Supplemental Information

The key parameters that govern translation efficiency

Dan D. Erdmann-Pham[1], Khanh Dao Duc[2] and Yun S. Song[2,3,4,*]

1 Department of Mathematics, University of California, Berkeley, CA 94720, USA
2 Computer Science Division, University of California, Berkeley, CA 94720, USA
3 Department of Statistics, University of California, Berkeley, CA 94720, USA
4 Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

* Lead Contact and Corresponding Author: yss@berkeley.edu

This Supplemental Information contains:

- Supplementary Figures S1–S6
- Supplementary Table S1

Supplementary Figures and Tables

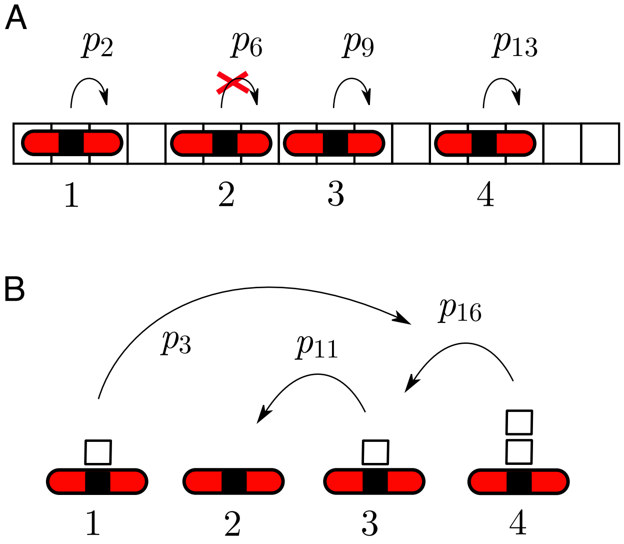

Figure S1. (*Related to Figure 1A,B and* The Hydrodynamic Limit of the Inhomogeneous ℓ-TASEP *of STAR Methods*). **Correspondence between inhomogeneous ℓ-TASEP (A) and the ZRP (B).** ℓ-TASEP particles (rods) correspond to ZRP sites, and holes (empty squares) become ZRP particles.

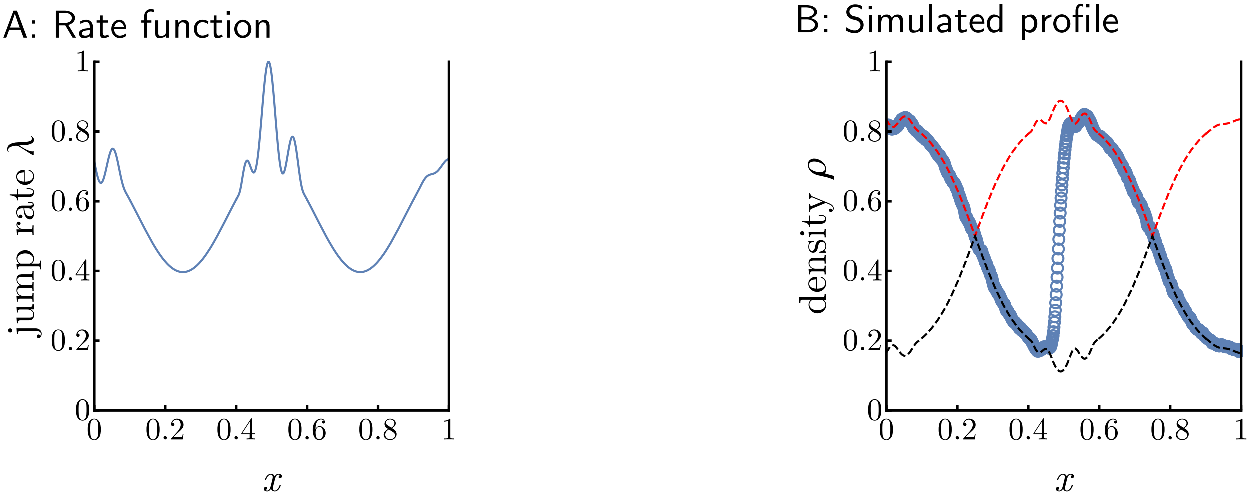

Figure S2. (*Related to* Particle densities, currents and phase transitions *and* Phase Diagram Analysis *of STAR Methods*). **Atypical behaviour of MC branch switching in the presence of two global minima.** Hydrodynamic predictions suggest that branch switching is bound to occur between any two global minima, but do not provide explicit information about the precise location of these singularities. Simulations indicate that branch switching is preferentially situated around local maxima. **A:** Elongation rates. **B:** Circles are averaged counts over 5×10^7 Monte-Carlo steps after 10^7 burn-in cycles on a lattice of size $N = 2000$ with parameters $\alpha = \beta = \ell = 1$ and elongation rate function shown in **A**. We compare these simulated densities to the theoretical profile obtained from the upper (red) and lower (black) branch solutions (described in ()).

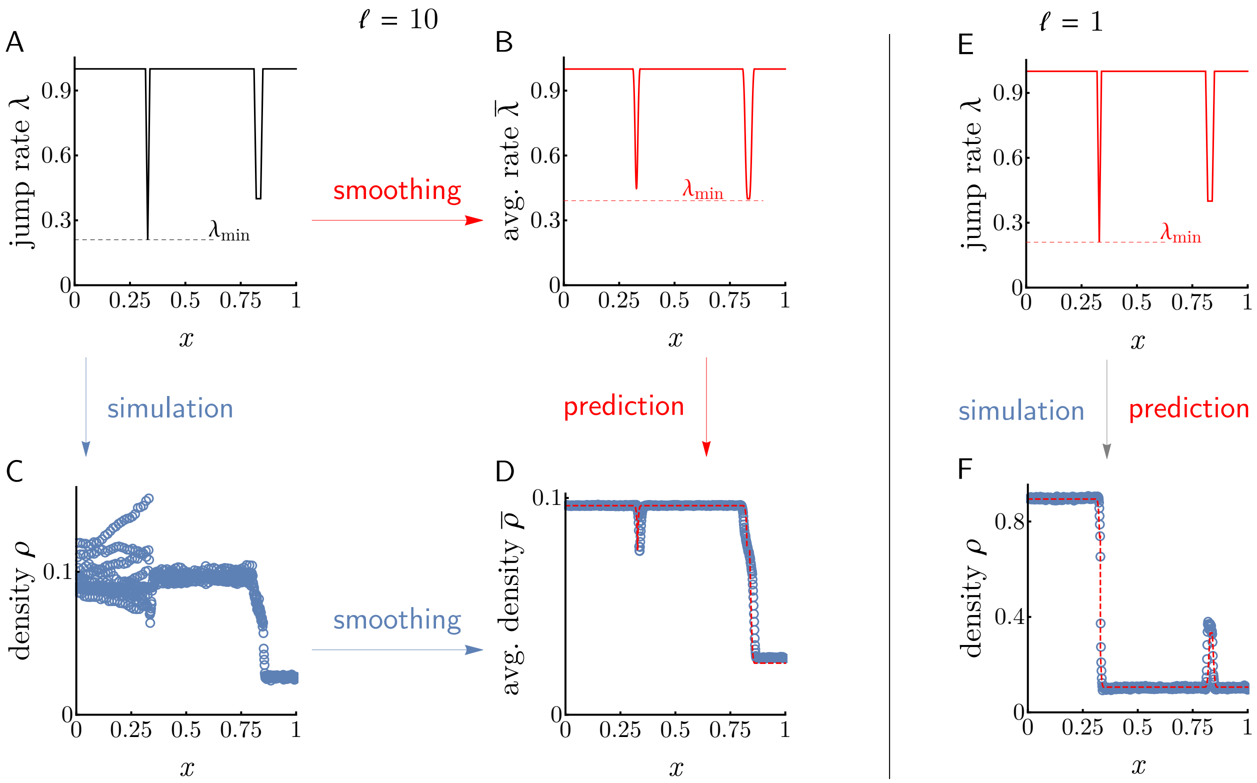

Figure S3. (*Related to Figure 2 and Phase transitions and profiles of STAR Methods*). **MC branch switching is determined by locally averaged elongation rates rather than raw elongation rates (A-D), and the averaging scale depends on the particle size (E,F).** Both the value as well as the location of the minimal elongation rate may differ significantly when measured with respect to the discrete elongation profile (Panel **A**) or smoothed elongation profile (Panel **B**). Panels **C** and **D** demonstrate that our hydrodynamic prediction is very accurate, and show that MC branch switching is governed by the smoothed elongation profile rather than its discrete counterpart. Several of the yeast transcripts we analyzed are affected by this phenomenon, suggesting that a codon's local neighborhood is a stronger determinant of translation dynamics than the absolute elongation rate at that site. Whether smoothed elongation rates (as opposed to unsmoothed rates) describe the translation dynamics more accurately is strongly linked to the particle size (ℓ) and the long-range correlations (in particular, ℓ-periodicity after traffic jams) it induces. To demonstrate this point, we performed the same analysis as in Panels **A-D** using particles of size $\ell = 1$. We found that our hydrodynamic predictions based on the raw, unsmoothed elongation rates (Panel **E**) does indeed provide an accurate approximation of simulated densities (Panel **F**). In short, the fact that the ribosome occupies 10 codons (i.e. the "particle" size is $\ell = 10$) provides another reason (in addition to alleviating the irregularity of elongation rates that cause analytical difficulty) for why smoothing the elongation rates is the right thing to do when applying the hydrodynamic limit of the ℓ-TASEP to study mRNA translation.

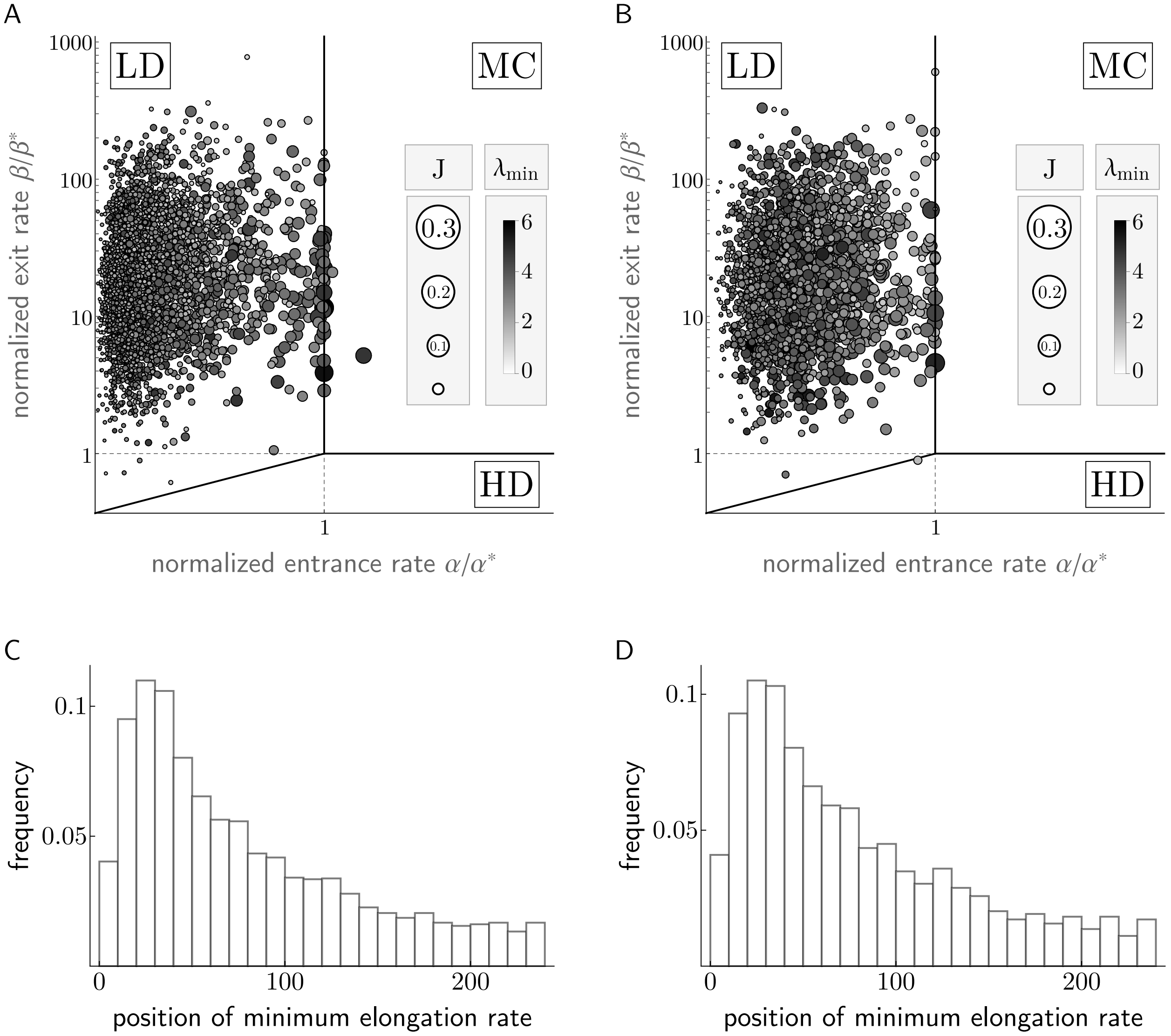

Figure S4. *(Related to* Empirical Study: Translational Efficiency in Yeast *and Figure 4).* **Inferences on the efficiency of yeast's translational system are consistent across datasets.** To test the replicability of our analysis using the previously inferred elongation rates in Dao Duc and Song (2018) and to exclude any possible systematic biases, we repeated our inference on elongation rates obtained by inverting (2) on two independent ribosome profiling datasets: One compiled by Williams et al. (2014) (**A, C**, total of 3098 genes), and one by Pop et al. (2014) (**B, D**, total of 2536 genes). The clear localization of genes within LD and at the LD/MC boundary, together with a characteristic ramp-shaped distribution of the minimum elongation location remain apparent, lending support to our proposed design principles holding true not only on the 850 genes analyzed in the main manuscript, but more generally as a framework governing translation efficiency.

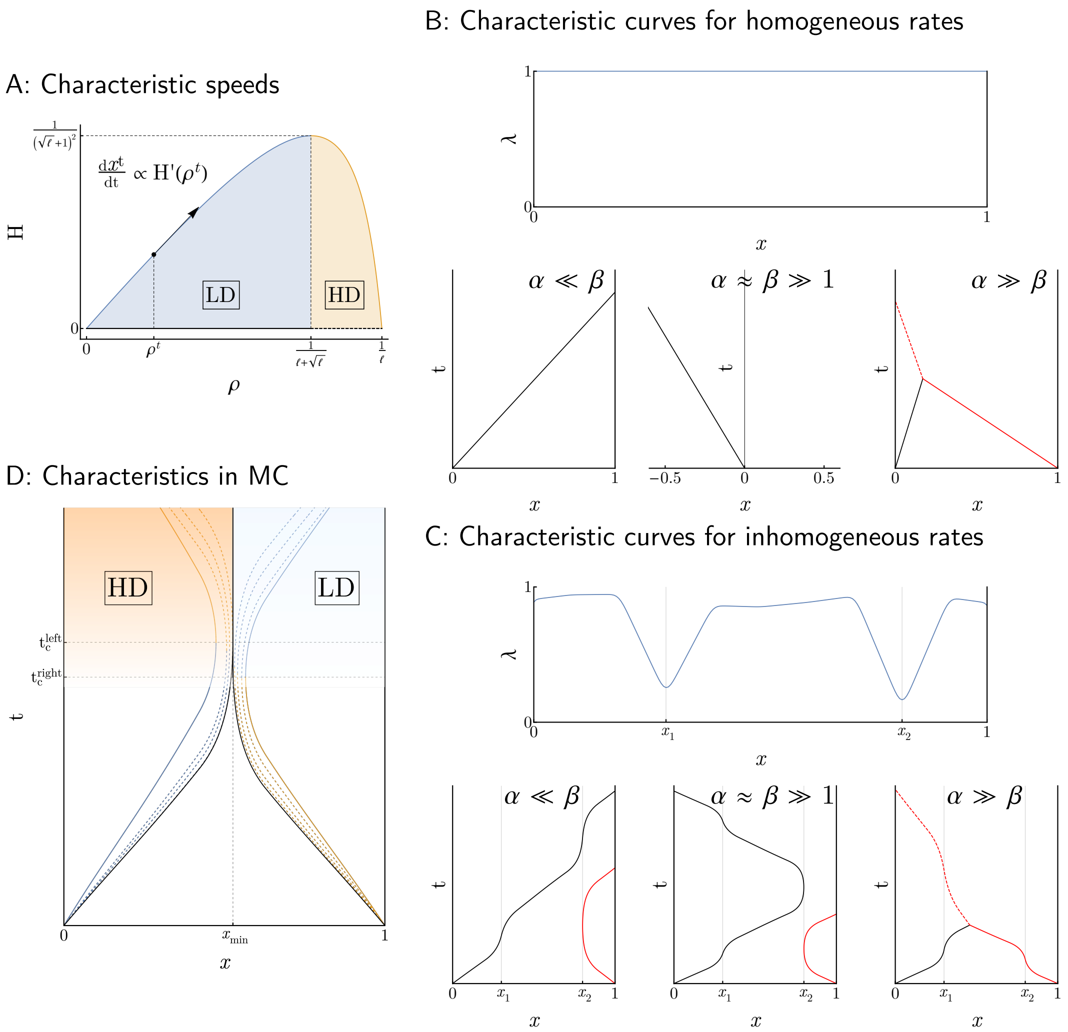

Figure S5. (*Related to Phase transitions and profiles of STAR Methods*). $H(\rho)$ **and its effect on characteristic curves. A:** The rate-normalized flux $H(\rho) = J(\rho, x)/\lambda(x)$ is depicted in solid blue and orange, with characteristic velocity of x^t indicated. If $J(\rho^0, x^0) < J_{\max}$, the characteristic density ρ^t stays within the regions marked LD (blue) or HD (orange), depending on the sign of $\rho^0 - (\ell + \sqrt{\ell})^{-1}$. Otherwise, ρ^t may cross $(\ell + \sqrt{\ell})^{-1}$ forcing the characteristic x^t to return to its origin x^0. **B,C:** Characteristic curves starting at lattice start (black solid curves) and end (red solid curves) for different regions of the phase diagram. Dotted curves represent shock fronts, with colors indicating which characteristic drives the shock. **B:** Homogeneous rates give rise to straight line characteristics with speed $\partial_\rho J(\rho_0)$ and $\partial_\rho J(\rho_1)$, respectively. **C:** Inhomogeneous rates produce more complicated behavior, with characteristics curves slowing down (and potentially reversing direction) near the troughs (x_1 and x_2) of λ. **D:** If $J(\rho^0, x^0) > J_{\max}$ the characteristic curves x^t (left and right colored solid curves) reverse directions at critical times t_c and return to their origin. At t_c, the density ρ^t switches from LD (blue) to HD (orange) (if associated with $x^0 = 0$) or HD to LD (if associated with $x^0 = 1$, cf. **A**). The same happens on all associated rarefaction waves (dashed curves), which interpolate between x^t and the stationary shock of $x^t_{\max}$ (solid black curve).

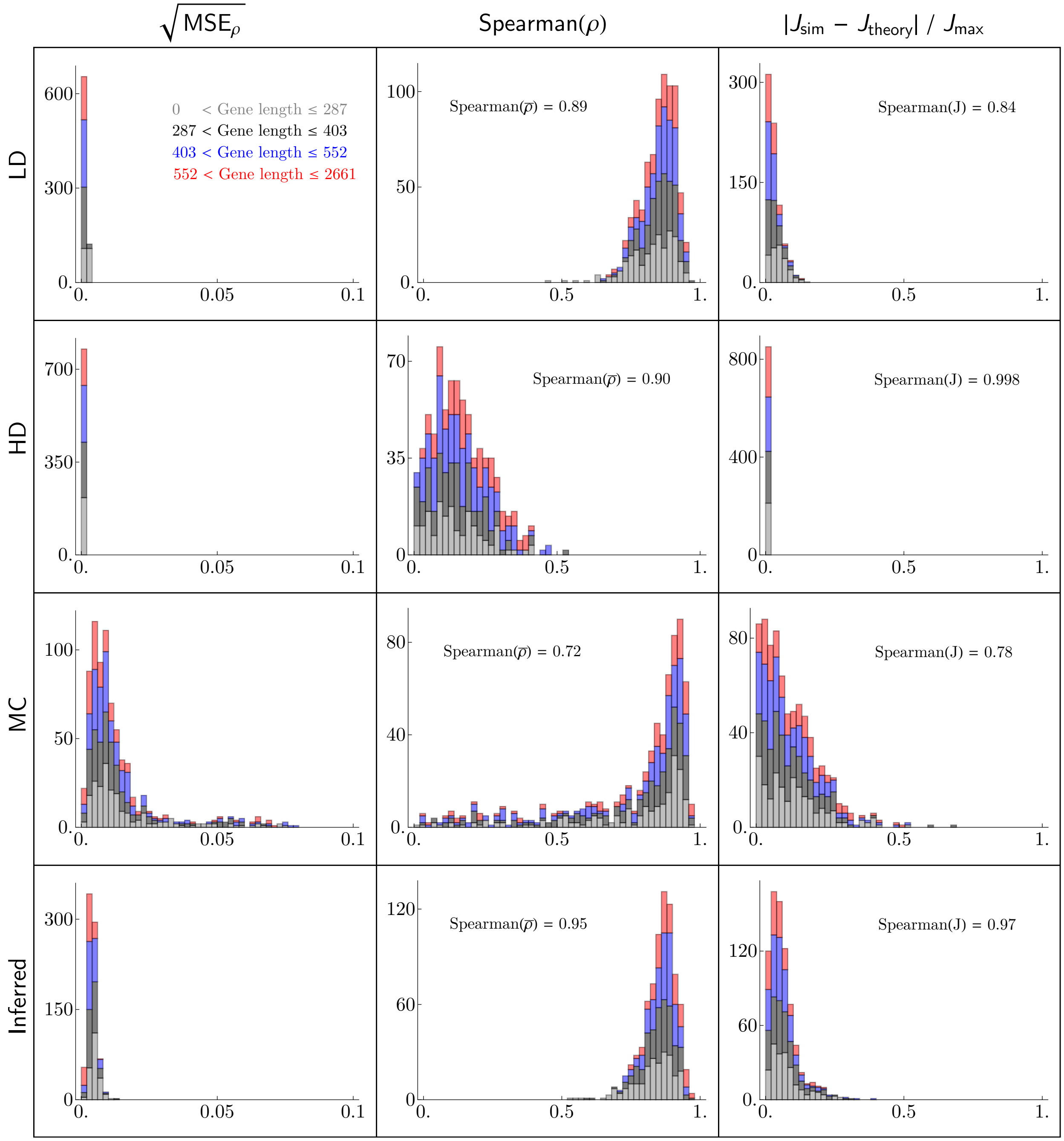

Figure S6. *(Related to* Agreement between Theoretical Prediction and Simulation *of STAR Methods).* **Comparison between simulation and theoretical prediction of our hydrodynamic approximation.** Absolute errors in position-specific ribosome densities ρ (first column) and currents J (third column) are low for all gene lengths (colored), for different regimes in the phase diagram (first three rows), and for biologically relevant initiation and termination rates (last row) inferred in Dao Duc and Song (2018). Moreover, transcript-by-transcript ribosome density profiles ρ and mean ribosome occupancies $\overline{\rho}$ correlate well between simulated data and our hydrodynamic predictions (middle column), as does currents (figure inset in third column).

Table S1. (*Related to* Boundary Conditions *of STAR Methods*). **Boundary conditions by phase.** Expected densities at the left ($x = 0$) and right ($x = 1$) end boundaries of the lattice.

Phase	ρ_0	ρ_1^+	ρ_1^-
LD	$\frac{\alpha}{\lambda_0 + (\ell - 1)\alpha}$	$\frac{1}{\beta}\left[\frac{\alpha(\lambda_0 - \alpha)}{\lambda_0 + (\ell - 1)\alpha}\right]$	$\frac{1}{\lambda_1}\left[\frac{\alpha(\lambda_0 - \alpha)}{\lambda_0 + (\ell - 1)\alpha}\right]$
HD	$\frac{1}{\ell} - \frac{1}{\ell\alpha}\left[\frac{\beta(\lambda_1 - \beta)}{\lambda_1 + (\ell - 1)\beta}\right]$	$\frac{\lambda_1 - \beta}{\lambda_1 + (\ell - 1)\beta}$	$\frac{1}{\lambda_1}\left[\frac{\beta(\lambda_1 - \beta)}{\lambda_1 + (\ell - 1)\beta}\right]$
MC	$\frac{1}{\ell} - \frac{1}{\ell\alpha}\left[\frac{\lambda_{\min}}{(1 + \sqrt{\ell})^2}\right]$	$\frac{1}{\beta}\left[\frac{\lambda_{\min}}{(1 + \sqrt{\ell})^2}\right]$	$\frac{1}{\lambda_1}\left[\frac{\lambda_{\min}}{(1 + \sqrt{\ell})^2}\right]$