Electronic Supplementary Material (ESI) for Molecular BioSystems.

The relative importance of transcription and translation – next-generation analysis of gene expression regulation

Joel McManus[1], Zhe Cheng[2], Christine Vogel[2]

1 Carnegie Mellon University, Department of Biological Sciences, Pittsburgh, PA, USA

2 New York University, Department of Biology and Center for Genomics and Systems Biology, New York NY, USA

Corresponding authors: cvogel@nyu.edu, mcmanus@andrew.cmu.edu

Supplementary Information

Table S1. Large-scale methods to characterize stages of gene expression

Recently developed methods are highlighted as examples for approaches to system-wide analyses. Ref. [1] also provides a comprehensive overview of methods to study the different dimensions of gene expression analysis.

Measurement	Methods	Comment
Concentrations		
RNA concentrations	RNA-seq, microarrays	RNA-sequencing provides transcript abundance data in form of 'reads' for the entire genome. Typical datasets comprise millions of reads at nucleotide-level resolution. RNA abundances are estimated based on normalized read counts, e.g. by RPKM or FPKM. Several hundreds of datasets are publically available in respective databases (e.g. GEO).
Protein concentrations	Mass spectrometry	Liquid chromatography coupled to tandem mass spectrometry collects tens of thousands of spectra for a fraction (or sometimes the complete) proteome. Spectra are mapped to peptide sequences using vector multiplication algorithms [31]; precursor ion intensities inform on peptide abundance, which in turn provide estimates of protein concentrations. If estimates of copy numbers per cell are desired, observed abundances need to be scaled to internal or spike-in reference concentrations. Label-free quantification is now widely accepted; relative concentration or concentration changes are estimated by intensity ratios from experiments involving Stable Isotopic Labeling of Amino acids in Cell culture (SILAC)[2].
	Fluorescence-based imaging	GFP-tagged proteins are imaged in high-throughput and protein concentrations estimated from fluorescence intensity, e.g. [3, 4].
	Antibody-based immunohistochemistry	With thousands of antibodies available, immunohistochemistry based imaging reveals presence and subcellular localization of proteins across mammalian cells [5]. However, the method is still only semi-quantitative.
Processes		
Transcription rates	GRO-seq	Global run-on sequencing (GRO-seq) measures the genome-wide location, amount, and orientation of elongating RNA polymerase II. Purified and washed nuclei are incubated with 5 bromo UTP (BrU) and limiting amounts of CTP. Elongating RNAP II incorporates BrU into nascent transcripts, which are purified by immunoprecipiation with anti-BrdU beads. The subsequent RNA molecules are cloned and subjected to high-throughput sequencing.
	NET-seq	Native elongating transcript sequencing (NET-seq) identifies the density of RNA Polymerase II genome-wide by immunoprecipitating FLAG-

		tagged RNAP II and sequencing the 3' ends of nascent transcripts [6].
mRNA degradation rates	cDTA BRIC-seq RATEseq	Many studies have measured global turnover rates by shutting off transcription with drugs or temperature-sensitive alleles of RNAP II subunits – posing additional stress on the cells. More recent work has used pulse-chase labeling of transcripts with 4-thiouracil [7], cDTA [8], and RATE-seq [9] or 5 bromo uridine (BRIC-seq)[10]. During the chase period, mRNA are sampled at multiple time points, immunoprecipitated, and quantified with high-throughput sequencing to measure degradation rates.
Translation efficiency	Polysome profiling	Polysome profiling uses a sucrose gradient to separate actively translating mRNAs according to their density determined by the number of ribosomes bound [11]. Unbound (ribosome-free) mRNAs and monosomal mRNAs are thought to be less actively translated than polysomal mRNAs. While the method is simple and robust [12], the approach is labor-intensive: several fractions from the gradient have to be analyzed to achieve a sensitive measurement of translation efficiency. In addition, polysome profiling cannot provide information on ribosome positioning within the transcript; it only estimates density (number of ribosomes per mRNA) and occupancy (number of mRNAs with and without ribosomes).
	Ribosome profiling	Ribosome profiling (also called Ribo-seq, **Figure 2**) uses deep sequencing of ribosome-protected mRNA fragments to determine the position of ribosomes on mRNA sequences at sub-codon resolution. Established in both yeast and mammalian cells [13, 14]. Ribosome profiling involves sequencing fragments of RNA that are protected by ribosomes from nuclease digestion (**Figure 2**). Biological samples are treated with translation inhibitory drugs (e.g. cycloheximide) and ribosome-bound RNAs are extracted. The resulting extracts are treated with nuclease to remove unprotected RNA fragments, and monoribosomes are purified by density centrifugation. Short ribosome protected fragments of RNA (RPFs) are then size selected and converted to cDNA libraries for high-throughput sequencing. Ribosome profiling has been widely appreciated for its information-rich data and single-nucleotide level resolution. However, ribosome profiling is very challenging in particular for labs with limited experience in RNA biology. It is difficult to implement due to the requirement of sucrose gradient fractionation, analysis of multiple samples, complex purification steps of tiny amounts of material which include procedures to eliminate rRNA.
Translation rates	Pulsed-SILAC BONCAT	Proteomics methods can measure translation rates using pulsed metabolic labeling and analysis of the incorporation of the isotopic label

	QuanCAT	into proteins over time [15, 16]. Recently developed methods such as BONCAT and QuanCAT, combine isotopic labeling and mass spectrometry with enrichment of newly synthesized peptides via incorporation of a methionine analog [15, 17]. However, proteome coverage is relatively small. In addition, the methods do not provide absolute translation rates unless these rates are estimated from the time-resolved data by regression.
	PUNCH-P	PUNCH-P [18] is a creative proteomics approach to investigate translation via the analysis of nascent protein chains. The methods relies on use of a biotin-tagged elongation inhibitor (puromycin) and extraction of active ribosomes. Nascent proteins are then determined by mass spectrometry. Similarly to the above methods, absolute translation rates can only be estimated if PUNCH-P is combined with quantitative, time-resolved analyses.
	Ribosome profiling with inhibitors	Ribosome profiling alone cannot provide rate information, but in a time-resolved combination of different translation inhibitors, speed of translation has been estimated [14].
Protein degradation rates	Dynamic SILAC	Like translation rates, protein degradation can be measured by mass spectrometry through monitoring the disappearance of isotopic labels over time in variations of dynamic SILAC [1, 19-21]. Degradation rates are then inferred by regression for several thousands of proteins.
	Differential, time-resolved fluorescence imaging	Double-tagged proteins are imaged in high-throughput and analyzed for differential loss of the protein's fluorescence signal compared to the control [22]. While very creative, the approach has disadvantages due to the required library of genetic constructs and the partially artificial setup.

References

1 M. Larance and A.I. Lamond, *Nat Rev Mol Cell Biol*, 2015, **16**(5), 269-280.

2 S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, and M. Mann, *Mol Cell Proteomics*, 2002, **1**(5), 376-386.

3 Z. Wang, X. Sun, Y. Zhao, X. Guo, H. Jiang, H. Li, and Z. Gu, *Genome Biol Evol*, 2015, **7**(4), 1155-1167.

4 A. Khmelinskii, P.J. Keller, A. Bartosik, M. Meurer, J.D. Barry, B.R. Mardin, A. Kaufmann, S. Trautmann, M. Wachsmuth, G. Pereira, W. Huber, E. Schiebel, and M. Knop, *Nat Biotechnol*, 2012, **30**(7), 708-714.

5 M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten, *Science*, 2015, **347**(6220), 1260419.

6 L.S. Churchman and J.S. Weissman, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, 2012, **Chapter 4**, Unit 4 14 11-17.

7 S.E. Munchel, R.K. Shultzaberger, N. Takizawa, and K. Weis, *Mol Biol Cell*, 2011, **22**(15), 2787-2795.

8 M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivière, K.C. Maier, M. Seizl, A. Tresch, and P. Cramer, *Genome Research*, 2012, **22**(7), 1350-1359.

9 B. Neymotin, R. Athanasiadou, and D. Gresham, *RNA (New York, NY)*, 2014, **20**(10), 1645-1652.

10 N. Imamachi, H. Tani, R. Mizutani, K. Imamura, T. Irie, Y. Suzuki, and N. Akimitsu, *Methods*, 2014, **67**(1), 55-63.

11 T. Masek, L. Valasek, and M. Pospisek, *Methods Mol Biol*, 2011, **703**, 293-309.

12 D. Melamed, L. Pnueli, and Y. Arava, *Rna*, 2008, **14**(7), 1337-1351.

13 N.T. Ingolia, S. Ghaemmaghami, J.R. Newman, and J.S. Weissman, *Science*, 2009.

14 N.T. Ingolia, L.F. Lareau, and J.S. Weissman, *Cell*, 2011, **147**(4), 789-802.

15 A.J. Howden, V. Geoghegan, K. Katsch, G. Efstathiou, B. Bhushan, O. Boutureira, B. Thomas, D.C. Trudgian, B.M. Kessler, D.C. Dieterich, B.G. Davis, and O. Acuto, *Nat Methods*, 2013, **10**(4), 343-346.

16 B. Schwanhausser, M. Gossen, G. Dittmar, and M. Selbach, *Proteomics*, 2009, **9**(1), 205-209.

17 D.C. Dieterich, A.J. Link, J. Graumann, D.A. Tirrell, and E.M. Schuman, *Proc Natl Acad Sci U S A*, 2006, **103**(25), 9482-9487.

18 R. Aviner, T. Geiger, and O. Elroy-Stein, *Genes Dev*, 2013, **27**(16), 1834-1844.

19 M.K. Doherty, D.E. Hammond, M.J. Clague, S.J. Gaskell, and R.J. Beynon, *J Proteome Res*, 2009, **8**(1), 104-112.

20 B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, *Nature*, 2011.

21 M. Larance, Y. Ahmad, K.J. Kirkwood, T. Ly, and A.I. Lamond, *Mol Cell Proteomics*, 2013, **12**(3), 638-650.

22 H.C. Yen, Q. Xu, D.M. Chou, Z. Zhao, and S.J. Elledge, *Science*, 2008, **322**(5903), 918-923.