frontiers in
CELL AND DEVELOPMENTAL BIOLOGY

PERSPECTIVE ARTICLE
published: 14 August 2014
doi: 10.3389/fcell.2014.00039

Fingerprints of a message: integrating positional information on the transcriptome

Erik Dassi* and *Alessandro Quattrone**

Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Trento, Italy

Edited by:

Paolo Tieri, Consiglio Nazionale delle Ricerche, Italy

Reviewed by:

Amin R. Mazloom, University of Texas at Arlington, USA
Alessandro Giuliani, Istituto Superiore di Sanità (Italian NIH), Italy

****Correspondence:***

Alessandro Quattrone, Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, via delle Regole 101, Mattarello, Trento 38123, Italy
e-mail: alessandro.quattrone@unitn.it

The recent explosion of high-throughput sequencing methods applied to RNA molecules is allowing us to go beyond the description of sequence variants and their relative abundances, as measured by RNA-seq. We can now probe for RNA engagement in polysomes, for ribosomes, RNA binding proteins and microRNAs binding sites, for RNA secondary structure and for RNA methylation. These descriptors produce a steadily growing multidimensional array of positional information on RNA sequences, whose effective integration only would bring to decipher the regulatory interplay occurring between proteins, RNAs and their modifications on the transcriptome. This interplay ultimately dictates the degree of mRNA availability to translation, and thus the occurrence of cell phenotypes. However, several issues in data presentation are slowing down effective integration. A standardization effort for new dataset types produced should be urgently undertaken to solve these issues. Providing uniformed experimental details along with datasets processed to be directly usable and employing shared formats would greatly simplify integration efforts, strengthening hypotheses stemming from correlative observations and eventually bringing to mechanistic understanding.

Keywords: transcriptome, integration, post-transcriptional control, translation, RNA-seq, mRNA, data format, standards

PROBING THE BIOLOGICAL STATUS OF WHOLE TRANSCRIPTOMES

The last 15 years have witnessed, starting with the advent of microarray-based gene expression probing, an explosion of high-throughput technologies for the characterization of biological molecules. These technologies, affordable and relatively simple to apply, are steadily paving the way for routine multi-omics studies. The latest of such technologies, high-throughput sequencing (HTS) (Metzker, 2010), has quickly gained widespread acceptance and concurrently enabled several different types of measurements. Its sequence-based nature, permitting to pinpoint relevant features on the genome or transcriptome of interest (position-aware data), and its massively parallel data production capabilities are now indeed applied to the study of a wide array of biological questions. Applications focus on DNA (identification of sequence and copy number variants, mapping of chromatin binding sites by transcription factors and other proteins, chromatin topology studies in nuclei, etc.) (Koboldt et al., 2013) and on RNA (sequence variants of mRNAs and non-coding RNAs, expression levels, mapping of binding sites of RNA binding proteins (RBPs), post-transcriptional modifications etc.), (Ascano et al., 2013; Mutz et al., 2013). Translational regulation of gene expression, in particular, has lately been object of increasing interest: its role in profoundly reshaping transcriptome variations and being the determinant of plasticity in the nascent proteome (Vogel et al., 2010; Stevens and Brown, 2013) is increasingly appreciated. Consequently, omic approaches have been developed to investigate which features of an mRNA may influence its translation rate, which trans-factors play a role in such regulatory processes and how these two aspects combine to yield the final protein levels. We will focus on RNA-centered methods to examine the types of biological information they can provide; we will then look at how this information should be integrated to allow us a better understanding of both the global transcriptome dynamics and their effects on phenotype.

As shown in **Figure 1**, such methods can be classified by their descriptive capability, either *molecular* for the entire RNA or *submolecular* for specific RNA portions, and the kind of description they provide, *quantitative*, *qualitative* or both. The description of entire transcripts is provided by RNA-seq (Mutz et al., 2013), an HTS-based method which gives the sequence of coding and non-coding transcripts, including mapping of alternative transcription or termination sites, splice variants produced on the same locus and the presence of expressed sequence polymorphisms. Since different transcripts can be quantified in their relative abundance, this type of information is both qualitative and quantitative. The polysome profiling method (Arava, 2003; Gandin et al., 2014) is based on the separation by sucrose gradient centrifugation of cellular fractions containing polysomes and the subsequent quantification of their mRNA relative (to the total lysate or to the fractions not containing polysomes) abundance, which can be performed by RNA-seq or by the more conventional microarray analysis. The resulting information is a quantitative and qualitative description of the degree of polysomal engagement for every transcript (by which the molecular nature of this method), the so called translatome (Tebaldi et al., 2012); a calculation of

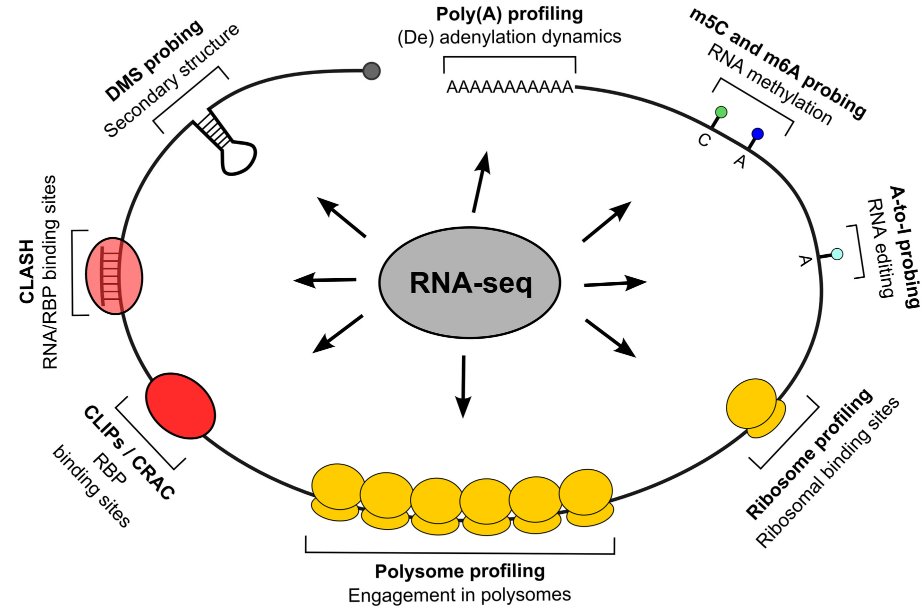

FIGURE 1 | Techniques for positional whole-transcriptome probing. The figure displays techniques allowing to study transcriptomes at various observational levels, with particular regard to positional information; all techniques, indicated by their representative feature on transcripts, are based on RNA-seq.

translational efficiency can be done by this assay. The qualitative component of polysome profiling is given by computational approaches which allow us to investigate the differential association of mRNAs produced by the same gene locus (splice and 5′/3′ variants) with the polysomes (Frac-seq, Sterne-Weiler et al., 2013), or which measure the effect of single-nucleotide polymorphisms on translational efficiency (Li et al., 2013). Ribosome profiling (Ingolia, 2014) aims at providing a snapshot of mRNAs under translation by scoring the transcript regions which are protected from nuclease attack by ribosomes. It is a RNA-seq-based method of the submolecular type: obtainable information can be integrated at the transcript level but has a positional content, so that translation initiation and termination sites, potential translation stalling events, upstream ORF translation, can be derived (Ingolia et al., 2011). Besides engagement in translation, another type of general, qualitative description of transcript status is the secondary structures pattern, recently become available to profiling through nucleotide accessibility probing coupled with RNA-seq (Ding et al., 2014; Rouskin et al., 2014; Talkish et al., 2014; Wan et al., 2014). Eventually, a transcript component which can be investigated is the poly(A) tail: two recent methods, PAL-seq (Subtelny et al., 2014) and TAIL-seq (Chang et al., 2014), exploit RNA-seq to characterize its length and potential modifications (such as uridylation and guanylation). The same principle of nuclease protection exploited in ribosome profiling is then systematically applied in locating RNA-associated "footprints" of RBPs. The CLIP techniques family: HITS-CLIP, PAR-CLIP, and iCLIP (Ule et al., 2003; Hafner et al., 2010; Konig et al., 2010) and the CRAC approach (Granneman et al., 2009) exploit an UV-induced crosslinking of RNA and associated proteins (with the option of using photoactivatable nucleotides, as done in PAR-CLIP) to enable the identification of RNA targets and binding sites for single, immunoprecipitated RBPs. These are therefore submolecular and essentially qualitative approaches. A variant method, CLASH (Helwak et al., 2013), introduces a RNA ligation step to locate sites where other RNAs are associated in trans in a protein complex, allowing to experimentally identify miRNA binding sites. CLIP methods can also be extended to consider many RBPs at once: "global CLIP" approaches such as protein occupancy profiling (Baltz et al., 2012) and PIP-seq (Silverman et al., 2014) thus provide contact sites for all RBPs at once on a transcriptome.

Coming finally to the most submolecular level, that of single nucleotides, mRNA editing events (such as adenosine to inosine conversions) can be revealed either by inosine chemical erasing (ICE), as in Sakurai et al. (2014), or by directly looking for sequence variants in RNA-seq reads (St. Laurent et al., 2013; Bazak et al., 2014). Eventually, RNA 5-methylcytosine and N6-methyladenosine nucleotide methylation can be detected with single-nucleotide precision, respectively by bisulfite conversion (Squires et al., 2012; Edelheit et al., 2013) and immunoprecipitation (Dominissini et al., 2012; Meyer et al., 2012; Khoddami and Cairns, 2013) or by other biochemical methods (Hussain et al., 2013; Liu et al., 2013).

APPROACHES FOR THE INTEGRATION OF TRANSCRIPT-CENTERED OMICS

Currently, several hundred papers employing the described transcriptome-based omics methods have been published, including a considerable number of pure RNA-seq datasets, secondary structure probing, editing and methylation profiles for the most common cell lines and organisms (see **Figure 1**), and at least 40 different CLIP or CLIP-like datasets (Dassi et al., 2014). With such a huge amount of data available, the naturally arising question is how to integrate these different types of information to obtain more insights than if considering single datasets in isolation. Several works have approached this problem so far. As shown in **Table 1**, they can be classified according to the different perspectives adopted in doing so.

A first, post-experimental way of integrating these heterogeneous data sets consists in building a database presenting all the collected data together, thus allowing users to prioritize and validate potential connections. Mining the data, superimposed on a reference genome, can be approached by looking for single genes (as happens in genome browsers) or by studying interesting gene lists (e.g., through functional enrichment or co-regulation analyses). This road was taken by AURA/AURA2 (Dassi et al., 2012, 2014), DORiNA (Anders et al., 2012), and starBase (Li et al., 2014). The first provides RBP and miRNA binding sites, cis-elements sites, RNA editing, and methylated nucleotides; the second offers RBP binding sites and predicted miRNA targets; the last includes RBP binding sites and miRNA interactions with coding and non-coding RNAs. While these databases are of general interest and can be useful for a broad spectrum of preliminary investigations, they still mostly contain data obtained in a limited set of particularly common model systems or cell lines (e.g., HEK293 cells): users will then likely need to trust this information to hold in their system of interest or validate the interaction in their specific conditions (e.g., for an RBP-mRNA interaction, by integrating expression data to check whether it could indeed occur, or by performing a RIP-qPCR assay in their system).

The second, most reliable method is obviously measuring several mRNA features in the system under study, focusing on a specific biological question, and then proceed by intersecting the obtained data to generate hypotheses stemming from the correlation of specific features. An intuitive example of this approach is in profiling the transcriptome and the translatome (the last through polysomal profiling, for instance) in various conditions (e.g., drug treatment vs. control) to identify which genes are subjected to translational control and the impact the treatment may have on translational efficiency (computed as the translatome vs. transcriptome ratio): this has already been done in a number of works (Genolet et al., 2011; Bates et al., 2012; Fu et al., 2012; Tebaldi et al., 2012; Courtes et al., 2013; Dudek et al., 2013; Willimott et al., 2013). A variation on this theme could include, in parallel, a miRNAs profiling in the system to correlate differences in their levels with differences in translational efficiency, generating candidate determinants of the latter changes (Clarke et al., 2012). Another example is the secondary structure and translational efficiency profiling of mRNAs in the system under

Table 1 | Current approaches for positional information integration on the transcriptome.

Name	Description	Scope	Potential issues	References
Integrated databases	Collecting and presenting available datasets of heterogeneous types and biological sources; allowing users to mine the data types in combination	Global over a vast number of different data types	Data quality and processing assessment not always possible; achieving database completeness and constant content update is particularly time-intensive	Anders et al., 2012; Dassi et al., 2012, 2014; Li et al., 2014
Multi-level profiling	Performing various types of measurements (i.e., mRNA levels, RNA secondary structure, RNA methylation) in the same system of interest (e.g., cell line) to derive correlative patterns	Global over a limited number of data types	Need very different experimental and data analysis expertise; results applicability is limited to the studied system	Genolet et al., 2011; Bates et al., 2012; Clarke et al., 2012; Dominissini et al., 2012; Fu et al., 2012; Tebaldi et al., 2012; Courtes et al., 2013; Dudek et al., 2013; Willimott et al., 2013; Zheng et al., 2013; Ding et al., 2014; Mao et al., 2014; Wang et al., 2014a
Measurements & public data exploitation	Performing a small number of measurements (i.e., mRNA levels only) in the system of interest, and exploiting public data to study genes derived from these measurements (i.e., presence of translational regulation) to infer and validate potential regulatory mechanisms and patterns	Over a small number (dozens) of interesting genes	Publicly available data on the system one wants to use may not be available; further validation and/or mechanistic experiments may be needed	Mazza et al., 2013; Avery-Kiejda et al., 2014; Schueler et al., 2014; Wang et al., 2014b

The table describes currently applied approaches to the integration of position-aware RNA datasets. Scope of the various approaches and associated potential issues are outlined along with the references of works employing them.

study, aiming at the identification of structural patterns conferring translational advantages to the mRNAs containing them (Ding et al., 2014; Mao et al., 2014). Along the same line is coupling m6A methylation probing and RNA-seq measurements in the same system: this allows us to understand whether methylation alters mRNA level, stability and splicing patterns in the conditions under investigation (Dominissini et al., 2012; Zheng et al., 2013; Wang et al., 2014a).

The last integration method we describe is based on bridging the previous two approaches: combining a limited number of direct measurements performed in the system of interest with the wealth of data available in public databases such as the ones described above (even though these data may not be produced in the same model). One may thus investigate whether, for instance, an RBP or a miRNA is controlling a group of mRNAs, whether the gene set under analysis is enriched with a particular feature (e.g., a 3′UTR cis-element in the form of a secondary structure, methylated nucleotides, etc.) or match observed patterns for one feature type (e.g., presence of a secondary structure feature) with public data (e.g., presence of trans-factor binding sites) to deduce general rules (e.g., preference of a trans-factor for that given structural feature). While this method leads to hypotheses that need validation as they may not hold in the system of interest, it allows speeding up the investigation and reducing the hypotheses space, consequently lowering experimental uncertainty, time and cost. This approach has been enabled just recently, due to the availability of the databases discussed above. However, in the few published works adopting it, it is usually applied to the integration of data focused on a few specific mRNAs, which have been previously selected for their behavior as observed in the ongoing study (Mazza et al., 2013; Avery-Kiejda et al., 2014; Wang et al., 2014b). One exception is the recent work by Schueler and colleagues, in which protein contact sites obtained by a global PAR-CLIP on two cell lines are integrated with known RBP binding sites to infer differential protein occupancy patterns (Schueler et al., 2014).

Summing up, even though the approaches we have discussed are useful examples of data integration applied to the structure and the behavior of mRNAs, it is evident that these are still early and limited efforts. Indeed, as also testified by the small number of published works, there still is a significant lack of accepted practices and standard procedures which could render these approaches of effective routine usage. Having built a database focused on post-transcriptional regulation (Dassi et al., 2014), we realized that processed data, as submitted by the authors, vary widely in their processing level: if we take CLIPs datasets as an example, some datasets include the definition of sites bound by the studied RBP while others are limited to, for instance, the indication of T > C conversions (for PAR-CLIP); obviously this marked differences put additional burden on whoever wants to use multiple datasets, produced in different experiments, together, in order to generate new hypotheses. Furthermore, methods are often described in many ways, with different levels of detail, representing further obstacles in individuating steps needed to make these datasets truly comparable. A last general issue is the absence of a systematic way to evaluate data quality and robustness, considering for example the presence of replicates, the number of supporting reads and other parameters linked to specific techniques.

THE NEED FOR STANDARDIZATION

Given the outlined issues, we asked which steps could be taken to improve the exploitability and the integration potential of the RNA-centered high throughput data. We propose two simple, preliminary actions. The first is the enforced use of standard file formats with precisely defined fields, a relatively simple goal to achieve. The second is the enforced provision of a minimal set of information—enhancing dataset description, uniformity and allowing quality evaluations—at submission time (similarly to what was established and is currently enforced for microarrays with MIAME and related initiatives; Brazma et al., 2001; Rayner et al., 2006). This could be straightforwardly imposed by repositories commonly used for high-throughput datasets submission such as GEO (Barrett et al., 2013), ArrayExpress (Parkinson et al., 2005), and SRA (Wheeler et al., 2008).

Concerning the first requirement, we need to deal with two types of data: intervals (such as RBP and miRNA binding sites obtained through CLIPs) and per-nucleotide intensities (continuous values such as the ones produced by RNA methylation or secondary structure probing assays). Intervals are most often represented by means of the Browser Extensible Data (BED) format: its main advantage lies in the extreme simplicity of fields definition, which nevertheless allows a certain degree of detail, making it also feasible to represent several datasets in a single file (by for instance using the name field to distinguish the RBP/miRNA and possibly specifying methods and data source publication in the description field). Furthermore, BED files can be converted to bigBed (Kent et al., 2010), the associated binary indexed format that is efficient to process and use with genome browsers even for huge datasets. Concerning continuous values, they are most often stored by means of either a format similar in nature to BED, called bedGraph, or through another common option called Wiggle (Kent et al., 2010). Both formats are stripped down to the essential and are not really intended to allow mixing different datasets in the same file; the file header however leaves room for some description to be added; furthermore, both can be converted to the binary indexed bigWig format (Kent et al., 2010), similarly to what mentioned above for bigBed. Given the versatility and already widespread use of these two formats, coupled with the storage and display efficiency, we propose that they should be deemed as de facto standards and systematically required for new data submissions.

For the second requirement (minimal set of parameters describing a dataset), which information should be considered as essential for the data to be exploited at their full potential? First of all, in the case of CLIP datasets intervals representing binding sites should be provided, rather than including raw per-nucleotide data only. Many scientists would not or cannot go the extra mile to compute intervals out of per-nucleotide data by themselves, and would thus loose the opportunity to use them. Furthermore, methods employed for data analysis should be described, at least briefly, indicating how intervals or per-nucleotide intensities (e.g., in the case of secondary structure data) were computed from raw reads. Eventually, basic quality metrics such as the number of replicates and the read depth supporting a given interval/position, along with call significance *p*-values (where appropriate) should also be provided to let the users judge on the data robustness, eventually allowing the application of homogeneous stringency filters when integrating multiple datasets. We believe that this "information package" could be enough to describe the data under study to an extent that will eventually make going back to the raw data unnecessary: we therefore propose that these information should be required when submitting a dataset of this sort.

Pushing further on this proposal, we may also consider the need for a dedicated repository storing transcriptome-centric positional data. Similarly to what major journals ask for microarrays-containing works, submission to this repository could be a de facto requirement for publication and have an unique ID assigned, to which direct reference could be made in publications further easing data traceability. Using one of the currently available databases as a repository of this sort could also have the advantage of allowing us to display various datasets together, integrated in a transcript-oriented way, thus providing a first glimpse of the data along with the possibility to retrieve them. Of course, this collection of proposals, which goes along the lines of several other "reproducible research" initiatives, can become a reality only if the majority of scientists in the field agree and commit to sustain it by complying with these recommendations.

CONCLUSION

The availability of techniques based on high-throughput sequencing is fostering the investigation of the biological behavior of transcriptomes with an unprecedented level of detail and a continuously increasing amount of available data types: the very nature of this technology effectively allows us to pinpoint the location of features responsible for known and unknown biochemical properties of mRNAs and non-coding RNAs which may ultimately influence mRNA translation. However, the integration of these datasets is still in its infancy, with only a few approaches and applications in the literature and a lot of room for improving and making these efforts much easier and useful. We think that this process could be eased by committing to the introduction of standardization measures involving file formats, minimal information to be provided for dataset description and, possibly, the setup of a dedicated data repository. The choice to advance a proposal limited to transcripts biological features is justified in our opinion by the momentum gained by studies in post-transcriptional regulation of gene expression, by the several RNA-seq-based techniques introduced in the last 2 years, and the exponential growth of datasets of this type being released. We therefore think that the effort needed to implement such proposal could be worthy and fruitful. While certainly requiring coordination between laboratories studying the topic, initiatives like OBO (Smith et al., 2007), MIAPE (Taylor et al., 2007), and BioBricks (Smolke, 2009) have shown that it is possible to implement and sustain a standardization effort aimed in our case at a better exploitation of high-throughput data. Given the pace at which these data are accumulating, we need for sure to urgently push their integrated exploitation to its fullest extent.

REFERENCES

Anders, G., Mackowiak, S. D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., et al. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* 40, D180–D186. doi: 10.1093/nar/gkr1007

Arava, Y. (2003). Isolation of polysomal RNA for microarray analysis. *Methods Mol. Biol.* 224, 79–87. doi: 10.1385/1-59259-364-X:79

Ascano, M., Gerstberger, S., and Tuschl, T. (2013). Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr. Opin. Genet. Dev.* 23, 20–28. doi: 10.1016/j.gde.2013.01.003

Avery-Kiejda, K. A., Braye, S. G., Mathe, A., Forbes, J. F., and Scott, R. J. (2014). Decreased expression of key tumour suppressor microRNAs is associated with lymph node metastases in triple negative breast cancer. *BMC Cancer* 14:51. doi: 10.1186/1471-2407-14-51

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690. doi: 10.1016/j.molcel.2012.05.021

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193

Bates, J. G., Salzman, J., May, D., Garcia, P. B., Hogan, G. J., McIntosh, M., et al. (2012). Extensive gene-specific translational reprogramming in a model of B cell differentiation and Abl-dependent transformation. *PLoS ONE* 7:e37108. doi: 10.1371/journal.pone.0037108

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., et al. (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* 24, 365–376. doi: 10.1101/gr.164749.113

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371. doi: 10.1038/ng1201-365

Chang, H., Lim, J., Ha, M., and Kim, V. N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. *Mol. Cell* 53, 1044–1052. doi: 10.1016/j.molcel.2014.02.007

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., et al. (2012). Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 13:656. doi: 10.1186/1471-2164-13-656

Courtes, F. C., Vardy, L., Wong, N. S., Bardor, M., Yap, M. G., and Lee, D. Y. (2013). Understanding translational control mechanisms of the mTOR pathway in CHO cells by polysome profiling. *N. Biotechnol.* doi: 10.1016/j.nbt.2013.10.003. [Epub ahead of print].

Dassi, E., Malossini, A., Re, A., Mazza, T., Tebaldi, T., Caputi, L., et al. (2012). AURA: atlas of UTR regulatory activity. *Bioinformatics* 28, 142–144. doi: 10.1093/bioinformatics/btr608

Dassi, E., Re, A., Leo, S., Tebaldi, T., Pasini, L., Peroni, D., et al. (2014). AURA 2: Empowering discovery of post-transcriptional networks. *Translation* 2:e27738. doi: 10.4161/trla.27738

Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., and Assmann, S. M. (2014). *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700. doi: 10.1038/nature12756

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112

Dudek, K. M., Suter, L., Darras, V. M., Marczylo, E. L., and Gant, T. W. (2013). Decreased translation of Dio3 mRNA is associated with drug-induced hepatotoxicity. *Biochem. J.* 453, 71–82. doi: 10.1042/BJ20130049

Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., and Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* 9:e1003602. doi: 10.1371/journal.pgen.1003602

Fu, S., Fan, J., Blanco, J., Gimenez-Cassina, A., Danial, N. N., Watkins, S. M., et al. (2012). Polysome profiling in liver identifies dynamic regulation of endoplasmic reticulum translatome by obesity and fasting. *PLoS Genet.* 8:e1002902. doi: 10.1371/journal.pgen.1002902

Gandin, V., Sikstrom, K., Alain, T., Morita, M., McLaughlan, S., Larsson, O., et al. (2014). Polysome fractionation and analysis of mammalian translatomes on a genome-wide scale. *J. Vis. Exp.* doi: 10.3791/51455

Genolet, R., Rahim, G., Gubler-Jaquier, P., and Curran, J. (2011). The translational response of the human mdm2 gene in HEK293T cells exposed to rapamycin: a role for the 5′-UTRs. *Nucleic Acids Res.* 39, 989–1003. doi: 10.1093/nar/gkq805

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9613–9618. doi: 10.1073/pnas.0901997106

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.*Cell* 141, 129–41. doi: 10.1016/j.cell.2010.03.009

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665. doi: 10.1016/j.cell.2013.03.043

Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., et al. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* 4, 255–261. doi: 10.1016/j.celrep.2013.06.029

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* 15, 205–213. doi: 10.1038/nrg3645

Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi: 10.1016/j.cell.2011.10.002

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26, 2204–2207. doi: 10.1093/bioinformatics/btq351

Khoddami, V., and Cairns, B. R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31, 458–464. doi: 10.1038/nbt.2566

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., and Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38. doi: 10.1016/j.cell.2013.09.006

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., et al. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915. doi: 10.1038/nsmb.1838

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248

Li, Q., Makri, A., Lu, Y., Marchand, L., Grabs, R., Rousseau, M., et al. (2013). Genome-wide search for exonic variants affecting translational efficiency. *Nat. Commun.* 4, 2260. doi: 10.1038/ncomms3260

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856. doi: 10.1261/rna.041178.113

Mao, Y., Liu, H., Liu, Y., and Tao, S. (2014). Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 42, 4813–4822. doi: 10.1093/nar/gku159

Mazza, T., Castellana, S., Andriulli, A., Auffray, C., Vinciguerra, M., and Pazienza, V. (2013). Affinity analysis of differentially expressed genes in hepatocytes expressing HCV core genotype 1b or 3a. *Biosystems* 114, 64–68. doi: 10.1016/j.biosystems.2013.05.009

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003

Mutz, K. O., Heilkenbrinker, A., Lonne, M., Walter, J. G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* 24, 22–30. doi: 10.1016/j.copbio.2012.09.004

Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., et al. (2005). ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 33, D553–D555. doi: 10.1093/nar/gki056

Rayner, T. F., Rocca-Serra, P., Spellman, P. T., Causton, H. C., Farne, A., Holloway, E., et al. (2006). A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics* 7:489. doi: 10.1186/1471-2105-7-489

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 505, 701–705. doi: 10.1038/nature12894

Sakurai, M., Ueda, H., Yano, T., Okada, S., Terajima, H., Mitsuyama, T., et al. (2014). A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res.* 24, 522–534. doi: 10.1101/gr.162537.113

Schueler, M., Munschauer, M., Gregersen, L. H., Finzel, A., Loewer, A., Chen, W., et al. (2014). Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol.* 15:R15. doi: 10.1186/gb-2014-15-1-r15

Silverman, I. M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J. L., et al. (2014). RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* 15:R3. doi: 10.1186/gb-2014-15-1-r3

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255. doi: 10.1038/nbt1346

Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.* 27, 1099–1102. doi: 10.1038/nbt1209-1099

Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., et al. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40, 5023–5033. doi: 10.1093/nar/gks144

Sterne-Weiler, T., Martinez-Nunez, R. T., Howard, J. M., Cvitovik, I., Katzman, S., Tariq, M. A., et al. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* 23, 1615–1623. doi: 10.1101/gr.148585.112

Stevens, S. G., and Brown, C. M. (2013). *In silico* estimation of translation efficiency in human cell lines: potential evidence for widespread translational control. *PLoS ONE* 8:e57625. doi: 10.1371/journal.pone.0057625

St. Laurent, G., Tackett, M. R., Nechkin, S., Shtokalo, D., Antonets, D., Savva, Y. A., et al. (2013). Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. *Nat. Struct. Mol. Biol.* 20, 1333–1339. doi: 10.1038/nsmb.2675

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., and Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71 doi: 10.1038/nature13007

Talkish, J., May, G., Lin, Y., Woolford, J. L. Jr., and McManus, C. J. (2014). Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20, 713–720. doi: 10.1261/rna.042218.113

Taylor, C. F., Paton, N. W., Lilley, K. S., Binz, P. A., Julian, R. K. Jr., Jones, A. R., et al. (2007). The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 25, 887–893. doi: 10.1038/nbt1329

Tebaldi, T., Re, A., Viero, G., Pegoretti, I., Passerini, A., Blanzieri, E., et al. (2012). Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC Genomics* 13:220. doi: 10.1186/1471-2164-13-220

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215. doi: 10.1126/science.1090095

Vogel, C., Abreu Rde, S., Ko, D., Le, S. Y., Shapiro, B. A., Burns, S. C., et al. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* 6, 400. doi: 10.1038/msb.2010.59

Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709. doi: 10.1038/nature12946

Wang, W. T., Zhao, Y. N., Yan, J. X., Weng, M. Y., Wang, Y., Chen, Y. Q., et al. (2014b). Differentially expressed microRNAs in the serum of cervical squamous cell carcinoma patients before and after surgery. *J. Hematol. Oncol.* 7:6. doi: 10.1186/1756-8722-7-6

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z., and Zhao, J. C. (2014a). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198. doi: 10.1038/ncb2902

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36, D13–D21. doi: 10.1093/nar/gkm1000

Willimott, S., Beck, D., Ahearne, M. J., Adams, V. C., and Wagner, S. D. (2013). Cap-translation inhibitor, 4EGI-1, restores sensitivity to ABT-737 apoptosis through cap-dependent and -independent mechanisms in chronic lymphocytic leukemia. *Clin. Cancer Res.* 19, 3212–3223. doi: 10.1158/1078-0432.CCR-12-2185

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 June 2014; accepted: 30 July 2014; published online: 14 August 2014.

Citation: Dassi E and Quattrone A (2014) Fingerprints of a message: integrating positional information on the transcriptome. Front. Cell Dev. Biol. ***2****:39. doi: 10.3389/fcell.2014.00039*

This article was submitted to Systems Biology, a section of the journal Frontiers in Cell and Developmental Biology.