Cell Systems, Volume *7*

Supplemental Information

Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes

James C. Taggart and Gene-Wei Li

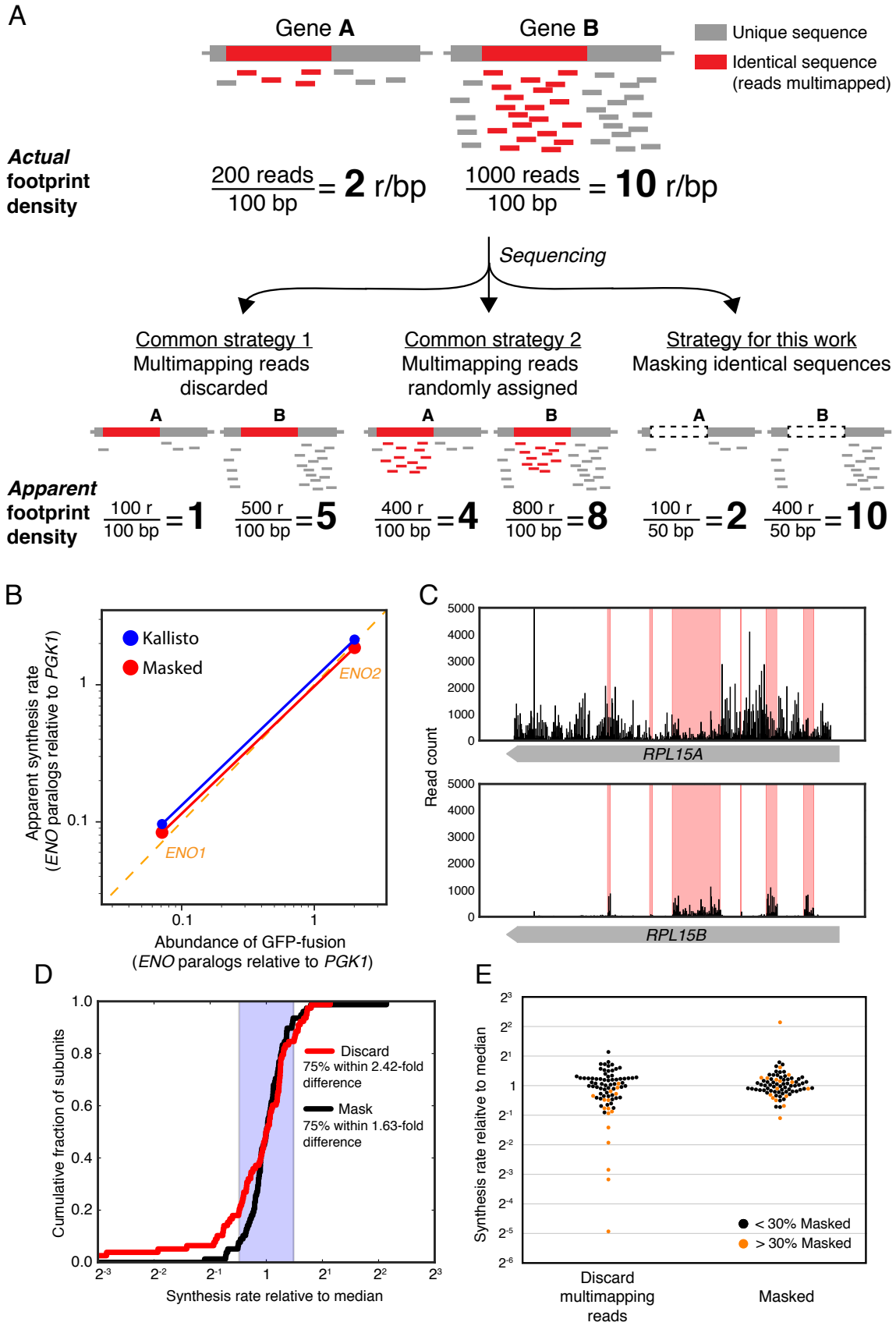
A
Gene A
Gene B
Unique sequence
Identical sequence
(reads multimapped)
Actual
footprint
density
200 reads / 100 bp = 2 r/bp
1000 reads / 100 bp = 10 r/bp
Sequencing
Common strategy 1
Multimapping reads
discarded
Common strategy 2
Multimapping reads
randomly assigned
Strategy for this work
Masking identical sequences
A
B
Apparent
footprint
density
100 r / 100 bp = 1
500 r / 100 bp = 5
400 r / 100 bp = 4
800 r / 100 bp = 8
100 r / 50 bp = 2
400 r / 50 bp = 10
B
Kallisto
Masked
ENO2
ENO1
Apparent synthesis rate
(ENO paralogs relative to PGK1)
Abundance of GFP-fusion
(ENO paralogs relative to PGK1)
C
Read count
RPL15A
RPL15B
D
Discard
75% within 2.42-fold
difference
Mask
75% within 1.63-fold
difference
Cumulative fraction of subunits
Synthesis rate relative to median
E
Synthesis rate relative to median
< 30% Masked
> 30% Masked
Discard
multimapping
reads
Masked

Figure S1. Inadequate read mapping prevents accurate counting of ribosome footprint density. Related to Fig. 1. A. Schematic showing common mistakes in dealing with multimapping reads for calculating ribosome density. Two genes (A and B) share regions of identical sequence (red) and produce different density of ribosome footprints, defined as the number of reads (r) per unit length of a gene (100 bp in this example). After deep sequencing, the origin of short footprints from identical regions (red) cannot be determined. A commonly used strategy discards multimapping reads, which underestimates ribosome footprint density. Another commonly used strategy randomly assigns multimapping reads, which leads to disproportionate density for genes sharing regions of similar sequence. In this work, we used a rational approach to determine the actual ribosome density by masking regions of identical sequence. **B.** Comparison of masking strategy to quantification with Kallisto (Bray et al., 2016). **C.** Example ribosomal protein paralogs in *S. cerevisiae* showing density when multimapping reads are randomly assigned (common strategy 2). Red regions correspond to positions which are removed in masking. Without masking, random assignment prevents quantitation of the synthesis from each paralog; the synthesis from *RPL15B* will be dramatically overestimated and *RPL15A* will be artificially underestimated. **D.** Cumulative distribution of synthesis rates for core subunits of the 80S ribosome using quantification methods that discard (common strategy 1, red) or mask (black) multimapping reads. Synthesis rates are shown relative to the logarithmically transformed median. Shaded region indicates a twofold spread (max to min). **E.** Data shown in D, highlighting proteins for which >30% of positions are masked in orange.

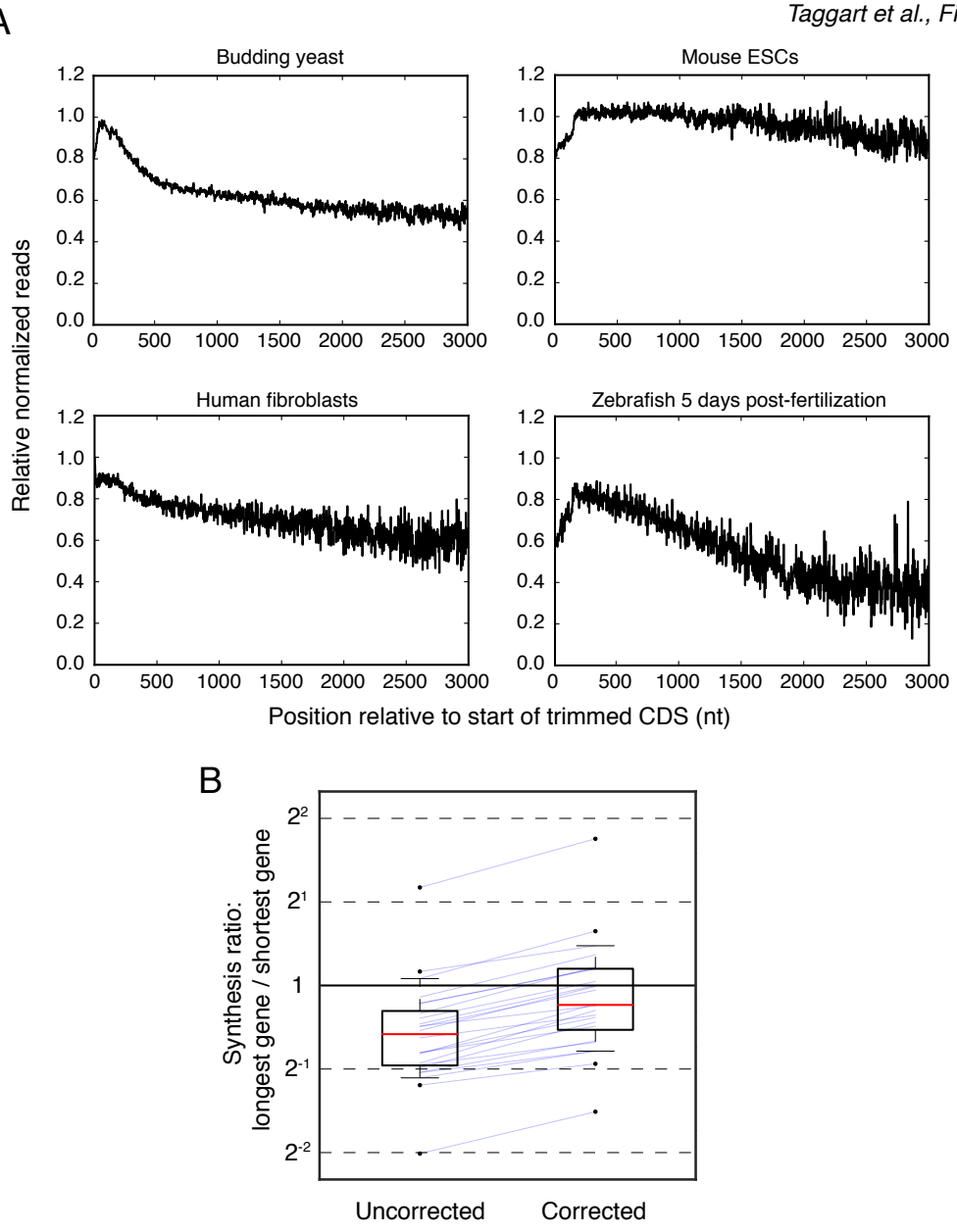

Figure S2. Metagene correction. Related to Fig. 1. A. Metagene analysis for the four species considered in this work. Metagene profiles were generated as described in STAR methods. Datasets correspond to those shown in Fig. 2 and Fig. 4. **B.** Impact of ramp correction on the apparent synthesis of complex subunits of extreme lengths. All complexes with both a subunit shorter than 250 amino acids and one longer than 500 amino acids were considered (n=24). Correction was calculated as described in Synthesis Rate Calculation section of methods.

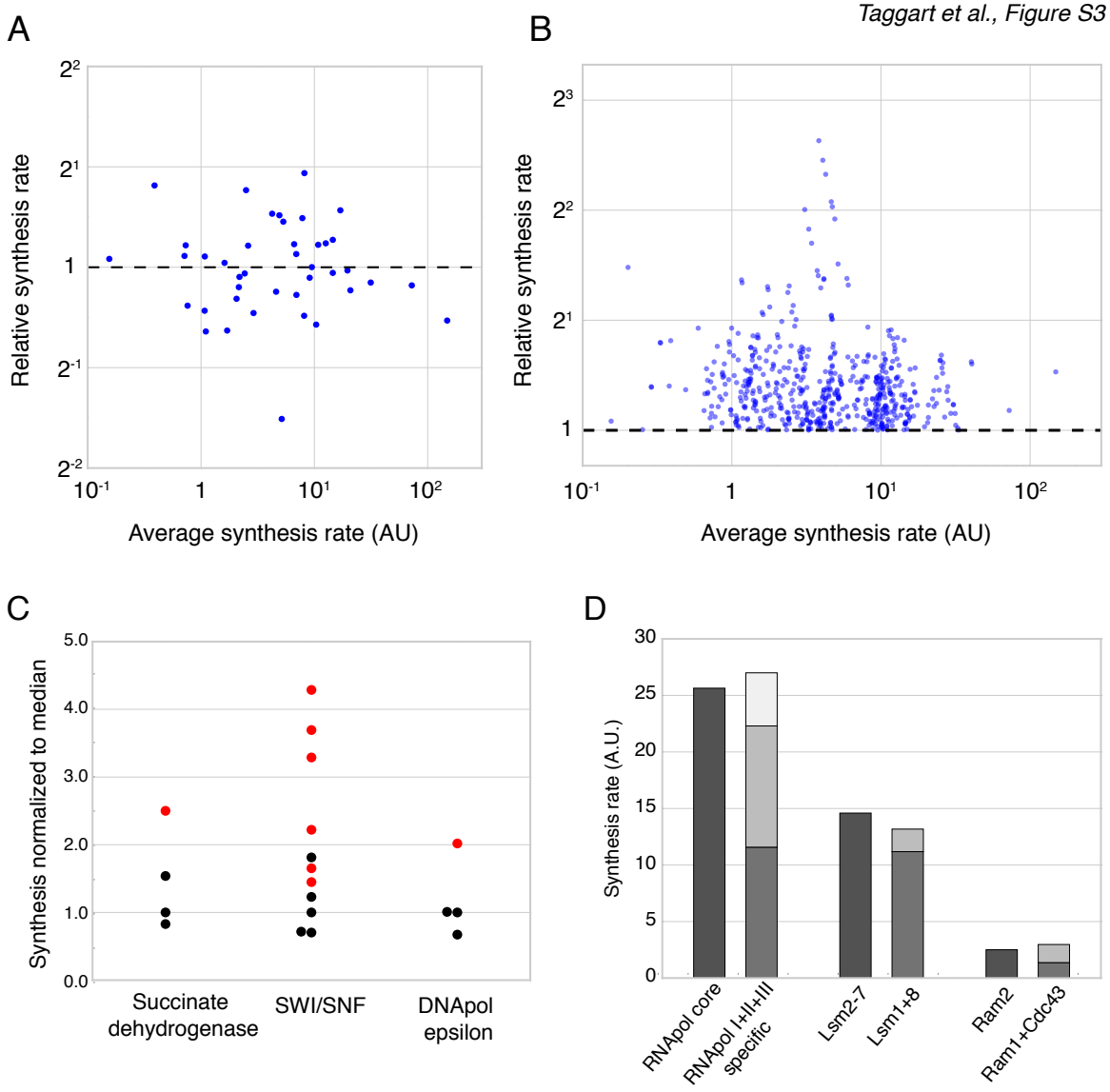

Figure S3. Related to Fig. 2. Influence of absolute expression and moonlighting functions on proportional synthesis. A. Comparison of relative synthesis rates to the geometric mean of synthesis rates for all equimolar 2-subunit complexes. Relative synthesis rates are calculated as Gene B / Gene A, preserving the ordering in Fig. 2A. **B.** Same as A, but including all pair-wise comparisons of subunits within complexes of size 2-10 subunits. Relative synthesis rates calculated as the ratio of the higher to lower synthesis rate. **C.** Synthesis rates for complexes with subunits removed due to secondary function. Synthesis rates are shown relative to the logarithmically transformed median of the included subunits (black points). Subunits excluded from our analyses are shown in red. Subunits with extra-complex functions appear to be

produced super-stoichiometrically. **D.** Synthesis rates for subunits associated with multiple defined complexes. Synthesis rates calculated as the logarithmically transformed median when multiple subunits are grouped and are normalized to the subunit(s) shared between all complexes.

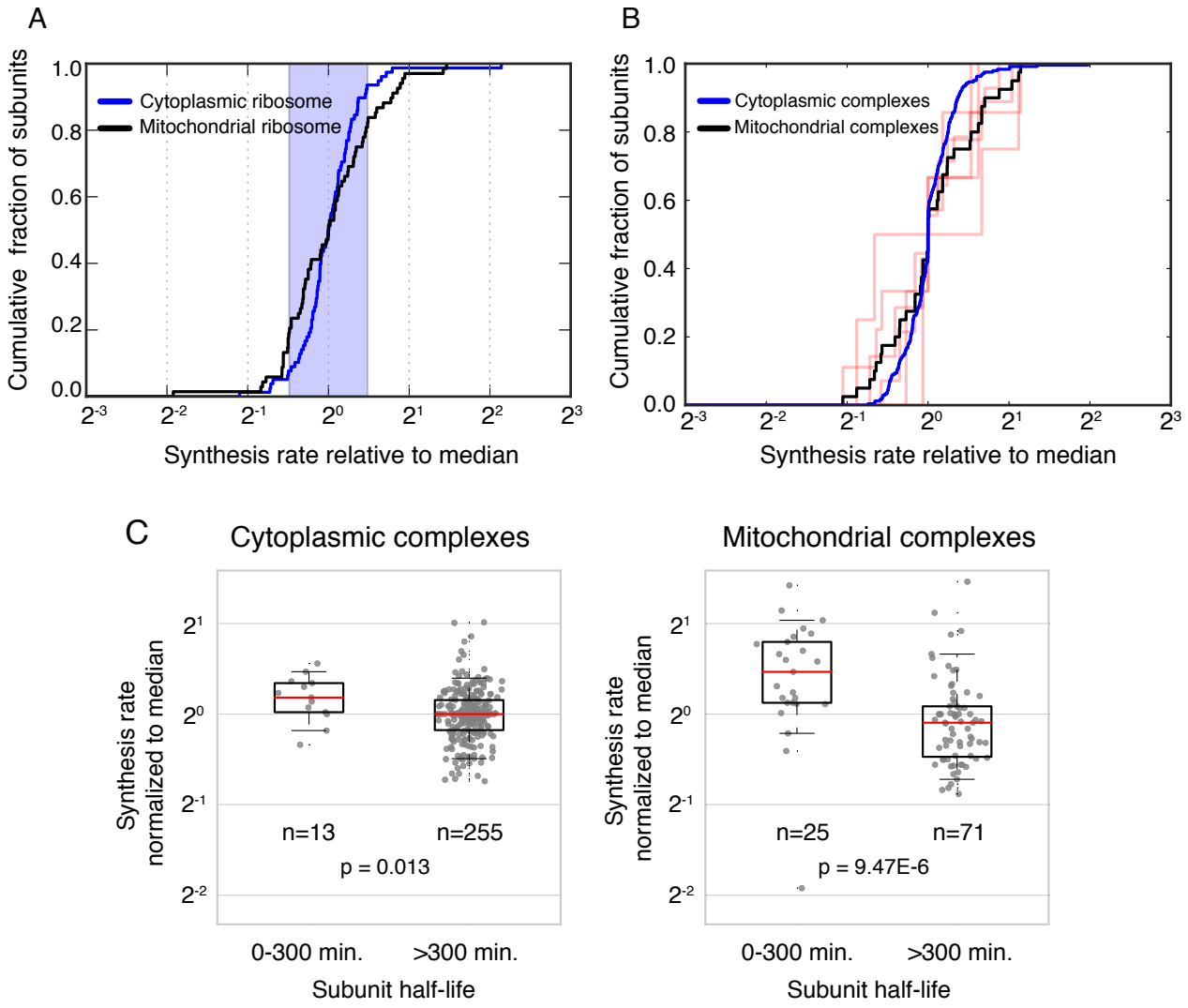

Figure S4. Related to Fig. 2. Mitochondrial complexes adhere less tightly to proportional synthesis. A. Cumulative distribution of synthesis rates for core subunits of the cytoplasmic (blue) and mitochondrial (black) ribosomes. Synthesis rates are shown relative to the logarithmically transformed median. Shaded region indicates a twofold spread (max to min). **B.** Cumulative distribution of synthesis rates for individual medium-sized stable mitochondrial protein complexes (3-10 members). Synthesis rates for subunits present in more than one copy per complex were divided by the stoichiometry, and shown relative to the logarithmically transformed median of each complex. Black line shows the aggregate distribution among all

medium-sized mitochondrial complexes. Blue line shows the aggregate distribution among all medium-sized cytoplasmic complexes, as in Fig. 2C. **C.** Boxplots showing the synthesis rates for subunits of cytoplasmic (left) and mitochondrial (right) complexes with 3 or more members, separated by half-life, as determined by (Christiano et al., 2014). 300 minutes corresponds to roughly two doubling times in the conditions of these half-life measurements. A greater fraction of mitochondrial subunits (25/96, 26%) exceeded this threshold than cytoplasmic subunits (13/268, 5%). As expected, these actively degraded proteins are, on average, synthesized more than their binding partners. Statistics were calculated as Wilcoxon rank-sum test between the half-life bins shown in boxplot.

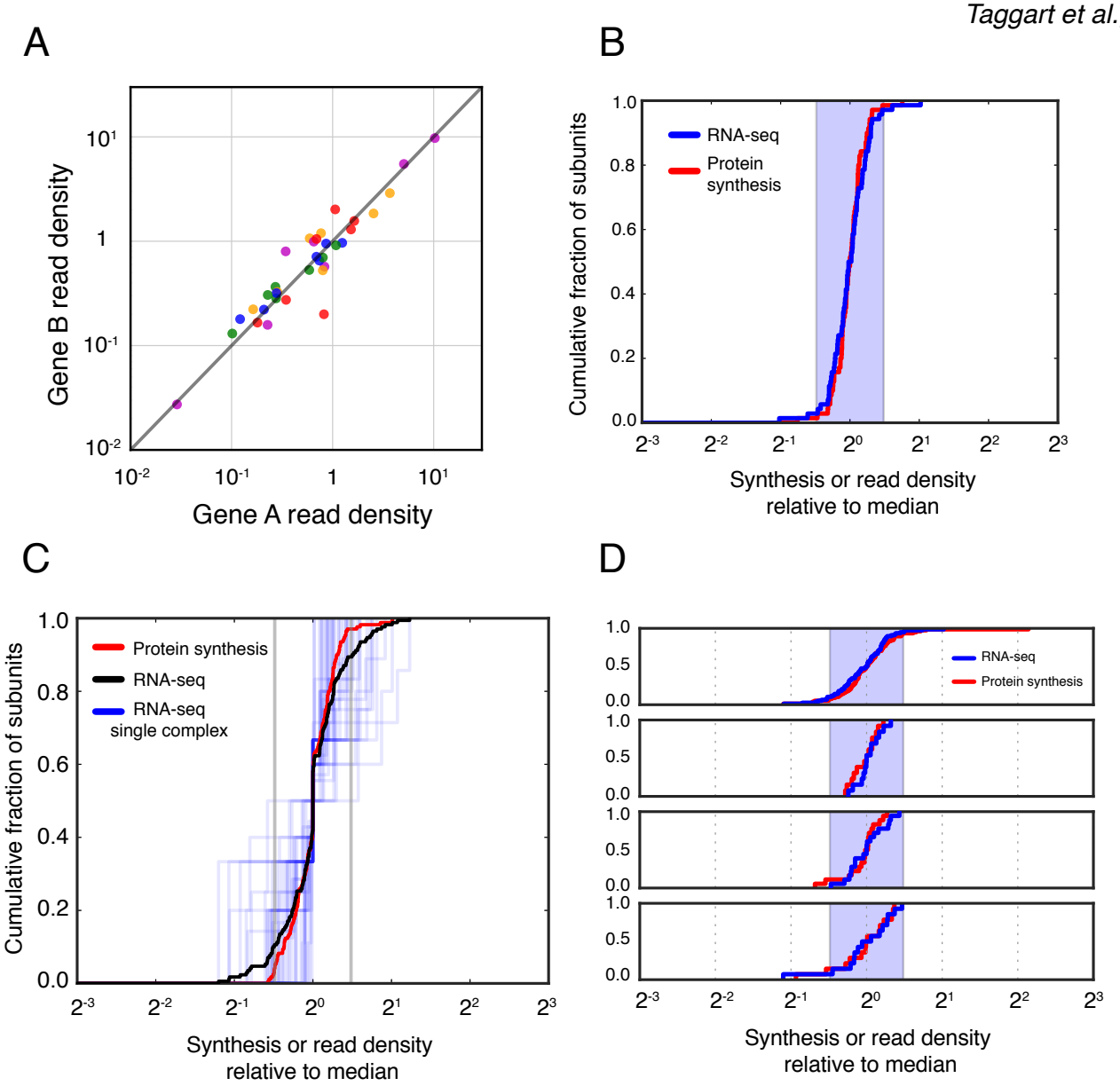

Figure S5. Related to Fig. 2. Transcript abundance for multiprotein complexes. A. RNA-seq read density for subunits of stable complexes composed of two different subunits (encoded by genes 'A' and 'B') with equal stoichiometry. **B.** Cumulative distribution of RNA-seq read density (blue) or synthesis rates (red) for equimolar complexes. Synthesis rates are shown relative to the logarithmically transformed median of each complex. Shaded region indicates a twofold spread (max to min). **C.** Cumulative distribution of RNA-seq read density for individual medium-sized stable protein complexes (3-10 members) (blue). Synthesis rates for subunits

present in more than one copy per complex were divided by the stoichiometry, and shown relative to the logarithmically transformed median of each complex. 170 subunits and 37 complexes are included. Black line shows the aggregate distribution among all medium-sized complexes, and red line shows equivalent aggregate distribution for synthesis rates. **D.** Cumulative distribution of RNA-seq read densities for core subunits of the 80S ribosome, 20S proteasome, 19S proteasome, and V-ATPase, respectively. Values are shown relative to the logarithmically transformed median of each complex. Paralogs of ribosomal proteins were grouped together. Shaded region indicates a twofold spread (max to min).