CellPress

Trends in Genetics

Spotlight

Moving towards sequencing-based metabolomics

Alia Clark-ElSayed [1], Andrew D. Ellington [1], and Edward M. Marcotte [1,*]

Metabolites are chemically heterogeneous and difficult to quantify in easily read formats. Recently, Tan and Fraser demonstrated that metabolites can be readily quantified by pairing aptamer function with DNA sequencing. This reflects a larger trend of sequencing for assessing biomolecule abundances, further leading to sequencing being a universal analytical tool.

Structure-switching aptamers are single-stranded nucleic acid molecules that change conformation upon binding to effectors. Aptamers have been used extensively for the detection of protein targets, small-molecule targets, and even cells [1]. Although small-molecule responsive aptamers account for a small subset of developed aptamers, they are among the most widely used in biosensing. Several hundred aptamers have been developed for the detection of metabolite targets, including amino acids and antibiotics [2].

While in many cases nucleic acids have been used to detect just one type of molecule, extensive research has been done to allow simultaneous detection of diverse analytes in sample mixtures using arrays [3,4]. Beyond one-to-one analyte-to-aptamer detection, studies have shown that anti-protein aptamers can be used for identifying signatures in molecular mixtures, including cell- and disease-specific patterns [5]. However, to date there has not been an easily usable method for quantifying many different metabolites in parallel within a mixture, either in a one-to-one fashion or for more complex signal identification.

Conformational changes have previously been coupled to fluorescent or electrochemical readouts [6,7], but recently researchers have developed a method that readily couples to DNA sequencing, called small-molecule sequencing (smol-seq) [8]. In this system, the structure-switching aptamer consists of two oligonucleotides, a ligand-binding oligo (LBO) and a short-release oligo (SRO) which is base-paired to the LBO (Figure 1). A ligand binds to the LBO, which is immobilized to a matrix, causing a conformational change in the stem–loop structure of the LBO. This results in the barcode-containing SRO to be displaced from the matrix, which can then be sequenced. Generating a barcode sequence on the SRO that is unique to one metabolite provides a signal readout proportional to the target amount in a mixture. Using this methodology, researchers demonstrate that detecting the barcodes released can quantify metabolite target levels across chemically diverse molecules, including ATP, glucose, ampicillin, lactate, and cortisol. Remarkably, this sequencing method can even differentiate between stereoisomers (D- and L-glucose) which has previously been a difficult task.

However, in order for smol-seq to be used for diagnostic or therapeutic applications, it must be useful and specific in complex mixtures. To this end, *Caenorhabditis elegans* nematodes were exposed to the antimalarial drug piperaquine, and cellular lysates were analyzed for the drug using a piperaquine-specific barcoded aptamer. Barcode quantification revealed that nematodes exposed to piperaquine could be distinguished from a control. Since smol-seq can detect many metabolite targets concurrently by parallel sequencing of target-specific barcodes, this may indicate that concomitant quantification of metabolites in cellular samples is possible, opening up the potential for systematic metabolic profiling. An important question still to be resolved is whether smol-seq can overcome the need for the multiple chemical extractions that are typical of conventional metabolomics protocols to measure both polar and non-polar metabolites.

The prospects for multiplex metabolic analyses are real, given that similar efforts with aptamers that recognize proteins have proved successful. Illumina's proteomics assay (Illumina Protein Prep) based on SOMAmer aptamers can detect 9000 unique human proteins using 11 000 highly specific SOMAmer aptamer reagents. Currently, the Illumina Protein Prep is being used in conjunction with the UK Biobank to analyze over 50 000 clinical samples. The success of proteome characterization with aptamers also reveals the chief issues facing any similar analysis of the metabolome: the development of large numbers of unique, specific aptamers that can recognize individual metabolites. The Human Metabolome Database contains 114 100 annotated metabolites, including peptides, lipids, nucleic acids, carbohydrates, vitamins, and pollutants [9]. SOMAmer arrays were the result of a multi-year, extensive effort to streamline and automate aptamer selection [5,10], and proteins are for the most part better targets for selection than small molecules. That said, machine learning has already proved successful in engineering protein binders and is beginning to show results for the development of high-affinity aptamers. Machine learning models have been trained on aptamer screening data from a single protein target to predict target affinity, enabling the prediction of sequence modifications and even the *de novo* prediction of aptamer sequences [11]. Transformer-based methods for aptamer design have also used structural data for sequence generation, enabling a hybrid structure–sequence

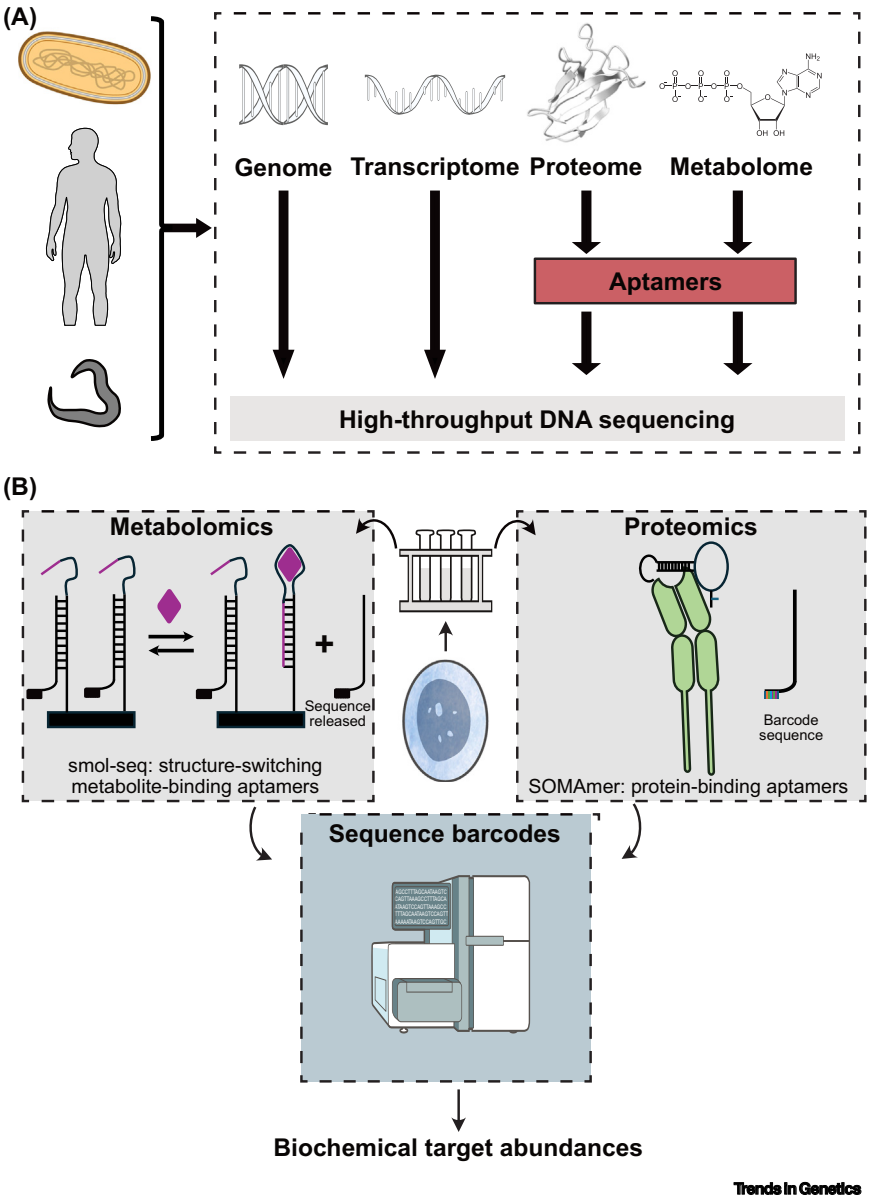

Figure 1. (A) High-throughput DNA sequencing has transformed genomics and transcriptomics, while proteomics and metabolomics were not coupled. Aptamers might now allow these fields to integrate via shared sequencing outputs. (B) Small-molecule sequencing (smol-seq) and SOMAmers enable the quantification of metabolites and proteins, respectively, by coupling the function of specific aptamers to barcode quantification using DNA sequencing.

method for *de novo* generation of RNA aptamers, at least where sequence data for a given target are already known [12].

Such generalizable deep-learning methods, combined with a plethora of data on antimetabolite aptamers (that does not yet exist), have the potential to scale aptamer design. If machine learning or other clever strategies can indeed better enable the parallel development of thousands of antimetabolite aptamers, then smol-seq could greatly streamline the acquisition of metabolomics data, enabling applications in human disease, environmental monitoring, and drug discovery.

Acknowledgments

The authors acknowledge financial support from the National Institutes of Health (T32GM139796 to A.C.E., 1R01 GM146093-01 A1 to A.D.E., and R35 GM122480 to E.M.M.), the Welch Foundation (F-1654 to A.D.E., F1515 to E.M.M.), and the Defense Advanced Research Projects Agency under award number N660012324041. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

Declaration of interests

The authors declare no competing interests.

[1]Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA

*Correspondence:
marcotte@utexas.edu (E.M. Marcotte).

https://doi.org/10.1016/j.tig.2025.04.006

References

1. Dunn, M.R. *et al.* (2017) Analysis of aptamer discovery and technology. *Nat. Rev. Chem.* 1, 0076
2. McKeague, M. and DeRosa, M.C. (2012) Challenges and opportunities for small molecule aptamer development. *J. Nucleic Acids* 2012, 748913
3. Hesselberth, J.R. *et al.* (2003) Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal. Biochem.* 312, 106–112
4. Pei, R. *et al.* (2009) High-resolution cross-reactive array for alkaloids. *Chem. Commun.* 2009, 3193
5. Gold, L. *et al.* (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 5, e15004
6. Nutiu, R. and Li, Y. (2003) Structure-switching signaling aptamers. *J. Am. Chem. Soc.* 125, 4771–4778
7. Xiao, Y. *et al.* (2005) Label-free electronic detection of thrombin in blood serum by using an aptamer-based sensor. *Angew. Chem. Int. Ed.* 44, 5456–5459
8. Tan, J.H. and Fraser, A.G. (2025) Quantifying metabolites using structure-switching aptamers coupled to DNA sequencing. *Nat. Biotechnol.*, Published online February 4, 2025. https://doi.org/10.1038/s41587-025-02554-7
9. Wishart, D.S. *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 46, D608–D617
10. Kraemer, S. *et al.* (2011) From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PLoS One* 6, e26332
11. Bashir, A. *et al.* (2021) Machine learning guided aptamer refinement and discovery. *Nat. Commun.* 12, 2366
12. Wong, F. *et al.* (2024) Deep generative design of RNA aptamers using structural predictions. *Nat. Comput. Sci.* 4, 829–839