

*This copy is for your personal, non-commercial use only.*

**If you wish to distribute this article to others**, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of July 29, 2010):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/329/5991/533>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/329/5991/533/DC1>

This article **cites 33 articles**, 15 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/329/5991/533#otherarticles>

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/329/5991/533#otherarticles>

This article appears in the following **subject collections**:

Molecular Biology

[http://www.sciencemag.org/cgi/collection/molec\\_biol](http://www.sciencemag.org/cgi/collection/molec_biol)

# Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells

Yuichi Taniguchi,<sup>1\*</sup> Paul J. Choi,<sup>1\*</sup> Gene-Wei Li,<sup>1,2\*</sup> Huiyi Chen,<sup>1,3\*</sup> Mohan Babu,<sup>4</sup> Jeremy Hearn,<sup>1</sup> Andrew Emili,<sup>4,5</sup> X. Sunney Xie<sup>1†</sup>

Protein and messenger RNA (mRNA) copy numbers vary from cell to cell in isogenic bacterial populations. However, these molecules often exist in low copy numbers and are difficult to detect in single cells. We carried out quantitative system-wide analyses of protein and mRNA expression in individual cells with single-molecule sensitivity using a newly constructed yellow fluorescent protein fusion library for *Escherichia coli*. We found that almost all protein number distributions can be described by the gamma distribution with two fitting parameters which, at low expression levels, have clear physical interpretations as the transcription rate and protein burst size. At high expression levels, the distributions are dominated by extrinsic noise. We found that a single cell's protein and mRNA copy numbers for any given gene are uncorrelated.

Gene expression is often stochastic, because gene regulation takes place at a single DNA locus within a cell. Such stochasticity is manifested in fluctuations of mRNA and protein copy numbers within a cell lineage over time and in variations of mRNA and protein copy numbers among a population of genetically identical cells at a particular time (1–4). Because both

manifestations of stochasticity are connected, measurement of the latter allows the deduction of the gene expression dynamics in a cell (5). We aim to characterize such mRNA and protein distributions in single bacteria cells at a system-wide level.

Single-cell mRNA profiling has been carried out with cDNA microarray (6) and mRNA-sequencing (mRNA-seq) (7); nevertheless, these

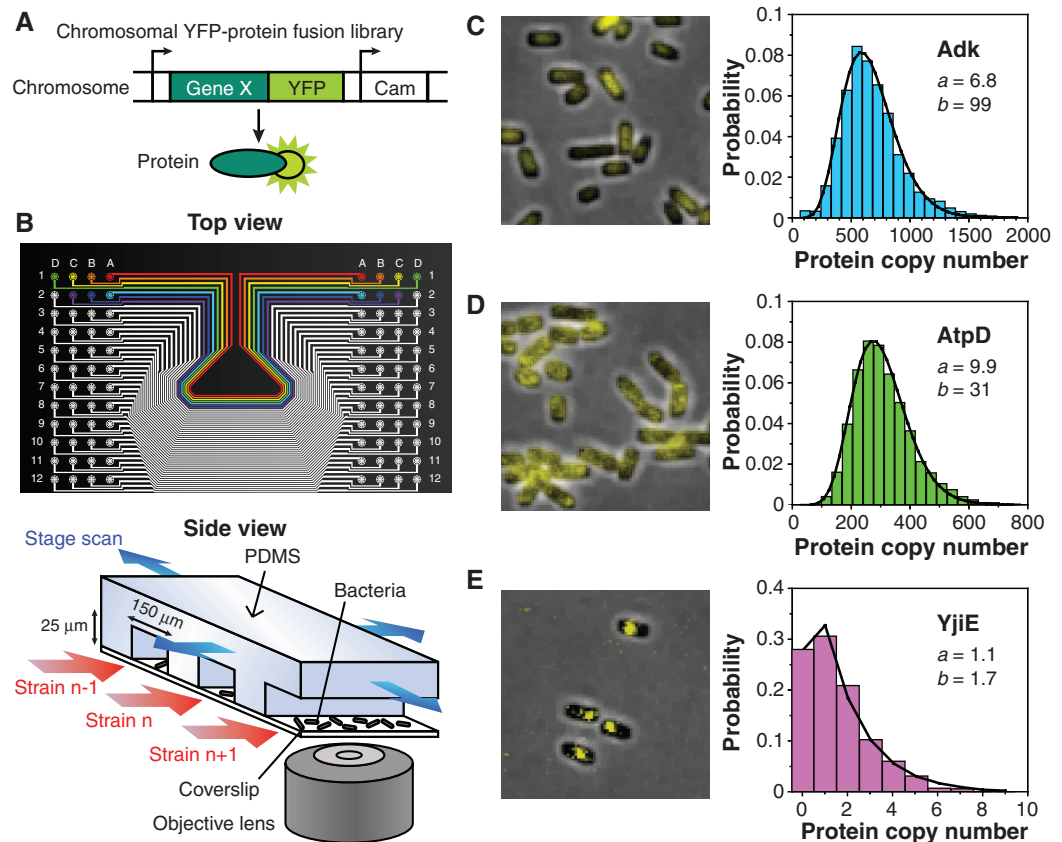
studies did not have single-molecule sensitivity and are not suitable for bacteria, which express mRNA at low copy numbers (8). A fluorescent protein reporter library of *Saccharomyces cerevisiae* (9) has proven to be extremely useful in protein profiling (10, 11). However, the lack of sensitivity in existing flow cytometry or fluorescence microscopy techniques prevented the quantification of one-third of the labeled proteins because of their low copy numbers. In recent years, single-molecule fluorescence microscopy has been used to count mRNA (12–16) or protein (8, 17) molecules in individual cells, especially in bacteria. Yet these methods have only been applied to a limited number of specific genes. Here we report single-cell global profiling of both mRNA and proteins with single-molecule sensitivity using a yellow fluorescent protein (YFP) fusion library for the model organism *Escherichia coli*.

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>2</sup>Department of Physics, Harvard University, Cambridge, MA 02138, USA. <sup>3</sup>Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. <sup>4</sup>Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto M5S 3E1, Canada. <sup>5</sup>Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: xie@chemistry.harvard.edu

**Fig. 1.** Quantitative imaging of a YFP-fusion library. **(A)** Each library strain has a YFP translationally fused to the C terminus of a protein in its native chromosomal position. **(B)** A poly(dimethylsiloxane) (PDMS) microfluidic chip is used for imaging 96 library strains. *E. coli* cells of each strain are injected into separate lanes and immobilized on a polylysine-coated coverslip for automated fluorescence imaging with single-molecule sensitivity. **(C to E)** Representative fluorescence images overlaid on phase-contrast images of three library strains, with respective single-cell-protein level histograms that are fit to gamma distributions with parameters  $a$  and  $b$ . Protein levels are determined by deconvolution (18). The protein copy number per average cell volume, or the concentration, was determined as described in the main text and the SOM (18). **(C)** The cytoplasmic protein Adk uniformly distributed intracellularly. **(D)** The membrane protein AtpD distributed on the cell periphery. **(E)** The predicted DNA-binding protein YjiE with clear intercellular localization. Single YjiE-YFPs can be visualized because they are localized. Note that, unlike **(C)** and **(D)**, the gamma distribution asymmetrically peaks near zero if  $a$  is close to or less than unity.



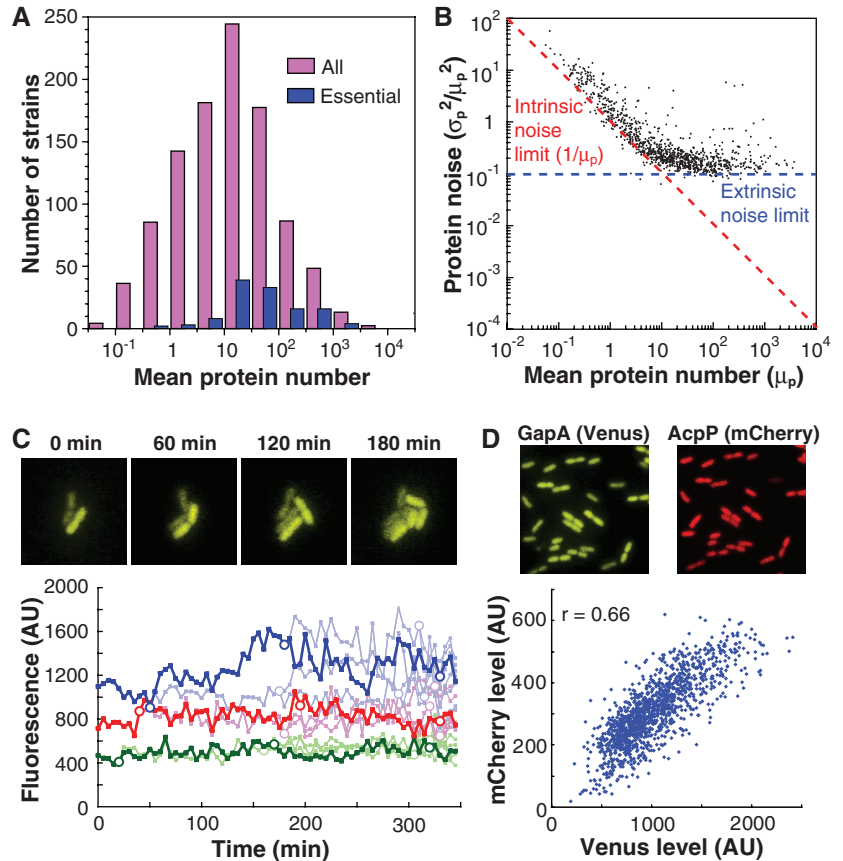
**Single-molecule imaging of a YFP reporter library.** We created a chromosomal YFP fusion library (Fig. 1A), in which each strain has a particular gene tagged with the YFP coding sequence. YFP can be detected with single-molecule sensitivity in live bacterial cells (8, 18). We converted the C-terminal tags of an existing chromosomally affinity-tagged *E. coli* library (19, 20) to *yfp*

translational fusions using  $\lambda$ -RED recombination (21). Out of the 1400 strains attempted, 1018 strains were confirmed by sequencing and showed no significant growth defects. The list of strains is given in table S1 (18).

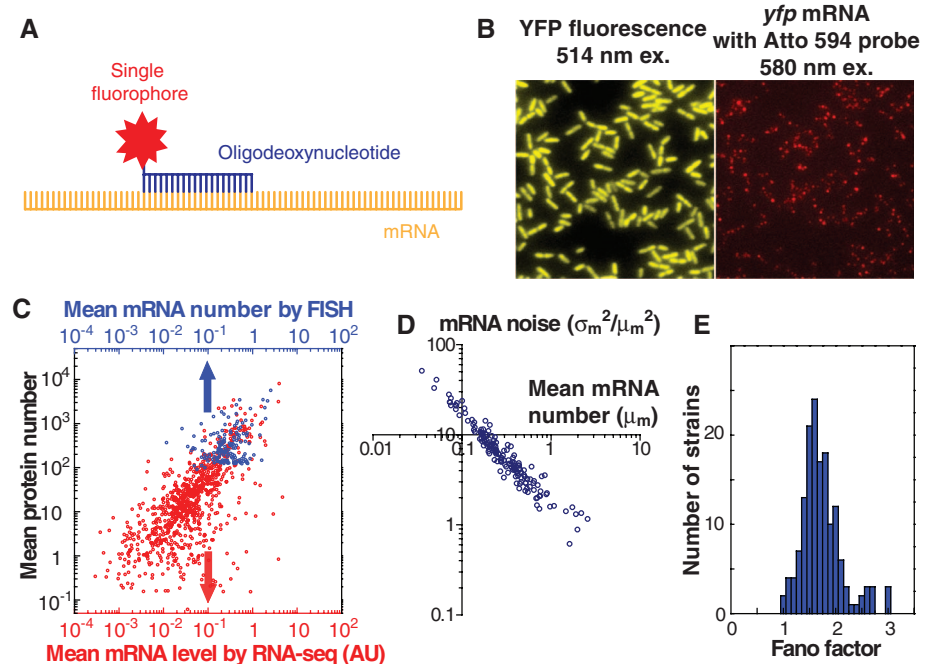
To facilitate high-throughput analyses of the YFP library strains, we implemented an automated imaging platform based on a microfluidic device

(Fig. 1B) (22) that holds 96 independent library strains attached to a polylysine-coated coverslip. Each device was imaged with a single-molecule fluorescence microscope at a rate of  $\sim 4000$  cells in 25 s per strain (Fig. 1C). Single-molecule sensitivity was confirmed by abrupt photobleaching of membrane-bound YFPs expressed at a low level (fig. S14) (8, 18, 23). Automated im-

**Fig. 2.** Profiling of protein abundance and noise. **(A)** Average copy numbers of essential proteins (blue) and all proteins (pink) for 1018 library strains. Almost all essential proteins are expressed at more than an average of 10 copies per cell, although some nonessential proteins have lower abundances. **(B)** Protein expression noise ( $\sigma_p^2/\mu_p^2$ ) versus the mean copy number per cell ( $\mu_p$ ). When  $\mu_p < 10$ , protein expression noise is close to the intrinsic noise limit, which is inversely proportional to the mean (red dashed line). When  $\mu_p > 10$ , noise becomes independent of the mean and is above a plateau of  $\sim 0.1$  (blue dashed lines), which is the extrinsic noise limit. **(C)** Real-time observation of the slow fluctuation of protein levels, at a time scale longer than a cell cycle, originating from extrinsic noise. Each time trace of fluorescence is normalized by cell size and represents a cell lineage of a strain expressing AcpP-YFP. The dark line follows a single lineage; the rest of the descendants have a lighter colored line. Each circle represents a cell division event. The variation among different cells arises primarily from slowly varying extrinsic noise, because the fluctuations within one cell over a cell cycle are comparatively small. **(D)** Two-color measurements of correlations of two different proteins in the same cell. Two highly expressed proteins, GapA and AcpP, are respectively labeled with Venus (YFP) and mCherry (RFP) in the same *E. coli* strain. The protein levels are correlated, with a correlation coefficient of 0.66, which supports the hypothesis that the dependence on global extrinsic factors like ribosomes, rather than gene-specific factors, dominates the extrinsic noise at high expression levels [see (18)].



**Fig. 3.** mRNA profiling of the YFP-fusion library with single-molecule sensitivity in single cells. **(A)** The mRNA of a tagged gene can be detected by FISH against the *yfp* mRNA sequence by using a DNA oligomer probe that is labeled with a single Atto594 fluorophore. **(B)** (Left) Protein and (right) mRNA of the same gene are detected simultaneously in the same fixed cells. **(C)** Mean mRNA number, measured by RNA-seq (red) and by FISH (blue), and mean protein number are correlated. The respective Pearson correlation coefficients ( $r$ ) are 0.54 and 0.77. Each dot is the average of a gene. The FISH data were taken for genes that express  $>100$  copies of proteins per cell, whereas the RNA-seq data include all expressed mRNAs, which are not fused to the *yfp* tag. **(D)** mRNA noise ( $\sigma_m^2/\mu_m^2$ ) scales inversely with mRNA mean number ( $\mu_m$ ) and is higher than expected for Poisson distributions. **(E)** mRNA Fano factors for 137 highly expressed genes. The mRNA Fano factors ( $\sigma_m^2/\mu_m$ ) of the measured strains have similar values centered around 1.6, which indicate non-Poissonian mRNA production or degradation.



age analysis was performed to determine the distribution of single-cell protein abundance normalized by cell size (Fig. 1D) (18). Normalization by cell size is necessary to account for cell size and gene copy number variation due to the cell cycle. We removed the contribution of cellular autofluorescence by deconvolution (fig. S14) (18). The absolute protein level was obtained by calibration with single-molecule fluorescence intensities (fig. S1) (18) to determine the protein concentration

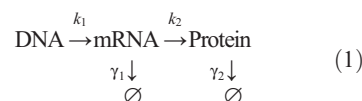
(copy numbers per average cell volume). Independent reporter assays, based on the Miller assay, mass spectrometry, and Western blotting, confirmed that the resulting fluorescence accurately reports on native protein abundance (18) (figs. S4, S21, and S23).

The fluorescence images show the intracellular localization of protein (Fig. 1, C to E). Most cytoplasmic proteins were uniformly distributed intracellularly (Fig. 1C), whereas many membrane-

bound or periplasmic proteins showed localization along the outer contours of the cell (Fig. 1D and table S3). Other proteins, including some DNA-bound proteins and low-copy membrane proteins, showed punctate localization (Fig. 1E).

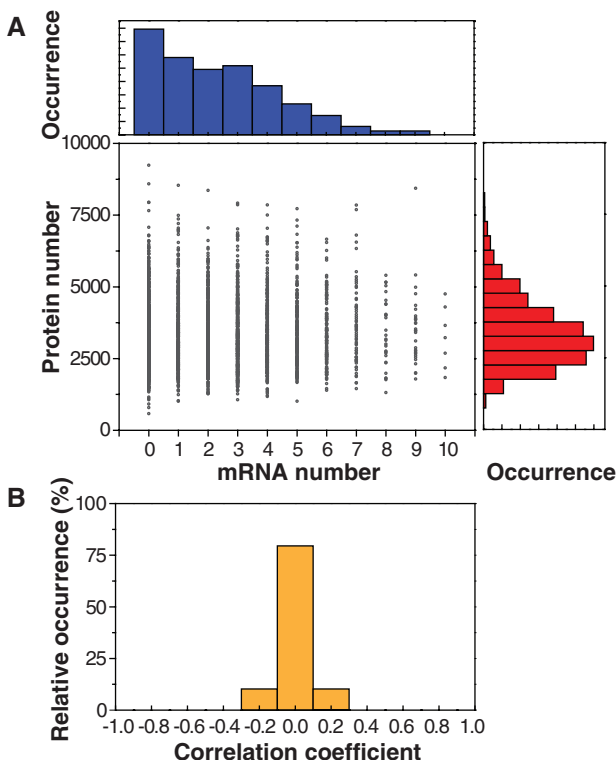
Average protein abundances span five orders of magnitude, ranging from  $10^{-1}$  to  $10^4$  molecules per cell (Fig. 2A). The average protein abundances of essential genes are higher than those for all genes. Of the 121 essential proteins in the library (24), 108 express at 10 or more molecules per cell (Fig. 2A), whereas about half of all the measured proteins are present at fewer than 10 molecules per cell (18). Of the low-expression genes, 60% have been annotated to date (18), and at least 25% were found to have a genetic interaction in a recent double-knockout study (25). The prevalence of proteins with very low copy number suggests that single-molecule experiments are necessary for bacteriology.

**Analysis of protein distributions.** To obtain intrinsic properties of gene expression dynamics, we analyzed the protein expression distributions of different genes. We consider the kinetic scheme

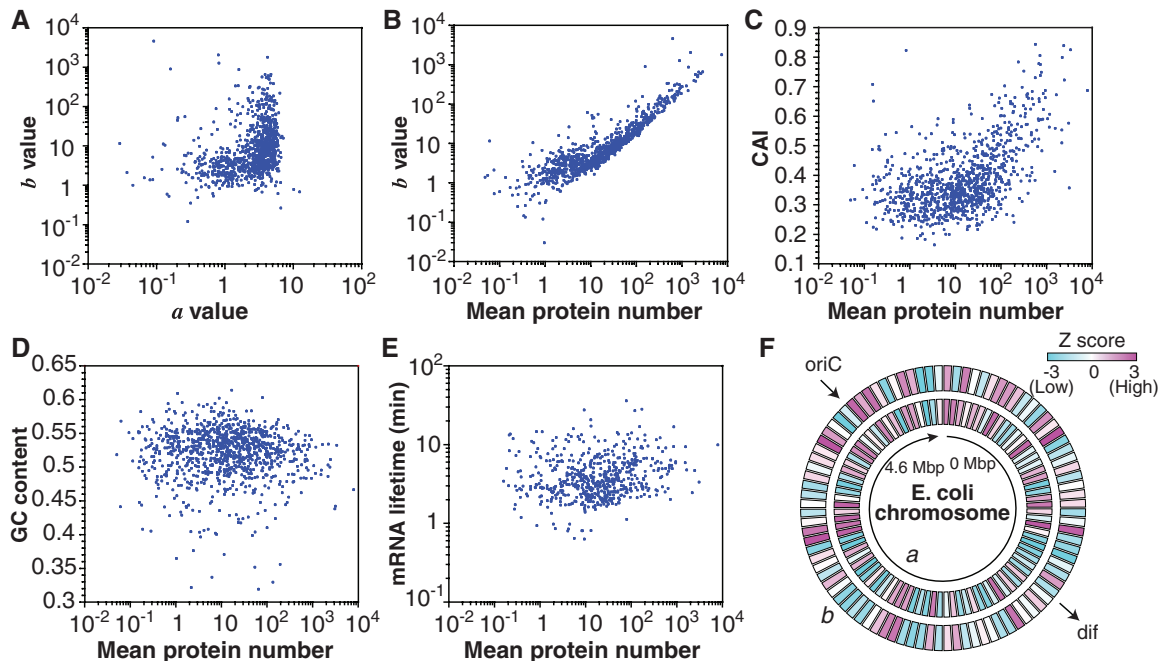


Here,  $k_1$  and  $k_2$  are the transcription and translation rates, respectively.  $\gamma_1$  is the mRNA degradation rate, and  $\gamma_2$  is the protein degradation rate. For stable proteins, including fluorescent protein fusions,  $\gamma_2$  is dominated by the rate of dilution due to cell division and is insensitive to protein lifetime, which could be different for the fusion and native protein. The number of mRNAs produced per cell cycle is given by  $a = k_1/\gamma_2$ ,

**Fig. 4.** No correlation between mRNA and protein levels in a single cell at a particular time. **(A)** (Top) mRNA and (right) protein levels. Protein versus mRNA copy number plot for the TufA-YFP strain, in which TufA is tagged with YFP. Each point represents a single cell of the strain. The correlation coefficient is  $r = 0.01 \pm 0.03$  (mean  $\pm$  SD,  $n = 5447$ ). **(B)** Correlation coefficients from 129 strains with highly expressed labeled genes whose sampling error for the correlation coefficient is  $<0.1$ . The histogram indicates that the lack of correlation between mRNA and protein levels in a single cell is a general phenomenon.



**Fig. 5.** Correlation between expression and gene characteristics. **(A)** Correlation plots of  $a$  and  $b$  ( $r = 0.01$ ) and **(B)** mean protein expression versus  $b$  ( $r = 0.72$ ).  $a$  and  $b$  values are calculated as  $a = \mu_p^2 / \sigma_p^2$  and  $b = \sigma_p^2 / \mu_p$ , respectively, using the mean,  $\mu_p$ , and standard deviation,  $\sigma_p$ , of the protein number histograms. Correlation plots of **(C)** mean protein expression versus CAI ( $r = 0.40$ ), **(D)** GC content ( $r = -0.06$ ), and **(E)** mRNA lifetime ( $r = 0.08$ ). **(F)** Chromosomal dependence of  $a$  and  $b$  values. Z scores of more than 3 (indicated by red) represent a significantly larger value compared with the whole-genome distribution with  $>99.9\%$  confidence; Z scores less than  $-3$  (indicated by blue) represent a significantly smaller value. oriC is the origin of replication, and dif is the resolvase locus.



**Table 1.** Trends in expression levels and protein localization. Table of Z scores of subsets of gene classes characterized by protein and RNA mean, RNA lifetime,  $a$ ,  $b$ , ratio of fluorescence detected on the edge compared with that on the inside of the cell ( $E/I$ ), and the degree of punctate protein localization ( $DP$ ). Leading strand corresponds to transcription in the same direction as the replication fork. PPI indicates protein-protein interactions. Z scores of more than 3 (indicated by red) represent a significantly larger value compared with the whole-genome distribution with >99.9% confidence; Z scores less than -3 (indicated by blue) represent a significantly smaller value.

Category	n	Protein					RNA	
		Mean	a	b	E/I	DP	Mean	Lifetime
All	1018	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Essential gene	121	7.54	8.70	5.34	-2.19	-3.16	5.61	1.38
Nonessential gene	894	-2.74	-3.16	-2.02	0.74	1.30	-1.99	-0.32
Enzyme	410	3.98	4.28	1.85	-5.45	-3.04	3.30	3.15
Translation	17	4.04	3.36	3.30	-0.86	-3.52	3.30	-2.44
Turnover, degradation	13	3.05	3.60	2.08	-0.99	-1.76	2.37	2.14
Transcription factor	98	0.61	1.27	0.71	-2.95	4.29	-0.53	-2.53
Transporter	88	-0.84	0.45	-1.44	7.78	0.96	2.28	3.29
Lagging strand	425	-1.42	-1.95	-0.82	0.74	-0.07	-4.42	-3.63
Leading strand	593	1.12	1.62	0.61	-0.64	0.02	3.96	3.17
Gene length < 500 bp	193	4.06	1.09	3.34	-1.06	-2.98	2.67	-1.65
Gene length >= 500 bp	825	-1.92	-0.55	-1.72	0.57	1.45	-1.26	0.73
Known PPI	603	3.73	3.75	1.37	-3.69	-3.77	1.90	0.58
No known PPI	415	-4.68	-4.65	-1.73	4.28	4.43	-2.25	-0.61

and the protein molecules produced per mRNA is given by  $b = k_2/\gamma_1$ . It was shown theoretically (5, 26) that, under the steady-state condition of Poissonian production of mRNA and an exponentially distributed protein burst size, as previously observed (8, 17), Eq. 1 results in a gamma distribution of protein copy numbers,  $x$ , which is normalized by the average cell volume.

$$p(x) = \frac{x^{a-1} e^{-x/b}}{\Gamma(a)b^a} \quad (2)$$

Here,  $\Gamma$  is a gamma function. The gamma distribution has the property that  $a$  is equal to the inverse of noise ( $\sigma_p^2/\mu_p^2$ ) and  $b$  is equal to the Fano factor ( $\sigma_p^2/\mu_p$ ), where  $\sigma_p^2$  and  $\mu_p$  are the variance and mean of the protein number distributions, respectively. Specific cases have provided experimental support for gamma distribution, but it has not been verified in a system-wide manner (17).

The distributions for 1009 out of the 1018 strains can be well fit by the gamma distribution, Eq. 2 (fig. S20) (18). Consistent with the gamma distribution, the observed distributions are skewed with the peak at zero for low-abundance proteins and have nonzero peaks for high-abundance proteins (Fig. 1, C to E). We note that the bimodal distribution of *lac* permease was observed in *E. coli* under certain inducer concentrations (23, 27). We did not observe clear bimodal distributions among the 1018 strains under our growth conditions, which indicates that bimodal distributions are generally rare.

We note that an alternative mathematical solution to Eq. 1 gives a negative binomial dis-

tribution of protein copy numbers (26). However, the gamma distribution offers a more robust fit of experimental data at low expression levels, because the negative binomial fits are very sensitive to measurement error (18). The two distributions have similar fitting at high expression levels. Other functions, such as log-normal distributions, have been used phenomenologically to fit unimodal distributions (10, 18). However, the gamma distribution fits better than the log-normal distribution for proteins with low expression levels (fig. S20) (18) and fits similarly well for proteins with high expression levels. Most important, the gamma distribution allows extraction of dynamic information from easy measurements of the steady-state distribution at low expression levels. The  $a$  and  $b$  values and the goodness of fits for the 1018 strains are given in table S6.

**Global scaling of intrinsic and extrinsic protein noise.** The protein noise ( $\eta_p^2 \equiv \sigma_p^2/\mu_p^2$ ) exhibits two distinct scaling properties (Fig. 2B). Below 10 molecules per cell,  $\eta_p^2$  is inversely proportional to protein abundance, indicative of intrinsic noise. In contrast, at higher expression levels (>10 molecules per cell), the noise reaches a plateau of  $\sim 0.1$  and does not decrease further, which suggests that each protein has at least 30% variation in its expression level.

For proteins expressed at low levels, simple Poisson production and degradation of mRNA and protein, commonly termed intrinsic noise, are sufficient to account for the observed scaling of  $\sigma_p^2/\mu_p^2 \propto 1/\mu_p$  (Fig. 2B) (10, 11, 28–30). This

scaling property has also been observed for highly expressed yeast proteins (10, 11). We verified Poisson kinetics by monitoring real-time protein production in single cells for several genes whose expression levels were low (table S4) (18); the result agrees with previous work on the repressed *lac* operon (8, 17). The observed noise is always greater or equal to  $1/\mu_p$ , which suggests that specific regulatory methods do not decrease noise substantially below this limit.

For abundant proteins, the  $1/\mu_p$  scaling no longer applies, and a large noise floor overwhelms the intrinsic noise contribution (Fig. 2B). This means that the interpretation of the two parameters  $a = \mu_p^2/\sigma_p^2$  and  $b = \sigma_p^2/\mu_p$  as the burst frequency ( $k_1/\gamma_2$ ) and burst size ( $k_2/\gamma_1$ ) applies well only at low expression levels, whereas the protein distributions at high expression levels are dominated by other factors extrinsic to the above model. We found that the noise floor does not result from cell size effects, nor did it arise from measurement noise (18).

We attribute the additional noise to extrinsic noise (3), that is, the slow variation of the values of  $a$  and  $b$ , which we confirm with real-time observation of protein levels for four randomly selected high-copy library strains. The high-expression noise fluctuates more slowly than the cell cycle (Fig. 2C) (18), so that the rate constants in Eq. 1 can be considered to be heterogeneous among cells.

If we assume that static or slowly varying heterogeneities of  $a$  and  $b$  exist with distributions  $f(a)$  and  $g(b)$ , respectively, the protein distribution is

$$p(x) = \int_0^\infty \int_0^\infty \frac{x^{a-1} e^{-x/b}}{\Gamma(a)b^a} f(a)g(b) da db \quad (3)$$

Even if the normalized variances of  $f(a)$  and  $g(b)$ ,  $\eta_a^2$  and  $\eta_b^2$ , are 0.1, Eq. 3 can still be approximated as a gamma distribution, which explains the generality of the gamma distribution fit of the data (18).

The noise plateau in Fig. 2B can be explained by calculating the expected noise from Eq. 3 (18, 26, 31)

$$\eta_p^2 = \frac{\langle b \rangle + \langle b \rangle \eta_b^2}{\mu_p} + \eta_a^2 + \eta_a^2 \eta_b^2 + \eta_b^2 \quad (4)$$

The extrinsic noise in the last three terms in Eq. 4 might originate from fluctuations in cellular components—such as metabolites, ribosomes, and polymerases (30, 32)—and dominates the noise of high-copy proteins ( $\mu_p \gg 1$ , Eq. 4).

We further demonstrate that the extrinsic noise is global to all high-expression genes by analyzing the correlations between expression levels of 13 pairs of randomly selected genes. Using YFP and red fluorescent protein (RFP) fusions as a pair of reporters (Fig. 2D), we observed statistically significant correlations between the expression levels of all gene pairs,

which confirmed the existence of a global noise factor. The observed correlation is quantitatively predicted by the observed noise floor (18).

**Single-molecule RNA counting.** To examine single-cell mRNA expression, we performed fluorescence in situ hybridization (FISH) with single-molecule sensitivity (33) (Fig. 3A), using a single universal Atto594-labeled 20-oligomer nucleotide probe targeting the *yfp* mRNA in our library. Because the same probe is used for all strains, the optimized hybridization efficiency is unbiased for every measured gene (18). We confirmed the validity of our transcript measurements with RNA-seq (table S6) (18).

We show that the YFP (yellow) and the mRNA (red) of the same gene can be simultaneously detected, and spectrally resolved, within a single fixed cell (Fig. 3B). Because of their low copy numbers, mRNA molecules are sparsely distributed within a cell, independent of YFP locations. By measuring the intensity of each fluorescent spot and counting the number of spots per cell, we determined mRNA copy numbers for individual cells. We used this single-molecule FISH method to quantify mRNA abundance and noise for 137 library strains with high protein expression (>100 proteins per cell).

At the ensemble level, the mean mRNA abundances among these 137 genes range from 0.05 to 5 per cell, and are moderately correlated with the corresponding mean protein expression level at the gene-by-gene basis (correlation coefficient  $r = 0.77$ ) (Fig. 3C). The lack of complete correlation, as reported previously in other organisms, is often attributed to differences in posttranscriptional regulation. Here, with the ability to determine the absolute number of molecules per cell, we determined the ratio between the mean protein abundance and the mean mRNA abundance to range from  $10^2$  to  $10^4$ .

At the single-cell level, the mRNA copy number distributions were broader than the Poisson distributions expected by the random generation and degradation of transcripts with constant rates (18). The mRNA noise scales in inverse proportion to the mean mRNA abundance (Fig. 3D), but mRNA Fano factor values ( $\sigma_m^2/\mu_m$ ), are close to  $\sim 1.6$  (Fig. 3E), rather than unity, as expected for the Poissonian case. We excluded gene dosage effects by gating with the cell size to select the cells that have not yet gone through chromosome replication (18). The non-Poisson mRNA distributions indicate that the rate constant for mRNA generation or degradation fluctuates on a time scale similar to or longer than the typical mRNA degradation time, which has an average of  $\sim 5$  to 10 min for our growth condition (18).

**Simultaneous RNA and protein measurements in single cells.** We now examine the extent to which the mRNA copy numbers and the protein levels are correlated in the same cells. We quantified single-cell mRNA and protein levels simultaneously (Fig. 3B). Figure 4A shows a two-dimensional scatter plot, in which each cell is plotted as a dot with its mRNA and protein

levels on the  $x$  and  $y$  axes, respectively, for the translation elongation factor EF-Tu in the TufA-YFP strain. mRNA and protein copy numbers in a single cell are not correlated ( $r = 0.01 \pm 0.03$ , SEM,  $n = 5447$ ). In fact, among many different highly expressed strains surveyed, the correlation coefficients are all centered on zero (Fig. 4B), which indicates a general lack of mRNA-protein correlation of the same gene within a single cell.

The lack of mRNA-protein correlation can be explained by the difference in mRNA and protein lifetime. In *E. coli*, mRNA is typically degraded within minutes (table S6) (18), whereas most proteins, including fluorescent proteins, have a lifetime longer than the cell cycle (18, 34). As a result, the mRNA copy number at any instant only reflects the recent history of transcription activity (a few minutes), whereas the protein level at the same instant represents the long history of accumulated expression (time scale of a cell cycle). However, additional factors such as extrinsic translational noise are necessary to explain fully the zero mRNA-protein correlation we observe (18). We note that the observed lack of correlation arises because the experiment only measured the copy numbers of protein and mRNA present at the moment of fixation of a single cell. This is not contradictory to the central dogma, which suggests that the mRNA level integrated over a long period of time should correlate with the protein level produced in the same cell, which is consistent with the notable correlation between the mRNA and protein levels averaged for many cells (Fig. 3C, 11). However, our result offers a cautionary note for single-cell transcriptome analysis and argues for the necessity for single-cell proteome analysis.

**Correlation of expression properties with biological factors.** The correlation between the expression parameters and selected gene characteristics is shown in Fig. 5. Small  $a$  values correspond to a narrow range of  $b$  values, and large  $a$  values correspond to a wide range of  $b$  values (Fig. 5A). Highly expressed proteins (mean > 10) had high  $b$  values, whereas low-expression proteins had  $b$  values of about 1 (Fig. 5B). The protein expression levels had a weak correlation with the codon adaptation index (CAI,  $r = 0.42$ ), but had little correlation with GC content ( $r = -0.06$ ) and the mRNA lifetime ( $r = 0.08$ ). The  $a$  and  $b$  values showed moderate dependence on the chromosome position (Fig. 5F). The correlation coefficients and  $Z$  scores between these two and additional parameters are summarized in table S2.

In addition, we characterized the statistical bias of the expression and localization parameters for functional gene categories, as measured by a  $Z$  score in Table 1 and table S3. Some functional categories are strongly correlated with parameters. For example, essential proteins have a strong correlation with high  $a$  ( $Z = 7.5$ ) and high  $b$  ( $Z = 5.3$ ). As expected, membrane transporters showed a high edge/inside ratio ( $Z = 7.3$ ), and

transcriptional repressors indicated high punctate localization ( $Z = 4.1$ ). Proteins with no known protein-protein interactions have significantly reduced expression ( $Z = -4.7$ ). We also found that shorter open reading frames may have higher protein expression levels ( $Z = 4.1$ ). RNA expression tends to be higher for genes transcribed from the leading strand parallel to the movement of the replication fork ( $Z = 4.0$ ). Thus, expression and localization properties can be significantly correlated with functional properties.

**Comparison between *E. coli* and yeast.** Protein abundance and noise have been investigated in yeast with flow cytometry for >2500 high-abundance proteins (10, 11). The single-molecule sensitivity in single bacterial cells allowed us to characterize the full range of protein copy numbers in *E. coli*, which has not been realized in yeast. We found that *E. coli* proteins generally had larger noise and Fano factors than yeast proteins, even for those present at similar copy numbers (fig. S6) (18). A noise plateau due to extrinsic factors is present for both, but the extrinsic noise is larger in *E. coli*.

**Conclusion.** We have provided quantitative analyses of both abundance and noise in the proteome and transcriptome on a single-cell level for Gram-negative bacteria *E. coli*. Given that some proteins and most mRNAs of functional genes are present at low copy numbers in a bacterial cell, the single-molecule sensitivity afforded by our measurements is necessary for understanding stochastic gene expression and regulation. We discovered large fluctuations in low-abundance proteins, as well as a common extrinsic noise in high-abundance proteins. Furthermore, we found that, in a single cell, mRNA and protein levels for the same gene are completely uncorrelated. This result highlights the disconnect between proteome and transcriptome analyses of a single cell, as well as the need for single-cell proteome analysis. Taken together, a quantitative and integral account of a single-cell gene expression profile is emerging.

#### References and Notes

1. C. V. Rao, D. M. Wolf, A. P. Arkin, *Nature* **420**, 231 (2002).
2. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
3. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, *Science* **307**, 1962 (2005).
4. J. M. Pedraza, A. van Oudenaarden, *Science* **307**, 1965 (2005).
5. N. Friedman, L. Cai, X. S. Xie, *Phys. Rev. Lett.* **97**, 168302 (2006).
6. I. Tietjen et al., *Neuron* **38**, 161 (2003).
7. F. Tang et al., *Nat. Methods* **6**, 377 (2009).
8. J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, *Science* **311**, 1600 (2006).
9. W. K. Huh et al., *Nature* **425**, 686 (2003).
10. A. Bar-Even et al., *Nat. Genet.* **38**, 636 (2006).
11. J. R. Newman et al., *Nature* **441**, 840 (2006).
12. J. M. Levsky, S. M. Shenoy, R. C. Pezo, R. H. Singer, *Science* **297**, 836 (2002).
13. A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, *PLoS Biol.* **4**, e309 (2006).
14. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
15. H. Maamar, A. Raj, D. Dubnau, *Science* **317**, 526 (2007).

16. D. Zenklusen, D. R. Larson, R. H. Singer, *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).
17. L. Cai, N. Friedman, X. S. Xie, *Nature* **440**, 358 (2006).
18. Methods and discussion are available as supporting material on Science Online.
19. G. Butland *et al.*, *Nature* **433**, 531 (2005).
20. P. Hu *et al.*, *PLoS Biol.* **7**, e96 (2009).
21. D. Yu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978 (2000).
22. J. C. McDonald *et al.*, *Electrophoresis* **21**, 27 (2000).
23. P. J. Choi, L. Cai, K. Frieda, X. S. Xie, *Science* **322**, 442 (2008).
24. T. Baba *et al.*, *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
25. G. Butland *et al.*, *Nat. Methods* **5**, 789 (2008).
26. J. Paulsson, M. Ehrenberg, *Phys. Rev. Lett.* **84**, 5447 (2000).
27. A. Novick, M. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 553 (1957).
28. J. Paulsson, *Nature* **427**, 415 (2004).
29. M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
30. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
31. V. Shahrezaei, P. S. Swain, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256 (2008).
32. P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
33. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, *Science* **280**, 585 (1998).
34. A. L. Koch, H. R. Levy, *J. Biol. Chem.* **217**, 947 (1955).
35. We thank L. Xun, N. K. Lee, D. Court, C. Zong, R. Roy, and J. Agresti for experimental assistance; and E. Rubin, L. Cai, and J. Elf for helpful discussions. This work was supported by the Gates Foundation (X.S.X.), the NIH Pioneer Director's Award (X.S.X.), and the Canadian

Institutes of Health Research (MOP-77639) (A.E.). Y.T. acknowledges additional support from the Japan Society for the Promotion of Science, the Uehara Memorial Foundation, and the Marubun Research Promotion Foundation, and P.J.C. from the John and Fannie Hertz Foundation.

### Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5991/533/DC1  
Materials and Methods  
SOM Text  
Figs. S1 to S23  
Tables S1 to S6  
References

16 February 2010; accepted 4 June 2010  
10.1126/science.1188308

# Tissue-Engineered Lungs for In Vivo Implantation

Thomas H. Petersen,<sup>1,2</sup> Elizabeth A. Calle,<sup>1</sup> Liping Zhao,<sup>3</sup> Eun Jung Lee,<sup>3</sup> Liqiong Gui,<sup>3</sup> MichaSam B. Raredon,<sup>1</sup> Kseniya Gavrilov,<sup>4</sup> Tai Yi,<sup>5</sup> Zhen W. Zhuang,<sup>6</sup> Christopher Breuer,<sup>5</sup> Erica Herzog,<sup>6</sup> Laura E. Niklason<sup>1,3\*</sup>

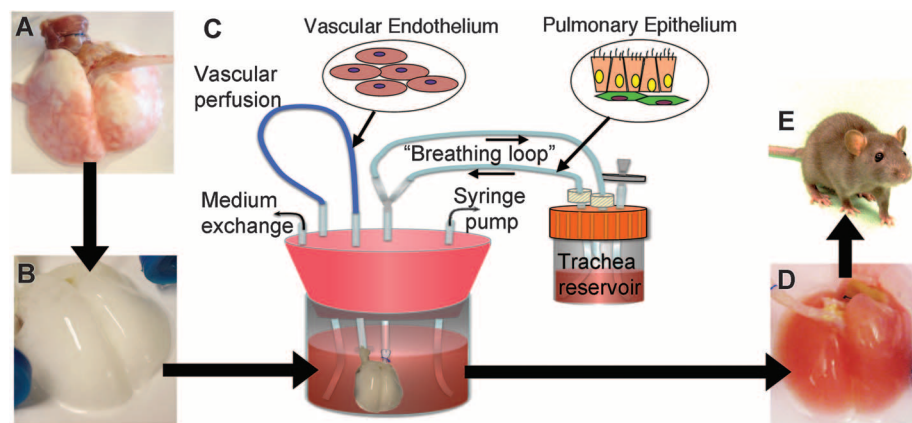
Because adult lung tissue has limited regeneration capacity, lung transplantation is the primary therapy for severely damaged lungs. To explore whether lung tissue can be regenerated in vitro, we treated lungs from adult rats using a procedure that removes cellular components but leaves behind a scaffold of extracellular matrix that retains the hierarchical branching structures of airways and vasculature. We then used a bioreactor to culture pulmonary epithelium and vascular endothelium on the acellular lung matrix. The seeded epithelium displayed remarkable hierarchical organization within the matrix, and the seeded endothelial cells efficiently repopulated the vascular compartment. In vitro, the mechanical characteristics of the engineered lungs were similar to those of native lung tissue, and when implanted into rats in vivo for short time intervals (45 to 120 minutes) the engineered lungs participated in gas exchange. Although representing only an initial step toward the ultimate goal of generating fully functional lungs in vitro, these results suggest that repopulation of lung matrix is a viable strategy for lung regeneration.

Lung diseases account for some 400,000 deaths annually in the United States (1). Human lungs do not generally repair or regenerate beyond the microscopic, cellular level. Currently, the only way to replace lung tissue is to perform lung transplantation, which is an expensive procedure that achieves only 10 to 20% survival at 10 years and one that is hampered by a severe shortage of donor organs (2). Recently, techniques have been developed to quantitatively decellularize complex organs such as heart, liver, and kidney (3–5). Acellular matrices can provide attractive scaffolds for repopulation with lung-specific cells for lung engineering because the extracellular matrix template should contain appropriate three-dimensional (3D) architecture and

regional-specific cues for cellular adhesion. To be functional in vivo, an engineered lung should (i) contain lung-specific cells, (ii) display the

branching geometry of the airways and contain a perfusing microvasculature, (iii) provide barrier function to separate blood from air, and (iv) have mechanical properties that allow ventilation at physiological pressures.

Here, we describe our progress toward the construction of a functional tissue-engineered lung, using rat as a model system. Our approach is summarized in Fig. 1. We first decellularized native lung tissue in order to remove all immunogenic cellular constituents (Fig. 1, A and B). We found that after careful decellularization, the tissue retained its alveolar micro-architecture, its ability to function as a barrier to particulates, and its tissue mechanics. Repopulation of the acellular lung matrix with mixed populations of neonatal lung epithelial cells resulted in regional-specific epithelial seeding in correct anatomic locations. To enhance the survival and differentiation of lung epithelium, we cultured the matrix in a bioreactor designed to mimic certain features of the fetal lung environment, including vascular perfusion and liquid ventilation (Fig. 1, C and D) (6). Lastly, we tested the functionality of the engineered lung tissue by implanting it for short time periods in a syngeneic rat model (Fig. 1E).



**Fig. 1.** Schema for lung tissue engineering. (A) Native adult rat lung is cannulated in the pulmonary artery and trachea for infusion of decellularization solutions. (B) Acellular lung matrix is devoid of cells after 2 to 3 hours of treatment. (C) Acellular matrix is mounted inside a biomimetic bioreactor that allows seeding of vascular endothelium into the pulmonary artery and pulmonary epithelium into the trachea. (D) After 4 to 8 days of culture, the engineered lung is removed from the bioreactor and is suitable for implantation into (E) the syngeneic rat recipient.

<sup>1</sup>Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA. <sup>2</sup>Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA. <sup>3</sup>Department of Anesthesia, Yale University, New Haven, CT 06520, USA. <sup>4</sup>Department of Cellular and Molecular Physiology, Yale University, New Haven, CT 06520, USA. <sup>5</sup>Department of Surgery, Yale University, New Haven, CT 06520, USA. <sup>6</sup>Department of Internal Medicine, Yale University, New Haven, CT 06520, USA.

\*To whom correspondence should be addressed. E-mail: laura.niklason@yale.edu