Supplementary Material

Protein abundances are more conserved than mRNA abundances across diverse taxa

Contents:

mRNA abundance data

Pseudomonas: growth conditions and methods

Strain and growth conditions. *Pseudomonas aeruginosa* UCBPP-PA14 was grown in Synthetic Cystic Fibrosis sputum medium (SCFM), which mimics the nutritional environment of the cystic fibrosis lung and has been previously described [1]. Bacterial growth in liquid media was monitored by measuring the optical density at 600 nm (OD_{600}) during growth at 37℃ with shaking at 250 rpm, and cells were harvested at an OD_{600} of 0.4 to 0.5. We performed two biological replications for each experiment. For proteomics analysis, cells were lysed by French press and prepared as described below.

Microarray experiments. Cultures were mixed 1:1 with RNAlater (Ambion), an RNA stabilizing agent. RNA was isolated using the RNeasy mini kit (Qiagen), and cDNA was prepared for hybridization to Affymetrix PA01 GeneChip microarrays. PCR amplification of the *P. aeruginosa* rplU gene [2, 3] was used to detect DNA contamination using the primers rplU-For (5'-CGCAGTGATTGTTACCGGTG-3') and rplU-Rev (5'-AGGCCTGAATGCCGGTGATC-3'). To assess RNA integrity, samples were subjected to agarose gel electrophoresis. GeneChips were washed, stained, and scanned using an Affymetrix fluidics station at the University of Iowa DNA core facility.

Rice: growth conditions and methods

Strain and growth conditions. Seeds of Kitaake, a rice (*Oryza sativa L*.) cultiver were germinated by imbibing for seven days in water, planting in clay soil, and maintaining in a greenhouse (16 h light/8 h dark). Healthy and well-expanded leaves from 6-week old rice plants were frozen and proteins and mRNA extracted as described below.

Microarray experiments. RNA extraction, clean-up, labeling and hybridization to DNA microarrays were performed using standard protocols. Briefly, total RNA was isolated from leaf tissue using TRIzol (Invitrogen) and RNAeasy (Qiagen). To eliminate traces of DNA, an on-column DNase1 digestion was performed according to manufacturer's protocol (RNase-Free_DNase Set, Qiagen). RNA quality was assessed by RNA Lab-On-A-Chip (Caliper technologies Corp) and Bioanalyzer 2100 (Agilent). Total RNA was labeled and hybridized according to manufacturer's protocols (Affymetrix Genechip 3' IVT Express, Affymetrix), and raw images scanned using a GeneChip Scanner 3000 (Affymetrix).

Microarray analyses

We preprocessed microarray .CEL files with the RMA method [4] by using the *affy* package (Version 1.18.2) [5] in R (Version 2.8.1) with default options (PM probe specific correction, quantile normalization, and expression measure by median polish).

Microarray datasets are deposited in the GEO database (**Table S1**).

Table S1. Data on absolute mRNA concentrations

Organism	Abbreviation	Method	Data source / reference	Comment
Escherichia coli	EC	Affymetrix	[6-8]	*E.coli* grown in minimal medium. Values averaged across datasets.
Pseudomonas aeruginosa	PA	Affymetrix	See above	See above, strain UCBPP-PA14. Deposited as GEO Accession # GSE21966.
Saccharomyces cerevisiae	SC	Affymetrix	[9]	
		RNAseq, Helicos	[10, 11]	
		SAGE	[12]	
Caenorhabditis elegans	CE	Affymetrix	[13]	Data kindly provided by Schrimpf et al.
		SAGE	[13]	
Drosophila melanogaster	DM	Affymetrix	[14]	Data kindly provided by Schrimpf et al.
		SAGE	[15]	
Homo sapiens	HS	Nimblegen single-channel array	[16]	Daoy medulloblastoma cell line.
Oryza sativa	OS	Affymetrix	See above	See above. Deposited as GEO Accession # GSE22788.

Table S2. Reliability of transcriptomic data – Comparison across platforms

mRNA concentrations based on signal intensity measured in single-channel microarrays correlate well with measurements from other methods, as shown here for published yeast data. RNA-seq and single-molecule sequencing (Helicos) [10, 11] are assumed to be the most accurate methods to estimate mRNA concentrations. All correlation coefficients have $p \leq 10^{-16}$. For data sources, see **Table S1**.

Dataset 1	Dataset 2	Spearman correlation (R_s)	Degrees of freedom
Affymetrix	**RNA-Seq**	0.81	1478
Helicos	**Affymetrix**	0.88	1773
Affymetrix	**SAGE**	0.62	1312
SAGE	**RNA-Seq**	0.57	2149
SAGE	**Helicos**	0.58	2489
Helicos	**RNA-Seq**	0.76	4701

Protein abundance data

Sample preparation. Except for *C. elegans* and *D. melanogaster*, we analyzed cellular lysates from each organism using the following general protocol: Cells were broken using glass beads or a homogenizer, and cellular lysate was extracted by 50 min centrifugation at 5,000g. Lysis buffer consisted of 25mM Tris HCL pH 7.5, 5mM DTT, 1.0mM EDTA, 1X CPICPS (Calbiochem protease inhibitor cocktail). Protein concentration was measured and lysate diluted to 2mg/ml with buffer (50mM Tris, pH 8.0). For a typical sample preparation (~10 injections on LTQ-Orbitrap) 50 µl of diluted cell lysate was mixed with 50 µl 100% trifluoroethanol and incubated at 55 °C for 45min (15mM DTT). The sample was cooled to room temperature and incubated with 55mM iodoacetamide in the dark for 30 min. The sample was then diluted to 1ml with buffer (50mM Tris, pH 8.0) and 1:50 w/w Trypsin was added to digest for 4.5hrs. Tryptic digest was halted by adding 20 µl in 1ml formic acid (resulting in 2% v/v). The sample was lyophilized to 20 µl, resuspended in buffer C (95% H_2O, 5% acetonitrile, 0.01% formic acid) and washed using a C18 tip (Thermo). The eluted sample was again lyophilized to 10 µl, resuspended in 120 µl buffer C and filtered through a Microcon-10 filter (for 50 min at 12,000 g). The sample was ready for MS/MS analysis and stored at -80 °C. Rice leaf proteins were extracted as described in [17], removing detergent by OrgoSol Detergent-OUT Kit (G-Biosciences, St. Louis, MO), then processed as above.

LC-MS/MS analysis. Samples were injected 3 to 4 times into an LTQ-Orbitrap Classic (Thermo Electron) mass spectrometer and analyzed in a 0 to 90% acetonitrile gradient over five hours via reverse phase chromatography on a Thermo BioBasic-18 column 100mm x 0.10mm ID. Each of the runs was analyzed independently with Bioworks/SEQUEST (Thermo Electron), searching a database of with the respective amino acid sequences. If splice variants were available, only the longest sequences were used as references. Results of replicate injections were combined for analysis by PeptideProphet [18] and ProteinProphet [19], and post-processed in the APEX pipeline [20, 21] to estimate absolute protein expression based on weighted spectral counts. We accepted proteins as confidently identified if their ProteinProphet probability was above a cutoff corresponding to <5% global FDR.

Raw and post-processed data files as well as detailed descriptions of the MS experiments are provided at http://marcottelab.org/MSdata/.

Table S3. Data on absolute protein concentrations

All protein datasets were obtained from shotgun proteomics experiments.

Organism	Data source / reference	Comment	Total number of proteins with matching mRNA data (% of genome)
Escherichia coli	[16]	Dataset 03 on http://www.marcottelab.org/MSdata	389 (9.1%)
Pseudomonas aeruginosa	See above	Dataset 12 on http://www.marcottelab.org/MSdata	1148 (20.6%)
Saccharomyces cerevisiae	[16]	Dataset 02 on http://www.marcottelab.org/MSdata	array: 1952 (29.9%) RNAseq: 1767 (27.1%)
Caenorhabditis elegans	[22]	Data kindly provided by Schrimpf et al.	array: 8501 (42%) SAGE: 8421 (41.6%)
Drosophila melanogaster	[22]	Data kindly provided by Schrimpf et al.	array: 8393 (59.6%) SAGE: 7646 (54.3)
Homo sapiens	[16]	Dataset 05 on http://www.marcottelab.org/MSdata/	949 (4.4%)
Oryza sativa	See above	Dataset 11 on http://www.marcottelab.org/MSdata	613(1.4%)

Figure S1. Reproducibility of transcriptomic and proteomic data

The mRNA concentrations of orthologs correlate less well with each other than the corresponding protein concentrations. This trend could be due to a decreased technical reproducibility of measurements of mRNA concentrations compared to those of protein concentrations. To test whether this explanation holds true, we examined the technical reproducibility of the present datasets.

Figures S1A and **B** show the correlation of mRNA and protein measurements of technical replicates for the example of *P. aeruginosa* cell lysate, respectively. The data is shown only for genes with both mRNA and protein data (N=568). The Spearman rank correlation coefficients are *0.99* and *0.80* for mRNA (**A**) and protein (**B**), respectively.

This trend holds true for other organisms: reproducibility of mRNA measurements is slightly higher than that of protein measurements. The reproducibility of the single-channel transcriptome data is ranges between an R_s of *0.96* to *0.99* for *Pseudomonas aeruginosa, Oryza sativa, Homo sapiens* (*not shown*). The reproducibility of proteomics data is also generally high, with a squared Pearson correlation coefficient of R^2*>0.85* of log-transformed data [20].

This observation implies that lack of reproducibility of mRNA measurements does not account for the lower correlation amongst mRNA concentration than amongst protein concentrations (across organisms).

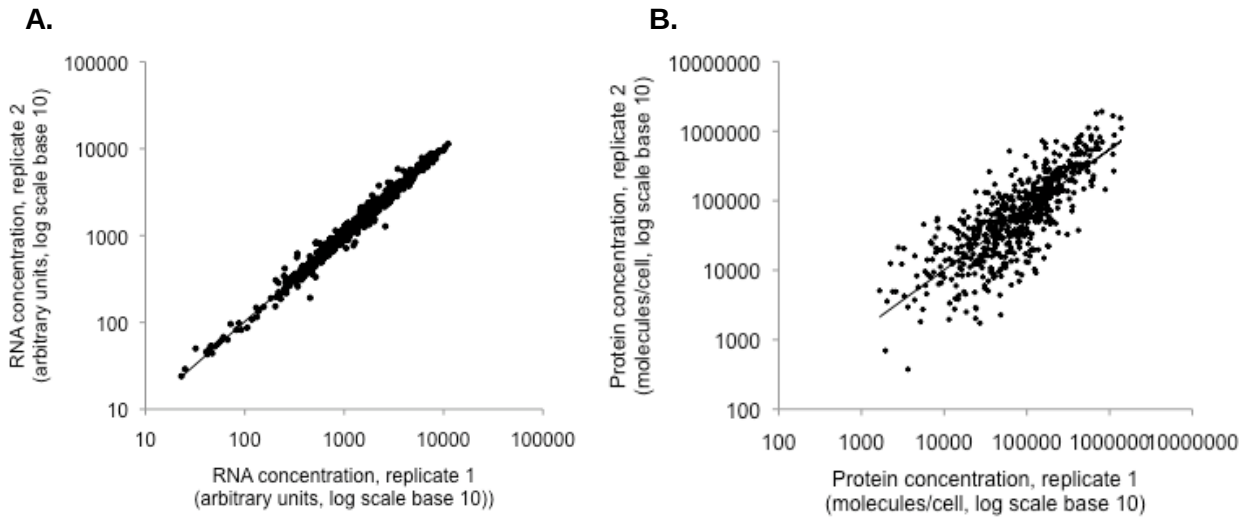
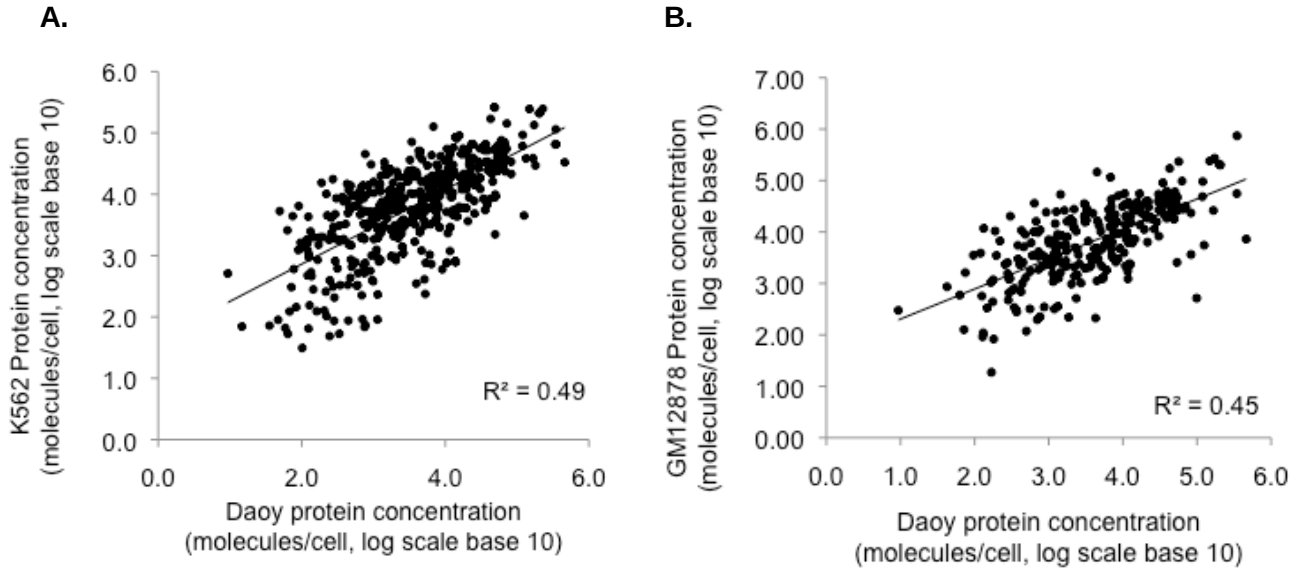


Figure S2. Comparison of the human Daoy medulloblastoma cell line to other human cell lines

To test for the ability of the Daoy medulloblastoma cell line to represent human protein expression, we compare the Daoy protein concentrations to those from two other human cell lines (*unpublished*): **A**) to the mylogenous leukemia cell line K562 (N=476); **B**) to the lymphoblastoid cell line GM12878 (N=257). Both cell lines are part of the set selected for the Encode project. In both cases, protein concentrations correlate significantly (P-value<0.001) between the two cell lines. While immortalized cell lines can only partially represent gene and protein expression in real human tissues, the extent of correlation between the different cell lines suggests that the Daoy data is a representative snapshot of general expression regulation in humans.



Data analysis

Ortholog extraction.

(i) We created datasets for use in this analysis by first extracting all genes for which both protein and mRNA abundance data was available. For each pair of organisms, we calculated orthologs using the online tool or a local build of InParanoid v. 4.0 [23]. For each group of orthologs and the respective in-paralogs, we extracted the ortholog pair with the highest bootstrap value (usually 100%), removing extra in-paralog matches and resulting in just a single protein pair for any given orthologous relationship. In-paralogs are considered not true orthologs as they arose through gene duplication after speciation, and thus after formation of the true orthologs.

We selected only orthologs for whom protein and mRNA data was available in each of the two organisms under study. Results are presented in the main paper and **Table S4** and **S7**.

(ii) As an alternative method, we used OrthoMCL [24] to estimate orthologs. OrthoMCL produces groups of putative orthologs (ortho-groups) relating to an ortho-group in another organism. An ortho-group can have one or several members. If several paralogs were present in an ortho-group, we used data for all genes, duplicating, if necessary, the data for the respective single ortholog in the other organism. This method corresponds to an averaging of the expression values across all members in an ortho-group. The method contrasts the first approach using InParanoid in which we only counted the single best match between two genes. It includes the data for the in-paralogs.

The results for the OrthoMCL analysis are presented in **Table S5**. Both methods of ortholog estimation result in similar biases in the correlation coefficients.

Correlation analysis.

Correlation analyses were performed using the online R interface at http://www.wessa.net [25]. Spearman's rank correlation coefficients were employed to nullify the effects of large outliers for the main correlation analyses. We first carried out the correlation analyses using for the mRNA concentrations average values across the array and non-array based methods (**Table S4A**). To account for possible variation in mRNA concentrations due to differences in the methods, we also calculated correlations using mRNA data derived from array-based methods, only (main text). Third, we calculated correlations using mRNA data from non-array based methods only, *i.e*. SAGE or RNAseq. These datasets were available for yeast, *C. elegans*, *D. melanogaster* (**Table S4B**).

For the **length correction** analysis, we performed a Pearson product moment partial correlation using the ranks for either protein or RNA given the ranks of the average ortholog length, in base pairs (**Table S6**).

Table S4. Correlations using entire (mixed sources) mRNA and protein abundance data.

Yeast transcript abundance was also available as an average of ≥2 of 3 measurements, one using Affymetrix microarrays, one using two-channel microarrays (taken from [20]), and one using RNAseq [10]. *C. elegans* and *D. melanogaster* transcript abundances for this analysis were an average for each gene between a SAGE measurement and an Affymetrix measurement [13-15]. Only those genes where both abundances were available were used. Orthologs were calculated using the InParanoid method. Abbreviations see **Table S1.** Protein abundances correlations are significantly higher than mRNA abundance correlations in analysis A ($p < 0.0001$, paired Wilcoxon test). Protein abundance correlations are larger than mRNA abundance correlations in all three available cases in analysis B.

In **Tables S4 to S7**, reported Wilcoxon tests consider all observations regardless of correlation *p*-value. All Wilcoxon *p*-values are also significant ($p < 0.05$) when considering only the subset of observations in which either the protein or mRNA correlation has $p < 0.001$, as well as the case in which both correlations have $p < 0.001$.

A. Mixed Source mRNA

	Mixed mRNA				
Comparison	**Protein**	**mRNA**	**Orthologs**	**p-value protein**	**p-value mRNA**
EC-PA	0.66	0.57	248	<0.0001	<0.0001
EC-SC	0.42	0.25	106	<0.0001	0.0098
EC-CE	0.40	-0.01	110	<0.0001	0.9174
EC-DM	0.34	0.02	123	<0.0001	0.8262
EC-HS	0.58	0.11	26	0.0019	0.5927
EC-OS	0.15	-0.01	48	0.3089	0.9462
PA-SC	0.30	0.33	164	<0.0001	<0.0001
PA-CE	0.39	0.11	203	<0.0001	0.1182
PA-DM	0.26	0.14	236	<0.0001	0.0316
PA-HS	0.42	0.29	42	0.0056	0.0625
PA-OS	-0.08	-0.17	59	0.547	0.198
SC-CE	0.61	0.28	707	<0.0001	<0.0001
SC-DM	0.57	0.24	813	<0.0001	<0.0001
SC-HS	0.54	0.20	143	<0.0001	0.0166
SC-OS	0.40	0.37	72	0.0005	0.0014
CE-DM	0.78	0.46	2536	<0.0001	<0.0001
CE-HS	0.59	0.42	279	<0.0001	<0.0001
CE-OS	0.27	0.05	93	0.0089	0.6341
DM-HS	0.67	0.47	325	<0.0001	<0.0001
DM-OS	0.09	0.14	101	0.3708	0.1626
HS-OS	0.35	0.16	30	0.058	0.3983

B. SAGE/Seq mRNA

	SAGE/Seq				
Comparison	**Protein**	**mRNA**	**Orthologs**	**p-value protein**	**p-value mRNA**
SC-CE	0.60	0.36	700	<0.0001	<0.0001
SC-DM	0.58	0.37	774	<0.0001	<0.0001
CE-DM	0.77	0.22	2680	<0.0001	<0.0001

Table S5. Correlations with mRNA abundance from array-data only

Correlations for organism pairs, calculated using an alternative method for ortholog estimation (OrthoMCL [24]). Abbreviations see **Table S1.** Protein abundances correlations are significantly higher than mRNA abundance correlations (p = 0.025, paired Wilcoxon test.)

OrthoMCL, Array only					
Comparison	**Protein**	**mRNA**	**Orthologs**	**p-value protein**	**p-value mRNA**
EC-PA	0.66	0.59	222	<0.0001	<0.0001
EC-SC	0.26	0.17	83	0.0176	0.1244
EC-CE	0.30	0.09	91	0.0039	0.3962
EC-DM	0.10	0.11	89	0.3511	0.3048
EC-HS	0.65	0.66	12	0.0221	0.0195
EC-OS	-0.04	0.02	22	0.8597	0.9296
PA-SC	0.28	0.27	198	<0.0001	0.0001
PA-CE	0.35	0.36	297	<0.0001	<0.0001
PA-DM	-0.04	0.08	301	0.4893	0.1662
PA-HS	0.50	0.35	39	0.0012	0.0289
PA-OS	-0.10	-0.15	60	0.4471	0.2526
SC-CE	0.61	0.54	1032	<0.0001	<0.0001
SC-DM	0.55	0.50	1081	<0.0001	<0.0001
SC-HS	0.60	0.26	150	<0.0001	0.0013
SC-OS	0.21	0.23	98	0.0379	0.0227
CE-DM	0.49	0.29	3480	<0.0001	<0.0001
CE-HS	0.60	0.41	283	<0.0001	<0.0001
CE-OS	0.05	0.17	130	0.5721	0.0532
DM-HS	0.60	0.41	281	<0.0001	<0.0001
DM-OS	0.14	0.20	141	0.0978	0.0174
HS-OS	0.26	0.18	41	0.1007	0.2601

Table S6. Correlations after correction for sequence length.

Correlations after length correction using the Pearson product moment partial correlation of the ranks of protein or transcript abundance, given ranks of average ortholog length. Orthologs were calculated using the InParanoid method. Abbreviations see **Table S1.** Protein abundance correlations are significantly higher than mRNA abundance correlations (p = 0.018, paired Wilcoxon test) in analysis A. Protein abundance correlations are larger than mRNA abundance correlations in all three available cases in analysis B.

A.Array data

Length corrected, Array Only					
Comparison	**Protein | length**	**mRNA | length**	**Orthologs**	**p-value protein**	**p-value mRNA**
EC-PA	0.59	0.46	248	<0.0001	<0.0001
EC-SC	0.40	0.24	111	<0.0001	0.0112
EC-CE	0.31	0.21	124	0.0005	0.0192
EC-DM	0.24	0.37	129	0.0062	<0.0001
EC-HS	0.58	0.42	26	0.0005	0.0260
EC-OS	0.03	-0.01	48	0.8396	0.9700
PA-SC	0.28	0.26	172	0.0002	0.0006
PA-CE	0.39	0.30	223	<0.0001	<0.0001
PA-DM	0.23	0.25	241	0.0003	<0.0001
PA-HS	0.41	0.34	42	0.0045	0.0276
PA-OS	-0.19	-0.14	59	0.1495	0.2903
SC-CE	0.52	0.49	803	<0.0001	<0.0001
SC-DM	0.47	0.39	872	<0.0001	<0.0001
SC-HS	0.44	0.22	146	<0.0001	0.0061
SC-OS	0.23	0.43	77	0.0442	<0.0001
CE-DM	0.73	0.45	2922	<0.0001	<0.0001
CE-HS	0.45	0.37	301	<0.0001	<0.0001
CE-OS	0.19	0.06	109	0.0049	0.5100
DM-HS	0.55	0.45	326	<0.0001	<0.0001
DM-OS	-0.01	0.15	106	0.8775	0.1200
HS-OS	0.32	0.16	30	0.0770	0.4100

B. SAGE/RNA-seq

SAGE/Seq, length corrected					
Comparison	**Protein | length**	**mRNA | length**	**Orthologs**	**p-value protein**	**p-value mRNA**
SC-CE	0.50	0.30	700	<0.0001	<0.0001
SC-DM	0.46	0.29	774	<0.0001	<0.0001
CE-DM	0.73	0.22	2680	<0.0001	<0.0001

Table S7. Data from main text (Figure 1B).

Abundance correlation values using the Spearman rank correlation, presented with respective p-values. Orthologs were calculated using the InParanoid method. Abbreviations see **Table S1.** Data in this table are the same as that presented in Figures 1B and 2A.

		Array only			
Comparison	**Protein**	**mRNA**	**Orthologs**	**p-value protein**	**p-value mRNA**
EC-PA	0.66	0.57	248	<0.0001	<0.0001
EC-SC	0.44	0.29	111	<0.0001	0.0020
EC-CE	0.41	0.26	124	<0.0001	0.0035
EC-DM	0.32	0.43	129	<0.0001	<0.0001
EC-HS	0.58	0.11	26	0.0019	0.5927
EC-OS	0.15	-0.01	48	0.3089	0.9462
PA-SC	0.31	0.28	172	<0.0001	0.0002
PA-CE	0.41	0.31	223	<0.0001	<0.0001
PA-DM	0.25	0.28	241	<0.0001	<0.0001
PA-HS	0.42	0.29	42	0.0056	0.0625
PA-OS	-0.08	-0.17	59	0.547	0.1980
SC-CE	0.63	0.56	803	<0.0001	<0.0001
SC-DM	0.58	0.50	872	<0.0001	<0.0001
SC-HS	0.56	0.21	146	<0.0001	0.0110
SC-OS	0.40	0.41	77	0.0003	0.0002
CE-DM	0.77	0.50	2922	<0.0001	<0.0001
CE-HS	0.60	0.41	301	<0.0001	<0.0001
CE-OS	0.27	0.08	109	0.0045	0.4083
DM-HS	0.67	0.49	326	<0.0001	<0.0001
DM-OS	0.11	0.18	106	0.2616	0.0648
HS-OS	0.35	0.16	30	0.0580	0.3983

References

[1] Palmer, K. L., Aye, L. M., Whiteley, M., Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. *J Bacteriol* 2007, *189*, 8079-8087.

[2] Kuchma, S. L., Connolly, J. P., O'Toole, G. A., A three-component regulatory system regulates biofilm maturation and type III secretion in Pseudomonas aeruginosa. *J Bacteriol* 2005, *187*, 1441-1454.

[3] Yarwood, J. M., Volper, E. M., Greenberg, E. P., Delays in Pseudomonas aeruginosa quorum-controlled gene expression are conditional. *Proc Natl Acad Sci U S A* 2005, *102*, 9008-9013.

[4] Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M.*, et al.*, Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003, *31*, e15.

[5] Gautier, L., Cope, L., Bolstad, B. M., Irizarry, R. A., affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004, *20*, 307-315.

[6] Allen, T. E., Herrgard, M. J., Liu, M., Qiu, Y.*, et al.*, Genome-scale analysis of the uses of the Escherichia coli genome: model-driven analysis of heterogeneous data sets. *J Bacteriol* 2003, *185*, 6392-6399.

[7] Corbin, R. W., Paliy, O., Yang, F., Shabanowitz, J.*, et al.*, Toward a protein profile of Escherichia coli: comparison to its transcription profile. *Proc Natl Acad Sci U S A* 2003, *100*, 9232-9237.

[8] Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., Palsson, B. O., Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004, *429*, 92-96.

[9] Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I.*, et al.*, Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 1998, *95*, 717-728.

[10] Nagalakshmi, U., Wang, Z., Waern, K., Shou, C.*, et al.*, The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008, *320*, 1344-1349.

[11] Lipson, D., Raz, T., Kieu, A., Jones, D. R.*, et al.*, Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* 2009, *27*, 652-658.

[12] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., Serial analysis of gene expression. *Science* 1995, *270*, 484-487.

[13] McKay, S. J., Johnsen, R., Khattra, J., Asano, J.*, et al.*, Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. *Cold Spring Harb Symp Quant Biol* 2003, *68*, 159-169.

[14] Chintapalli, V. R., Wang, J., Dow, J. A., Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet* 2007, *39*, 715-720.

[15] Lee, S., Bao, J., Zhou, G., Shapiro, J.*, et al.*, Detecting novel low-abundant transcripts in Drosophila. *Rna* 2005, *11*, 939-946.

[16] Ramakrishnan, S. R., Vogel, C., Prince, J. T., Li, Z.*, et al.*, Integrating shotgun proteomics and mRNA expression data to improve protein identification. *Bioinformatics* 2009, *25*, 1397-1403.

[17] Park, C. J., Peng, Y., Chen, X., Dardick, C.*, et al.*, Rice XB15, a protein phosphatase 2C, negatively regulates cell death and XA21-mediated innate immunity. *PLoS Biol* 2008, *6*, e231.

[18] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002, *74*, 5383-5392.

[19] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003, *75*, 4646-4658.

[20] Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E. M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 2007, *25*, 117-124.

[21] Vogel, C., Marcotte, E. M., Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. *Nat Protoc* 2008, *3*, 1444-1451.

[22] Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H.*, et al.*, Comparative Functional Analysis of the Caenorhabditis elegans and Drosophila melanogaster Proteomes. *PLoS Biol* 2009, *7*, e48.

[23] Remm, M., Storm, C. E., Sonnhammer, E. L., Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 2001, *314*, 1041-1052.

[24] Chen, F., Mackey, A. J., Stoeckert, C. J., Jr., Roos, D. S., OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 2006, *34*, D363-368.

[25] Wessa, P., Office for Research Development and Education 2010.