Charting the "unknown unknowns" of cancer progression

Hani Goodarzi[1,2,3]

Integrated computational and experimental strategies reveal previously unknown regulatory programs underlying metastatic disease.

Metastasis, the process in which tumor cells acquire the ability to disperse to and grow in distal organs, is the ultimate cause of death in most cancer patients. During this process, tumor cells undergo gene expression reprogramming to acquire phenotypes that promote metastatic progression. My research over the past several years has focused on discovering regulatory programs that underlie pathologic changes in gene expression in highly metastatic cells. The unifying theme of my work is the development of integrated frameworks that enable a systems-level analysis of gene expression dynamics. My background in computational and experimental biology has provided me with the opportunity to set up an interdisciplinary research program focused on cancer systems biology. One of my recent publications, which reported the discovery and characterization of a pathway that promotes breast cancer metastasis through posttranscriptional destabilization of metastasis suppressor genes (1), integrated advanced methods in large-scale gene expression profiling, algorithm development, large data analysis, and focused in vivo and in vitro functional characterizations of the identified components of the pathway. The result was the discovery of a new mode of gene regulation mediated by an RNA binding protein interacting with stem-loop hairpins embedded in metastasis-regulating transcripts. Ultimately, it is this multidisciplinary approach to studying cancer that has defined my identity as a researcher.

The start of my graduate career coincided with the emergence of genome-wide data collection efforts, and I became interested in harnessing these data sets to provide a bird's-eye view of cellular states in normal and pathological conditions. Given its inherent complexity, cancer was a fitting biological problem to tackle. Gene expression profiles had been pub-lished for many different types of cancer, yet the accompanying analyses were often limited to signature gene identification and lacked a principled dissection of more complex expression patterns. To address this problem, I developed a computational pipeline called iPAGE to discover regulatory pathways involved in different types of cancer, together with O. Elemento and under the supervision of S. Tavazoie (2). We used iPAGE to generate a detailed picture of the regulatory states of tumors across the spectrum of 53 distinct human malignancies. We captured many known and previously unknown regulatory interactions—a subset of which were confirmed through experimental follow-up—that underlie oncogenesis in different cancers. This study was the first of its kind in providing a detailed portrait of regulatory perturbations in cancers (2). To provide access to this tool for the scientific community, we also developed the iGET web portal (https://iget.c2b2.columbia.edu), now hosted by Columbia University, which currently hosts ~1400 registered users.

A surprising finding uncovered by the aforementioned study was the abundance of putative regulatory sequence motifs in non-coding regions of transcripts. Most of these motifs were previously uncharacterized, with only a token number matching known RNA binding protein binding sites or microRNA recognition sites. This observation highlighted the paucity of data, both computational and experimental, on the action of posttranscriptional regulatory programs and their roles in the emergence of disease. The main challenge in discovering posttranscriptional regulatory interactions is accounting for the contribution of RNA secondary structure to modulating RNA interactions with proteins. There are many documented instances where the presence of an RNA structural element dic-tates key steps in the life cycle of RNA molecules. As such, characterization of posttranscriptional regulatory programs requires capturing information provided by both the local secondary structure and the underlying sequence of RNA molecules. In a multilaboratory collaborative effort, we addressed this challenge by taking advantage of optimization algorithms and highly parallel processing on supercomputers. The resulting computational approach, named TEISER, systematically explores the immense space of small structural elements to discover those that are informative of transcriptomic measurements (3). To test this framework, I devised a noninvasive experimental technique, based on pulse-chase metabolic labeling of cellular RNA using 4-thiouridine, to measure transcript stability genome-wide. By applying TEISER to this data set, I identified and validated a set of previously uncharacterized regulatory networks that modulate the stability of endogenous transcripts in the cell (3).

Given the success of TEISER when applied to RNA decay rates, my goal was to expand this approach to identify the regulatory drivers of pathologic cellular states. As a postdoctoral fellow at the Rockefeller University, I used TEISER to identify posttranscriptional regulators of metastatic breast cancer. I performed differential transcript stability measurements in in vivo–selected cells that were highly metastatic relative to their parental lines to identify the regulatory interactions that are co-opted by the cancer cells en route to metastatic progression (1). Using this approach, I discovered a new structural element that was sufficient for the differential destabilization of its host transcripts. With help from a team of scientists in the Tavazoie laboratory at the Rockefeller University, we subsequently identified the double-stranded RNA binding protein TARBP2 (TAR RNA binding protein 2) as a functional binding partner for these elements and used mouse models to show that silencing TARBP2 impaired metastasis. We successfully charac-terized two transcripts within the TARBP2

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA. [2]Department of Urology, University of California, San Francisco, San Francisco, CA 94158, USA. [3]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA. Email: hani.goodarzi@ucsf.edu

Goodarzi, *Sci. Transl. Med.* **9**, eaao0959 (2017)   26 July 2017

1 of 3

A TARBP2-mediated destabilization of metastasis suppressors
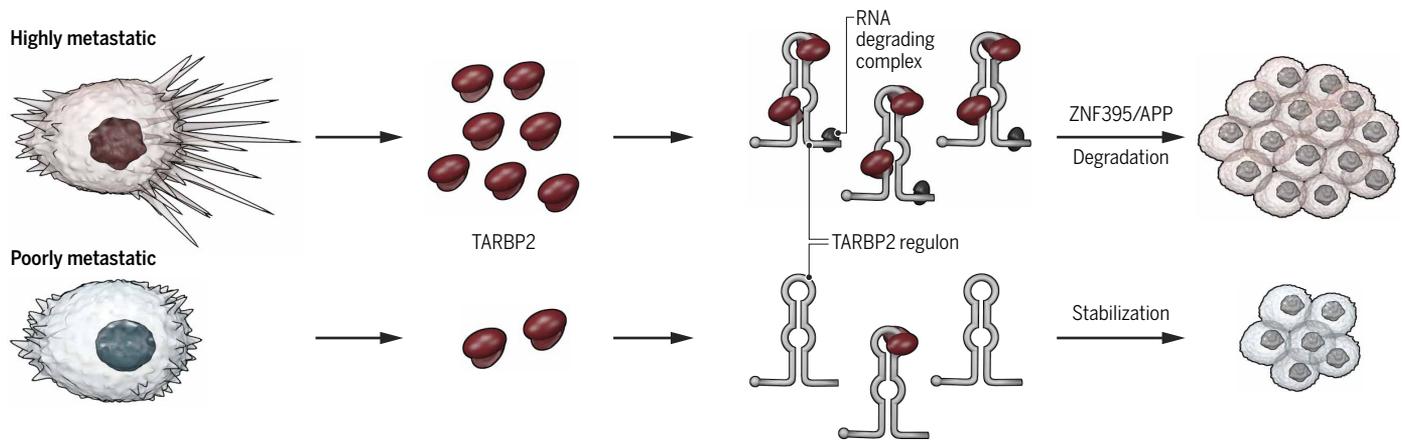
B Pathologic modulation of tRNAs drives cancer progression
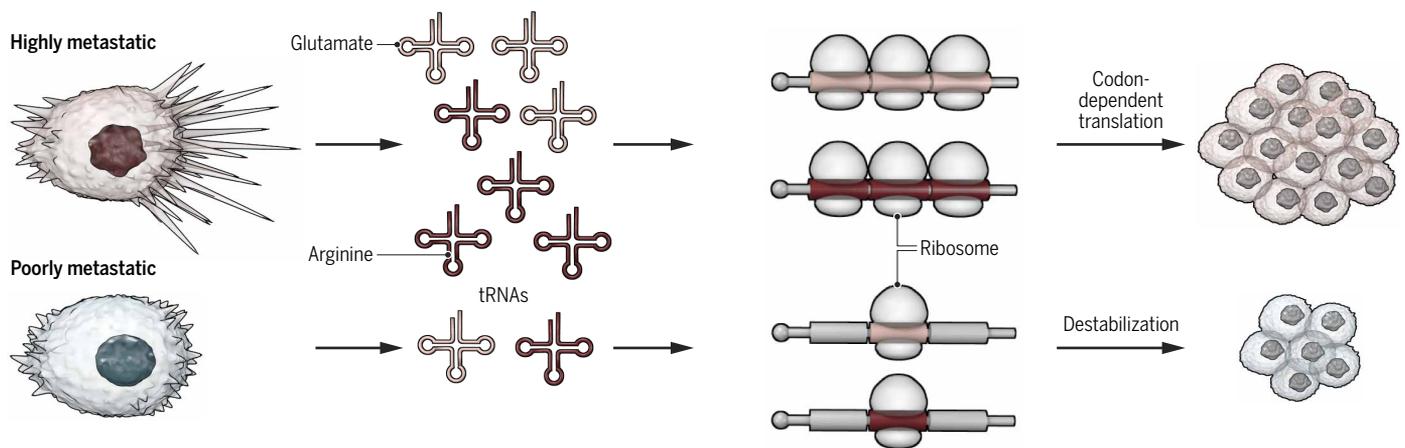
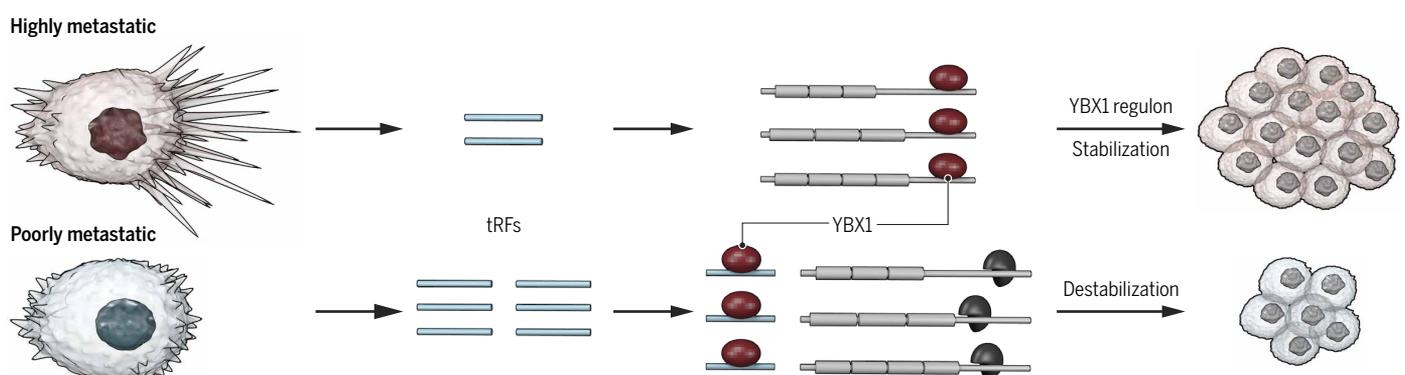## C Stress-induced tRNA fragments suppress metastasis

Fig. 1. Discovery of regulatory pathways that drive cancer metastasis. (**A**) In highly metastatic cells, TARBP2 is transcriptionally up-regulated. This increased expression, in turn, leads to a higher rate of degradation and lower expression for its target transcripts. Because key suppressors of metastatic progression, such as ZNF395 and APP, are among TARBP2 targets, hyperactivation of this posttranscriptional pathway results in higher metastatic colonization of the lung by breast cancer cells. (**B**) The abundance of specific tRNAs is dysregulated in metastatic breast cancers. Changes in tRNA expression in turn affect the translation and degradation rates of mRNAs based on their codon content. Cancer cells can hijack this pathway to modulate the expression of key promoters and suppressors of metastasis. (**C**) Under stress, specific tRNAs are fragmented to form tRFs, a class of small noncoding RNAs. A number of these fragments, which are generated in poorly metastatic breast cancer cells, compete with endogenous transcripts for binding to the protective RNA binding protein YBX1. As a result of this reduced YBX1 activity, a large number of its target transcripts are destabilized and degraded. Because the YBX1 regulon includes key promoters of metastatic progression, tRF induction acts as a suppressive mechanism, which is largely blunted in highly metastatic cells.

Goodarzi, *Sci. Transl. Med.* **9**, eaao0959 (2017)   26 July 2017

2 of 3

regulon—using in vivo and in vitro epistasis experiments—as suppressors of breast cancer metastasis that are destabilized by TARBP2 binding. Analysis of clinical samples and data sets further supported the role of this TARBP2-mediated pathway in driving breast cancer metastasis. The discovery of this regulatory network exemplifies the power of effective data analysis, when combined with rigorous experimentation, to reveal previously unknown regulatory interactions.

One facet of my research has been centered on identifying regulators that play major roles in normal cell physiology and disease. Early in my graduate studies, we observed that codon usage (the frequency of each codon in a gene) is highly informative of gene ontology and that genes involved in the same pathways show similar trends in their codon usage (4). In a subsequent study, in collaboration with H. Najafabadi, we demonstrated that these codon usage profiles can be used to annotate unknown coding sequences in newly sequenced genomes. By expressing a combinatorial library of transfer RNAs (tRNAs) and performing genetic screens in *Escherichia coli*, we successfully revealed that tRNA copy number associates with the emergence of specific phenotypes (5). These observations raised the possibility that tRNAs play an active role in modulating gene expression and can promote different cellular states and phenotypes. However, in the absence of reliable approaches for tRNA quantification, this question could not be effectively addressed. Therefore, during my postdoctoral training in S. Tavazoie's laboratory, we set out to develop technologies for measuring the tRNA content of a cell genome-wide. By combining new algorithmic design with an experimental approach devised to harness the power of high-throughput sequencing, we developed and benchmarked a method for tRNA quantification (6). By applying this method to poorly and highly metastatic cells, we demonstrated that the tRNA landscape is dramatically changed in highly metastatic breast cancer cells and that these changes drive the expression of key metastasis

promoters. We showed that two specific tRNAs, coding for glutamate and arginine, are genomically amplified and that their higher expression affects the stability and translation of genes in a codon-dependent manner. We also demonstrated that exogenous modulation of these tRNAs in the cell affects the ability of breast cancer cells to metastasize (6). In addition to revealing a regulatory mechanism hijacked by cancer cells, this study challenged the traditional view of tRNAs as a static component of the protein synthesis machinery.

Another aspect of my research has focused on the role of small noncoding RNAs in driving breast cancer metastasis. Recent studies have established specific microRNAs as promoters or suppressors of metastasis. However, our knowledge of posttranscriptional regulatory programs mediated by other classes of small RNAs is limited. For example, tRNAs are cleaved under stress to give rise to tRNA-derived fragments (tRFs) (7), yet the regulatory programs mediated by these tRFs are poorly characterized. By performing comparative small RNA sequencing, we identified a group of tRNA fragments that are produced under hypoxia in poorly metastatic but not in highly metastatic breast cancer cells. Affinity purification of synthetic hypoxia-induced tRFs followed by mass spectrometry identified the oncogenic RNA binding protein YBX1 (Y-box binding protein 1) as an interaction partner of these tRFs. These hypoxia-induced tRFs, upon induction, reduce YBX1 binding to other endogenous transcripts by competing for YBX1. As a result, the YBX1-bound transcripts, identified using cross-linking and immunoprecipitation of RNA (HITS-CLIP), are in turn down-regulated. This YBX1-tRF regulon contains hundreds of transcripts, including many known promoters of tumorigenesis and metastasis. Through in vitro assays and using in vivo xenograft mouse models, we demonstrated that these tRFs act as suppressors of metastasis and reduce cancer cell invasion and cell proliferation under starvation conditions (Fig. 1). We verified these results in multiple independent breast cancer cell lines. These

findings revealed that breast cancer cells attenuate the induction of these fragments en route to higher metastatic capacity (8).

Going forward, as an assistant professor at the University of California, San Francisco, I am building a multidisciplinary research program where we combine integrated technology development and sophisticated computational analyses with focused in vitro and in vivo experimentation to reveal regulatory programs that drive human disease. We are also applying our methodologies to other biological problems through collaborative efforts.

REFERENCES

1. H. Goodarzi, S. Zhang, C. G. Buss, L. Fish, S. Tavazoie, S. F. Tavazoie, Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. *Nature* **513**, 256–260 (2014).
2. H. Goodarzi, O. Elemento, S. Tavazoie, Revealing global regulatory perturbations across human cancers. *Mol. Cell* **36**, 900–911 (2009).
3. H. Goodarzi, H. S. Najafabadi, P. Oikonomou, T. M. Greco, L. Fish, R. Salavati, I. M. Cristea, S. Tavazoie, Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* **485**, 264–268 (2012).
4. H. Goodarzi, N. Torabi, H. S. Najafabadi, M. Archetti, Amino acid and codon usage profiles: Adaptive changes in the frequency of amino acids and codons. *Gene* **407**, 30–41 (2008).
5. H. S. Najafabadi, H. Goodarzi, R. Salavati, Universal function-specificity of codon usage. *Nucleic Acids Res.* **37**, 7014–7023 (2009).
6. H. Goodarzi, H. C. B. Nguyen, S. Zhang, B. D. Dill, H. Molina, S. F. Tavazoie, Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* **165**, 1416–1427 (2016).
7. M. M. Emara, P. Ivanov, T. Hickman, N. Dawra, S. Tisdale, N. Kedersha, G.-F. Hu, P. Anderson, Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J. Biol. Chem.* **285**, 10959–10968 (2010).
8. H. Goodarzi, X. Liu, H. C. B. Nguyen, S. Zhang, L. Fish, S. F. Tavazoie, Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* **161**, 790–802 (2015).

Goodarzi, *Sci. Transl. Med.* **9**, eaao0959 (2017)   26 July 2017

3 of 3

Science Translational Medicine

Charting the "unknown unknowns" of cancer progression

Hani Goodarzi

ARTICLE TOOLS	http://stm.sciencemag.org/content/9/400/eaao0959
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/9/397/eaan0026.full http://stm.sciencemag.org/content/scitransmed/9/384/eaai8504.full http://stm.sciencemag.org/content/scitransmed/8/364/364ra153.full http://stm.sciencemag.org/content/scitransmed/8/361/361ra138.full http://science.sciencemag.org/content/sci/360/6394/eaao4908.full http://stm.sciencemag.org/content/scitransmed/10/453/eaat8285.full http://stm.sciencemag.org/content/scitransmed/10/453/eaat8288.full http://stm.sciencemag.org/content/scitransmed/11/501/eaay1018.full
REFERENCES	This article cites 8 articles, 1 of which you can access for free http://stm.sciencemag.org/content/9/400/eaao0959#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions