNAP concentration and CFP and mCherry concentrations, respectively (Supplementary Fig. 4). This result shows that RNAP concentration does not fluctuate to an extent that significantly disrupts the correlation between RNAP and its downstream proteins at the sampling time.

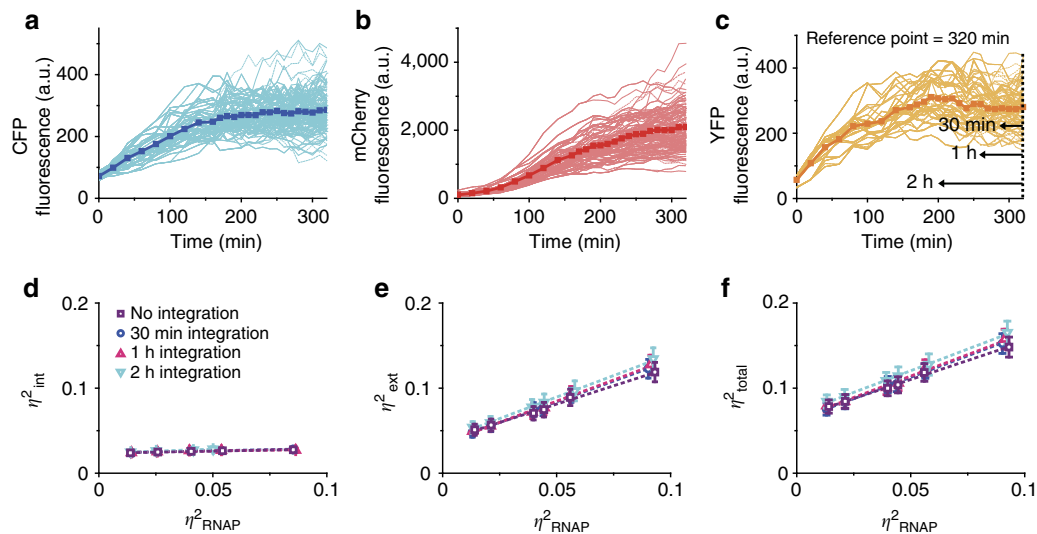
This result may also explain the observation that the intrinsic noise of downstream proteins is invariant under changes in the magnitude of the T7 RNAP concentration noise<sup>20</sup>. Because temporal fluctuations in T7 RNAP abundance remain small relative to variation in RNAP abundance between cells for extended time intervals, the autocorrelation lifetime of RNAP fluctuation is much longer than the average lifetime of intrinsic fluctuations of the downstream protein, and the adjustment of downstream protein gene expression to changes in RNAP concentration is effectively instantaneous. Therefore, the possible contribution by dynamic fluctuation of RNAP concentration<sup>24</sup> to intrinsic noise in downstream gene expression, as measured by a dual reporter system, could be negligible under changes in the magnitude of T7 RNAP expression noise when the mean of the T7 RNAP concentration is held constant.

Next, to quantitatively investigate the effect of the T7 RNAP concentration history on downstream noise, we measured the concentrations of the downstream proteins together with T7 RNAP concentration at the single-cell level on a gel pad for  $\sim 5$  h (Fig. 5a,b). We observed that CFP and mCherry concentrations reached a steady state after  $\sim 3$  and 4.5 h of induction, respectively. The difference in expression curve may originate from the difference between the maturation times of CFP and mCherry. To evaluate the effect of the previous T7 RNAP expression levels on downstream protein noises, we integrated the fluorescence intensity of T7 RNAP in each cell lineage for 0 min (no integration), 30 min, 1 h and 2 h ( $\sim 1.4$  cell cycles on a gel pad) prior to the reference point (320 min) (Fig. 5c). Integrating

T7 RNAP intensity for a specific period of time would account for the history of the RNAP concentration fluctuation. Next, we used the integrated fluorescence intensities of T7 RNAP for the random cell selection with fixed mean and variance as in Fig. 3a. We used the CFP and mCherry concentrations of the selected cells measured at 320 min for downstream noise analysis (Fig. 5d–f) (see Methods for a detailed description). We also tested CFP and mCherry concentrations at time points other than 320 min for the noise analysis in a similar manner (Supplementary Fig. 5). In all cases, we found that the magnitudes of intrinsic and extrinsic noise of the downstream proteins were nearly independent of the integration time of T7 RNAP concentration within the error bar, indicating that fluctuation of T7 RNAP is not significant. Importantly, the intrinsic noise was invariant to RNAP noise variation, while extrinsic noise was linearly proportional to the RNAP noise in the gel-pad experiment. These results are in line with the results in Fig. 3, which were obtained from cells grown in a shaking culture. These results consistently support our observation that RNAP concentration variation contributes mostly to extrinsic noise. In addition, selecting cells with particular RNAP concentrations is a valid approach for controlling extrinsic perturbation by RNAP in our gene-expression system, because the fluctuation history of the T7 RNAP concentration presented a minimal effect on downstream protein noises.

Calculating the dependence of downstream protein noises on RNAP noise from a single-cell lineage at the steady-state RNAP concentration is an alternative way of investigating RNAP noise effect (Supplementary Fig. 6). However, RNAP noise in a single-cell lineage at the steady state was extremely low (Fig. 4c) and thus it was difficult to observe any dependence of protein noise on RNAP noise (Supplementary Fig. 6a–c). The noise of RNAP measured from a single-cell lineage (Fig. 4c) was smaller than the noise from a cell population. This difference occurred because the observation duration for a single-cell lineage was not long enough to satisfy ergodicity, consistent with other works<sup>9,15</sup>. Thus, a subsampling approach is required in this study.

Then, because the real-time tracking data in Fig. 5a–c reported variation in RNAP concentration as well as that of the downstream protein concentrations of each single-cell lineage, we calculated RNAP noise and intrinsic, extrinsic and total noise of the downstream proteins of each single-cell lineage using concentration data from the induction point (0 min) to 320 min (total 729 cells) (Supplementary Fig. 6d–f). This noise includes the variation in the RNAP concentration caused by induction; thus, this noise is not the same as traditionally defined noise, which is typically obtained at steady state. However, this noise can be used to approximate the large dynamic effects of RNAP



**Figure 5 | Evaluation of the effect of the T7 RNAP concentration history on downstream protein noise at the single-cell lineage level.** Together with T7 RNAP intensity, the intensities of CFP and mCherry were measured for  $\sim 5$  h. We used 729 traces for the analysis but showed only 200 traces in **a–c** for clarity. Representative time traces of **(a)** CFP fluorescence intensity and **(b)** mCherry fluorescence intensity. The bold blue and red lines indicate the average. **(c)** Representative time traces of YFP-T7 RNAP fluorescence intensity. The integration times for YFP-T7 RNAP concentration of a single-cell lineage, 0, 30, 1 and 2 h from the reference point (320 min), are indicated. **(d–f)** Dependence of the intrinsic, extrinsic and total noise of the sub-sampled cells on RNAP noise. Here, we used the integrated YFP-T7 RNAP concentrations from Fig. 5c for the subsampling analysis, and the concentrations of CFP and mCherry of the selected cells measured at 320 min were used for noise analysis in a population. The dotted lines represent linear guidelines for clarity. Error bars indicate  $\pm$  s.d.,  $n = 1,000$ .

concentration fluctuation. Interestingly, intrinsic noise was nearly independent of RNAP noise, while extrinsic noise increased linearly as the RNAP noise of each cell lineage increased (Supplementary Fig. 6g); this result further supports our finding that RNAP noise propagates to the extrinsic noise of the downstream proteins.

**Model of RNAP noise propagation.** We note that the dependency of extrinsic noise on RNAP noise (slope in Fig. 3c) decreased as the expression level of T7 RNAP was increased: the magnitude of the slope determines the ‘propagation efficiency’ of RNAP noise to downstream proteins. This observation indicated that the propagation of RNAP noise was dependent on T7 RNAP concentration. As RNAP concentration was increased, the extent of the propagation of RNAP noise was reduced. To assess what determines the propagation efficiency of RNAP noise, we considered the kinetic scheme of transcription given by the following description:

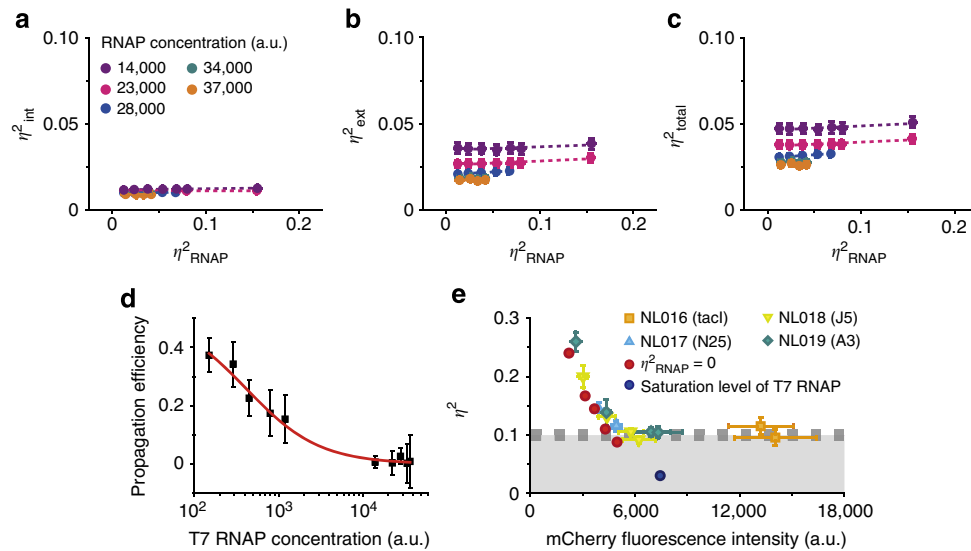


where  $k_1$  and  $k_{-1}$  are the rate coefficients for the association and dissociation reactions between RNAP and the promoter, respectively, and  $k_2$  is the transcriptional initiation rate for mRNA synthesis. According to Michaelis–Menten enzyme kinetics, the steady-state rate of the above transcriptional reaction is proportional to  $\theta = KN_{\text{RNAP}}/(1 + KN_{\text{RNAP}})$ , where  $K$  is defined by  $K = k_1/(k_2 + k_{-1})$  and  $N_{\text{RNAP}}$  is the RNAP concentration. Here,  $\theta$  and  $K$  can be interpreted as the fraction of the promoter occupied by RNAP during the transcriptional reaction and the affinity of RNAP for the promoter, respectively. Then, the overall transcriptional rate ( $R_{\text{TX}}$ ) can be defined as  $R_{\text{TX}} = k_{\text{TX}}\theta$  ( $0 < \theta < 1$ ), where  $k_{\text{TX}}$  denotes the overall transcriptional rate when  $\theta$  is unity. By definition,  $k_{\text{TX}}$  is independent of RNAP concentration. In contrast, the fraction of RNAP occupation (or the unoccupied fraction,  $1 - \theta$ ) is dependent on RNAP

concentration. Thus, only the fraction of promoter occupation by T7 RNAP is a potential factor responsible for the propagation of RNAP noise in this kinetic scheme.

If the fraction of occupation becomes unity (or the unoccupied fraction becomes zero), fluctuations in RNAP concentration would have no effect on the variation of transcription and protein expression. Thus, if the fraction of occupation is related to noise propagation, the slope, or the propagation efficiency of RNAP noise as shown in Fig. 3c, should converge to zero as the RNAP concentration reaches a saturation level, that is, when the fraction of occupation is close to unity. To test this hypothesis, we prepared a high-copy-number plasmid (pNL002) to express T7 RNAP at sufficiently high levels to saturate the expression levels of both CFP and mCherry (Supplementary Fig. 7). We performed the same random selection procedure for subsets of cells to achieve a gamma distribution as indicated in Fig. 3a (Supplementary Fig. 8) and quantified intrinsic and extrinsic noise values depending on RNAP noise at constant RNAP concentrations (Fig. 6a–c). Consistent with our prediction, extrinsic and intrinsic noise were nearly invariant with respect to variation in RNAP noise (slope  $< 0.03$ ) (Fig. 6b); that is, the propagation efficiency of RNAP noise was close to zero. As a consequence, the total noise was also invariant with respect to the RNAP noise (Fig. 6c). We found that the propagation efficiencies obtained from the extrinsic noise well fit the function of the unoccupied fraction,  $(1 - \theta) = a/(1 + KN_{\text{RNAP}})$  (Fig. 6d). These results demonstrated that the propagation of RNAP noise to downstream protein noise depends on the RNAP concentration. We propose that the fraction of occupation of the promoter determines the efficiency of noise propagation. For a given distribution of RNAP concentrations, a gene with a stronger promoter is less susceptible to RNAP noise, thus propagating RNAP noise to downstream protein expression less efficiently.

**Dominant source of the extrinsic noise in *E. coli*.** Recently, Taniguchi *et al.*<sup>9</sup> reported that the noise associated with bacterial



**Figure 6 | Analysis of RNAP-dependent intrinsic and extrinsic noise at saturation conditions.** (a) Intrinsic, (b) extrinsic and (c) total noises of the sub-sampled cells depending on RNAP noise using the high-copy-number plasmid (pNLO02). We expressed T7 RNAP at a sufficiently high level using pNLO02 to saturate the expression levels of CFP and mCherry (Supplementary Fig. 7), and generated sub-sampled cell distributions of RNAP concentrations with pre-determined means and s.d. from the total collection of cells (Supplementary Fig. 8). The dotted lines represent linear guidelines for clarity. Error bars indicate  $\pm$  s.d.,  $n = 1,000$ . (d) The slope (susceptibility to RNAP noise) of the extrinsic noise depending on the T7 RNAP concentration from Figs 3c and 6b. The red line is fitted to the function of the unoccupied fraction of the promoter by RNAP,  $\alpha/(1 + KR)$ , where  $R$  and  $K$  denote T7 RNAP concentration and the RNAP affinity for the promoter, respectively (least squares,  $R^2 = 0.99$ ). Error bars indicate  $\pm$  s.d.,  $n = 10$ . (e) Comparison of mCherry noise induced by various *E. coli* RNAP promoters with the noise induced by T7 RNAP promoters. The rectangle, triangle and diamond symbols represent the gene expression noise under various *E. coli* RNAP promoters<sup>20</sup>. Expression levels in each strain were modulated by applying different IPTG concentrations. When mean mCherry intensity is  $> 6,000$  (a.u.), noise becomes saturated at  $\sim 0.1$  (dashed line), consistent with the findings of Taniguchi *et al.*<sup>9</sup> Error bars represent the mean  $\pm$  s.d. of three independent experiments. The red-filled circles represent gene expression noise under the T7 promoter when the noise induced by the RNAP noise is removed, as estimated from the linear fitting shown in Fig. 3c. The blue-filled circle represents gene expression noise under the T7 promoter at saturation-level T7 RNAP concentrations using pNLO02.

proteins converged to a limit of 0.1 when they measured the noises of nearly 1,000 *E. coli* proteins. This noise limit, generally attributed to residual extrinsic noise<sup>1,9,13,26</sup>, has also been observed in yeast<sup>12</sup>. Because the noise measured by counting mRNA copy numbers is mostly intrinsic<sup>21,27</sup>, translational noise or the noise generated by other metabolites is attributed to the source of this extrinsic noise limit<sup>18</sup>.

We tested whether the extrinsic noise limit was retained when the portion of extrinsic noise induced by RNAP noise was minimized. We first tested whether the extrinsic noise limit expressed by endogenous *E. coli* RNAP was observed under our noise-measurement conditions. We incorporated the *mCherry* gene into the *E. coli* chromosome under various *E. coli* RNAP promoters and then measured the total noise. As expected, the noise reached a saturation floor at 0.1 for cells grown in M9 glucose minimal medium (Fig. 6e), consistent with the findings of Taniguchi *et al.*<sup>9</sup> However, when saturating levels of T7 RNAP were induced using pNLO02 to minimize the noise caused by RNAP fluctuations (Fig. 6c), the total noise was reduced to 0.03 (Fig. 6e, blue-filled circles). This result demonstrates that the noise generated by ribosomes and translational fluctuation, as well as the fluctuation of other metabolisms can be as low as 0.03 in *E. coli*. The red-filled circles in Fig. 6e represent the total noise observed after subtracting RNAP noise-induced extrinsic noise, which was consistently smaller than the noise of *E. coli* RNAP-driven expression at identical mCherry expression levels. Because the translation process in our bacterial system is identical to that of natural *E. coli*, the reduction of the minimum noise from 0.1 in natural *E. coli* to 0.03 in our cell system can be mostly ascribed to replacement of the bacterial transcription system with the viral system. A total of 70% of the minimum noise is a result of

fluctuation in the transcriptional process of highly expressed genes in *E. coli*, whereas the noise resulting from translational fluctuation is  $< 30\%$  of the minimum gene expression noise in natural *E. coli* genes.

Transcriptional bursting, the origin of which is still unknown in bacteria<sup>21,27</sup>, transcription factor dynamics<sup>28</sup>, transcriptional pausing or the complicated process of *E. coli* RNAP initiation with sigma factor binding, such as open complex formation and abortive initiation<sup>29,30</sup>, may contribute to the transcriptional noise of *E. coli* RNAP. Another possible mechanism for generating such noise is transient variation in the concentration of *E. coli* RNAP available for the specific promoter due to intensive RNAP use by other promoters. Generally, bacteria and higher organisms display high-copy-number RNAPs. However, they also exhibit high copy numbers of promoters for RNAP binding, and thus the available number of RNAPs near a specific promoter may transiently fluctuate. Our results indicate that such local effects, if present, should be transmitted to downstream protein expression noise. The probability of promoter occupation by RNAP may determine what fraction of RNAP noise is transmitted to downstream protein noise.

In summary, we have demonstrated that RNAP noise is transmitted to downstream protein noise as an extrinsic noise factor (Supplementary Fig. 9). The propagation of RNAP noise to downstream protein noise was attenuated and is inversely proportional to RNAP concentration. When the downstream noise induced by RNAP noise was removed by saturating promoter occupancy, the total noise was reduced to 30% of the lower limit of the noise in *E. coli*, suggesting that noise generated at the transcriptional level is the dominant source of the extrinsic noise of high-copy-number proteins in bacteria.

## Methods

**pNL001 and pNL002 plasmids for T7 RNAP expression.** T7 RNAP fused with YFP (*Venus*)<sup>31</sup> was expressed from a low-copy-number plasmid (pNL001), which was tightly regulated by the tet promoter and carried the ampicillin-resistant gene (*ampR*). The expression of T7 RNAP was induced by anhydrotetracycline (aTc). When a saturation level of T7 RNAP was required, YFP-fused T7 RNAP was expressed from a high-copy-number plasmid (pNL002). L-arabinose was used to induce T7 RNAP expression from pNL002, which also carries *ampR* for antibiotic selection.

**NL010 strain.** We constructed an *E. coli* strain (NL010) expressing CFP and mCherry via T7 RNAP. We replaced *intC* with the *mCherry* gene and *galK* with *cfp* (the CFP gene) in the *E. coli* chromosome using  $\lambda$ -RED recombination. The inserted region included two lac operators (O1), a T7 promoter, a ribosome-binding site, a gene encoding a fluorescent protein and the *cat* (chloramphenicol resistance) gene (Fig. 1a). We inserted two O1 sequences both upstream and downstream of the T7 promoter to reduce the background level of protein expression. The *cat* gene was removed using a FRT cassette following successful insertion of the target gene into genomic DNA.

**NL016–NL019 strains.** To test the lower noise limit in *E. coli*, we inserted the *mCherry* gene under various *E. coli* RNAP promoters. The promoter sequences employed are referred to in the literature<sup>32</sup>. The inserted sequences included a promoter, O1, ribosome-binding site, the *mCherry* gene and *cat*. We replaced the lac operon with these sequences using  $\lambda$ -RED recombination.

(promoter sequences: from -50 to +1)

NL016 (PtaC1): 5'-tctgaaatgagctgtgcaacaaatcatcgcgctcgtataatgtgtgga-3'

NL017 (PN25): 5'-tcataaaaattatttgcctcaggaaaattttctgtataatagattca-3'

NL018 (PJ5): 5'-atataaaacgctattgacacaggtggaatttagaataatctgttagta-3'

NL019 (PA3): 5'-ggtaaaacacggttgacaacatgaagtaaacacggtacatgtaccaca-3'

**Cell preparation.** All strains were grown overnight in LB medium at 37 °C from a single colony. The overnight cultures were re-inoculated into fresh M9 medium supplemented with 0.4% glucose, amino acids and vitamins at 1:200 dilution, unless otherwise specified. The cells were grown at 37 °C until reaching an OD<sub>600</sub> ~ 0.3, and a 1-ml aliquot of the cells was then pelleted by centrifugation for 1 min and resuspended in 500  $\mu$ l of fresh M9 medium for washing. The cells were pelleted again by centrifugation and resuspended in a final volume of 5  $\mu$ l fresh M9 medium. Then, 0.8  $\mu$ l of the concentrated cell culture was placed between a coverslip and a 3% low-melting-temperature agarose gel pad (Lonza, #50111) prepared with M9 medium. NL010 cells carrying the pNL001 plasmid were induced with aTc at final concentrations of 0, 0.01, 0.02, 0.05, 0.1 and 1 ng ml<sup>-1</sup> to modulate the expression level of T7 RNAP. To achieve full expression of CFP and mCherry, 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) was added together with aTc. For the NL010 carrying the pNL002 plasmid, M9 medium supplemented with 0.4% glycerol was used to achieve a high expression level of the pNL002 plasmid, together with 0.1 and 0.2% of L-arabinose.

**Microscope setup.** Samples were placed on an inverted microscope (Olympus, IX-71) with a  $\times$  100 oil-immersed objective lens (Olympus). Phase contrast images and three fluorescence images for CFP, YFP and mCherry were acquired at multiple positions using a cooled EMCCD camera (Andor iXon DU897). For CFP, FF01-434/17-25 (excitation), FF01-479/40-25 (emission) and FF452-Di01-25  $\times$  36 (dichroic mirror) were used (purchased from Semrock). For mCherry, FF01-572/28-25 (excitation), FF01-641/75-25 (emission) and FF593-Di02-25  $\times$  36 (dichroic mirror) were used. For YFP, FF01-500/24-25 (excitation), FF01-550/32-25 (emission) and Q525LP (dichroic mirror) were used. Metamorph software (Molecular Devices) was employed to control the automated measurements and maintain focus during data acquisition.

**Image analysis.** The images were analysed using home-built software (Matlab). The program extracted the total fluorescence intensity of individual cells from the phase contrast and fluorescence images automatically and recorded the average fluorescence intensity (concentration) by normalizing the total intensity to the cell area. The intensity of autofluorescence was obtained from NL010 cells without the plasmid expressing YFP-T7 RNAP.

**Noise analysis by generating gamma distributions.** We generated gamma distributions based on the YFP fluorescence intensities in cells to obtain pre-determined means ( $\mu$ ) and s.d. ( $\sigma$ ). We prepared the total cell collection by combining all cells obtained from the measurements performed at different aTc concentrations (Fig. 3a). The number of cells in the entire collection was ~ 40,000. Then we randomly selected 500 cells each time from the total cell collection to satisfy a given gamma distribution:  $p(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/b}$  with  $a = \mu^2/\sigma^2$  and  $b = \sigma^2/\mu$  being fixed, following the procedure described below.

(1) Using input values for the mean ( $\mu$ ) and s.d. ( $\sigma$ ), we calculated  $a = \mu^2/\sigma^2$  and  $b = \sigma^2/\mu$  and obtained the gamma distribution,  $p(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/b}$ .

- (2) We set the size of the bin for the X-axis (YFP intensity) to  $\sigma/2$ . Then, we calculated the number of cells required for each bin to satisfy the given gamma distribution. We set the total number of the sub-sampled cells as 500.
- (3) We randomly selected the required number of the cells for each bin from the total cell collection.
- (4) The selected cells were used to calculate the intrinsic, extrinsic and total noises as follows, as defined by Elowitz *et al.*<sup>1</sup>:

$$\eta_{int}^2 \equiv \frac{\langle (c-m)^2 \rangle}{2\langle c \rangle \langle m \rangle}; \quad \eta_{ext}^2 \equiv \frac{\langle cm \rangle - \langle c \rangle \langle m \rangle}{\langle c \rangle \langle m \rangle}; \quad \eta_{tot}^2 = \eta_{int}^2 + \eta_{ext}^2,$$

where  $c$  and  $m$  represent the CFP and mCherry fluorescence intensity in each cell, respectively.

- (5) The mean and s.d. of the noises were obtained by repeating this process of random selection 1,000 times.

**Real-time observation of single-cell lineage on a gel pad.** For real-time experiments, NL010 cells carrying the pNL001 plasmid were grown in LB overnight. The overnight cultures were re-inoculated into fresh M9 medium supplemented with 0.4% glucose, amino acids and vitamins at 1:200 dilution. For maintenance of the basal expression level of CFP and mCherry, the cells were grown in the presence of 1 mM IPTG at 37 °C until they reached ~0.2 OD<sub>600</sub>. Then, a 500- $\mu$ l aliquot of the cells was pelleted by centrifugation and resuspended in a final volume of 50  $\mu$ l of fresh M9 medium, and 0.5  $\mu$ l of the cells was placed between a coverslip and a 1.5% low-melting-temperature agarose gel pad (Lonza, #50111) prepared with M9 media containing 1 ng ml<sup>-1</sup> aTc and 1 mM IPTG. The sample chamber was attached to a temperature controller (FCS2, Biopetech) and maintained at 37 °C during the experiments. To provide a continuous flow of fresh media, pre-warmed M9 media containing 1 ng ml<sup>-1</sup> aTc and 1 mM IPTG was supplied by a syringe pump at 0.26 ml min<sup>-1</sup>. Images were acquired every 20 min for 160 min and then every 10 min for the remaining time. The acquired images were analysed using the Schnitzcell MATLAB module (provided by the M. Elowitz group, California Institute of Technology)<sup>33</sup>.

**Analysis of the T7 RNAP effect on the protein noise.** To quantitatively investigate the effect of the T7 RNAP concentration history, we integrated YFP-T7 RNAP intensities for each cell lineage for a specific period of time (0, 30, 1 and 2 h) from the reference time point (320 min), as depicted in Fig. 5c. Then, we used the integrated YFP-T7 RNAP intensity of each cell for random cell selection with a fixed mean and variance, similar to our selection in Fig. 3a. We randomly selected 300 cells each time from the 729 total cells to satisfy a given gamma distribution:  $p(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/b}$  with  $a = \mu^2/\sigma^2$  and  $b = \sigma^2/\mu$  being fixed, and  $x$  = the integrated YFP-T7 RNAP intensity. The downstream protein noise was calculated using CFP and mCherry fluorescence intensities from measurements of the selected cells at the reference time point (320 min). We used the subsampling procedure described in 'noise analysis by generating gamma distributions' for the remaining procedure of the subsampling noise analysis. The mean and s.d. of the noise were obtained by repeating this process of random selection 1,000 times (Fig. 5d-f).

## References

1. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
2. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
3. Balazsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).
4. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
5. Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**, 526–529 (2007).
6. Singh, A. & Weinberger, L. S. Stochastic gene expression as a molecular switch for viral latency. *Curr. Opin. Microbiol.* **12**, 460–466 (2009).
7. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
8. Paulson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
9. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
10. Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell.* **45**, 483–493 (2012).
11. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).
12. Newman, J. R. S. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
13. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).



14. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
15. Hensel, Z. *et al.* Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nat. Struct. Mol. Biol.* **19**, 797–802 (2012).
16. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
17. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864 (2006).
18. Maheshri, N. & O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).
19. Raj, A. & van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38**, 255–270 (2009).
20. Shahrezaei, V., Ollivier, J. F. & Swain, P. S. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.* **4**, 196 (2008).
21. So, L. H. *et al.* General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* **43**, 554–560 (2011).
22. Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).
23. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
24. Hilfinger, A. & Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Natl Acad. Sci. USA* **108**, 12167–12172 (2011).
25. Bowsher, C. G. & Swain, P. S. Identifying sources of variation and the flow of information in biochemical networks. *Proc. Natl Acad. Sci. USA* **109**, E1320–E1328 (2012).
26. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
27. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
28. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
29. Campbell, E. A., Westblade, L. F. & Darst, S. A. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. *Curr. Opin. Microbiol.* **11**, 121–127 (2008).
30. Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
31. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90 (2002).
32. Deuschle, U., Kammerer, W., Gentz, R. & Bujard, H. Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* **5**, 2987–2994 (1986).
33. Young, J. W. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88 (2012).

### Acknowledgements

We thank Dr Paul Choi, Dr Dann Huh and Dr Jie Xiao for critical reading of the manuscript. This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MEST) NRF-2013R1A1A2063302, grants (K13040 and K14040) from the Korea Institute of Oriental Medicine and the Ministry of Science ICT and Future Planning, Korea, under the 'IT Consilience Creative Program' (NIPA-2014-H0201-14-1001) supervised by the National IT Industry Promotion Agency to N.K.L., NRF grants (2011-0016412) and the Priority Research Center Program through the NRF of Korea (2009-0093817) to J.S.

### Author contributions

S.Y. and N.K.L. designed the experiments; S.Y. and N.K.L. wrote the manuscript; S.Y. performed all measurements and analysis. S.K., C.K. and H.J.A. performed cloning and helped in measurements. J. S. provided a model and new analysis and Y.R.L., J.-H.K., and J.S. performed analysis.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

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

**How to cite this article:** Yang, S. *et al.* Contribution of RNA polymerase concentration variation to protein expression noise. *Nat. Commun.* **5**:4761 doi: 10.1038/ncomms5761 (2014).