

EXPLORING CANCER METASTASIS OUTSIDE THE GENOME

Interview with Dr. Hani Goodarzi



BY TIMOTHY JANG AND ANANYA KRISHNAPURA

Hani Goodarzi, PhD, is an assistant professor of the Departments of Biophysics & Biochemistry and of Urology at the University of California, San Francisco. He is also a member of the Helen Diller Family Comprehensive Cancer Center as well as the Institute for Computational Health Sciences. Dr. Goodarzi is the principal investigator of the Goodarzi Lab, which combines computational and experimental approaches in its study of cancer systems biology. The laboratory's current research is largely focused on the metastasis of different cancers and neurodegenerative disease. In this interview, we discuss his research on post-transcriptional pathways affecting breast cancer metastasis to the lung as well as his current work on SARS-CoV-2.

BSJ: Much of your research focuses on the metastasis of different cancers and how to address the challenges this poses on a molecular level. What initially drew you to this topic in particular?

HG: I trained as a computational biologist, so I come from more of a theoretical background. I saw myself as a data scientist more than anything else before data scientists were even called data scientists. When I started graduate school, there was this explosion in biological data. There was this surge of the application of microarrays, which were pretty new at the time, to measure mRNA expression genome-wide in different organisms and for different conditions. Specifically, we were seeing the birth of these precision medicine applications of microarrays to profile different types of cancers, look at their gene expression patterns, and learn something from them. It was relatively good timing for someone with my background to start thinking about how to aggregate and integrate these types of data sets and learn something from them in a broad perspective. I ended up joining Saeed Tavazoie's lab at Princeton, which had a computational and experimental side. I started on the computational side, where I trained with a postdoc, Olivier Elemento, and we started this project, asking, "How can we make sense of the broad regulations that happen in the context of cancer? What is the identity of cancer cells from a molecular perspective?"

Essentially, my introduction to cancer was predominantly an accidental one, in the sense that I initially simply cared about data. Over time, the nature of my interactions with cancer changed quite a bit. I came to realize that there is no amount of computation and statistical analysis that would make an association into causation, so if I really believed in the way that I was doing cancer research, I owed it to myself to take it to the next level. I slowly geared towards picking up genomics and later multi-omics types of analysis of biological systems. When it came time to be a postdoc, I joined a traditional cancer biology lab, Sohail Tavazoie's lab at Rockefeller. My focus on metastasis was, in part, also accidental in that his lab studied metastasis under the understanding that if we want to limit mortality from cancer, we really have to pay attention to metastasis because, especially for cancers that are operable, the real cause of mortality is metastatic dissemination.

BSJ: What initially led you to explore the possibility of a post-transcriptional regulatory pathway for cancer progression and metastasis?

HG: The idea came from an intersection of two different perspectives that I developed as a postdoc. I had initially approached cancer under the operating idea that there are regulatory and signaling pathways in the cell that are hijacked by cancer cells

"I started to think about the possibility that cancer cells can step outside of that—maybe they can engineer their own regulatory pathways."

in order to achieve the types of dysregulation needed to elicit their growth and spread. However, I started to think about the possibility that cancer cells can step outside of that—maybe they can engineer their own regulatory pathways through rewiring and gene expression control mechanisms that do not exist in normal cells. This was one perspective I was thinking about at the time. Expanding on that, I also began to think about what you could possibly need for something like this to be true. One of these possibilities would be the existence of a pool of macromolecules with regulatory potential in cancer cells that are just normally not around.

Coincidentally, at the time, I was working on this other project on tRNA fragments. The way that we studied tRNA fragments was through this approach called small RNA sequencing, which captures all the small RNAs (not just the tRNA fragments). As I was poring over that data, I noticed quite a few of these other RNAs every now and then in the genome that were just not annotated. At some point, as we continued to perform small RNA sequencing for various projects, it clicked that we do not see this category of small RNAs as much when we look at normal tissues; however, we see them in cancer cells. This suggested that there was this population of small RNAs that are not annotated and are cancer-emergent.

BSJ: How did you isolate this set of small RNAs specific to cancer cells?

HG: We performed small RNA sequencing across cell lines from different breast cancer subtypes and compared them to human mammary epithelial cells (HMECs), which serve as non-cancer models. We then combined our results with data from The Cancer Genome Atlas (TCGA) data set and analyzed this data to find small RNAs specific to cancer cells. We called these molecules “orphan noncoding RNAs” (oncrRNA). We borrowed this terminology from bacterial genetics, where “orphan genes” refer to genes that uniquely appear in a given species. Here, there is a similar idea of these RNA molecules simply appearing in cancer cells. In turn, cancer cells can then learn to adapt them for new functions.

BSJ: In one of your papers, you describe how one oncrRNA, T3p, has a strong association with breast cancer progression. How did you demonstrate whether T3p directly affects cancer progression and metastasis?

HG: We first looked through our lists of orphan RNAs and searched for those that were associated with tumor progression such that they not only appear in cancer cells, but their levels increase as the tumor progresses. T3p came out of that process, but as I said before, association is never causation. In order to prove causation, we performed loss of function experiments to test whether, if we took away T3p, we would see an effect on the biology

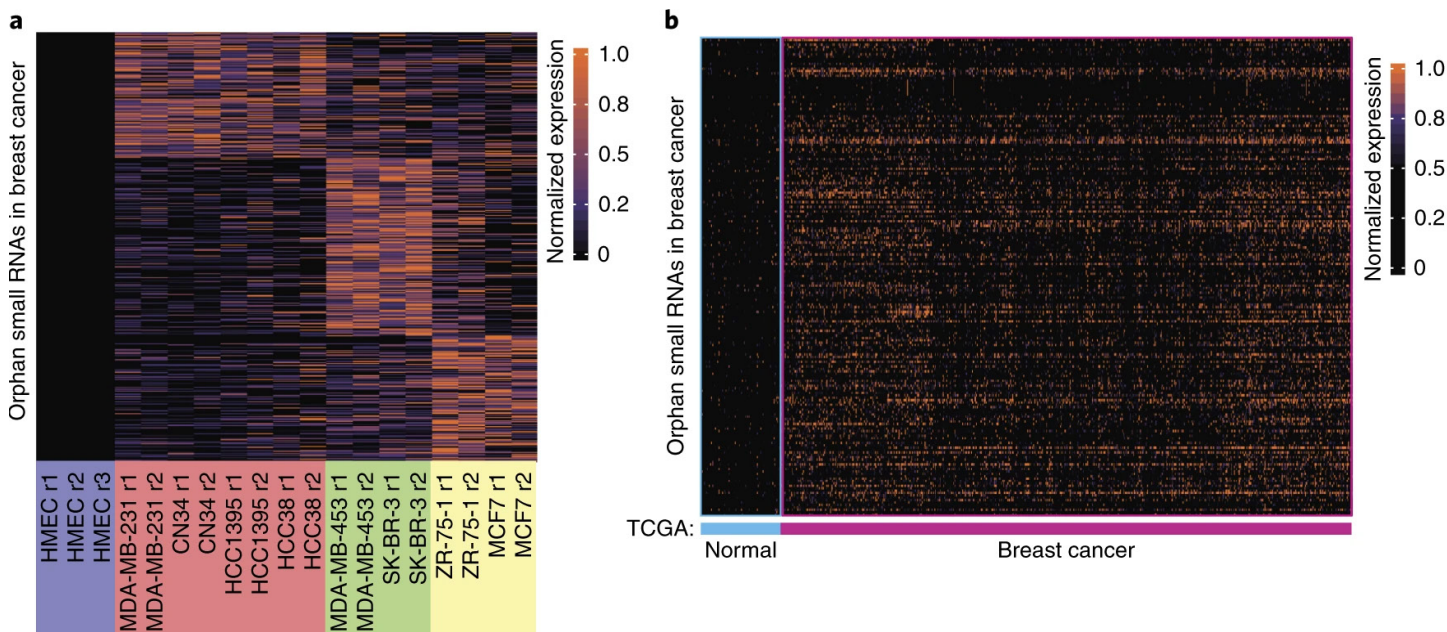


Figure 1: Figure 1a is a heat map depicting the significant expression of 437 small noncoding RNAs in breast cancer cell lines (red, green, and yellow groups) as compared to their non-significant expression in normal cell lines, as represented by human mammary epithelial cells (HMECs). In Figure 1b, The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) data collection was used to identify a subset of these smRNAs that was significantly expressed in breast cancer biopsies as well as absent in the surrounding normal tissue. The resulting 201 smRNAs are defined as orphan noncoding RNAs (oncrRNAs).²

of the cell cancer-related phenotypes. To do this, we used a class of antisense RNAs called locked nucleic acids (LNAs) that form very stable duplexes with small RNAs. They have been used historically to look at other small noncoding RNAs with regulatory functions, such as microRNAs. We used LNAs against T3p to see if, after we inhibit T3p, we can see specific changes in gene expression patterns in the cell and, more importantly, changes in their metastatic capacity. This was measured using xenograft mouse models, where we were able to implant or inject tumor cells into immunocompromised mice and measure how metastatic or aggressive the tumor is. We used these assays to measure the ability of cancer cells to colonize the lungs after perturbations of T3p. We were ultimately able to demonstrate that there is indeed a functional link between T3p expression and metastasis.

BSJ: In the article, you discuss T3p's relationship with the RISC complex. What is the general function of the RISC complex, and how does T3p interact with it?

HG: As I mentioned earlier, there is this class of small noncoding RNAs called microRNAs. These molecules are loaded into the RISC complex and serve to recognize target RNA molecules for degradation through base pairing. When the microRNA recognizes a complementary sequence on a target molecule, the RISC complex will cut this target RNA, leading to its degradation.

Regarding T3p, once we were able to show that T3p had a direct effect on cancer progression, the next question we had to answer was, "What is its mechanism of action?" Since we are a half-computational lab, we had already pre-built a lot of the tools and data sets necessary in order to analyze the interaction potential of RNAs. This included what are known as CLIP data sets, data sets generated for RNA-binding proteins in order to show where they bind. Through this analysis, we found that Argonaute 2 (AGO2), a key enzyme of the RISC complex, was bound to T3p. That meant one of two things. One was that T3p itself could potentially function as a microRNA. Alternatively, T3p could

instead be a target of the RISC complex, meaning that it interacts with a microRNA already loaded into the RISC complex.

In regards to the first possibility, T3p was already a bit too long to be a microRNA to begin with, and we could not find a seed sequence that would explain the gene expression changes as a result of its presence in cells. We thus ruled out the first possibility and landed on the second possibility, where the T3p is binding to the RISC complex in the context of other microRNAs. We then looked for the specific microRNAs that could target T3p, and we found a few. We tested them experimentally to see if they actually do form a complex, and we showed that two of them directly bind T3p. Additionally, we showed that T3p levels are modulating the gene expression of a few targets through these couple of microRNAs. That was how we landed on the link between T3p and the RISC complex.

BSJ: Have you been able to explore the clinical implications of this link between orphan noncoding RNAs and cancer progression?

HG: Yes, since the paper's publication, we have started a retrospective collaborative project with the I-SPY breast cancer trial at UCSF where we look at the oncrRNA content of serum samples from breast cancer patients and determine how it changes upon treatment or how it relates to the size of the tumor or residual disease. This is one of the directions we are pursuing in order to find out if we can use liquid biopsies built around the detection of oncrRNAs to stratify patients by risk.

Outside of just diagnostics, I want to add that there is also a possibility of having orphan RNAs serve as therapeutic targets. Since they are not traditional therapeutic targets, we are still in the early days of exploring what is possible, but the bottom line is that they can serve as novel targets that likely have limited toxicity. This is due to the fact that most of the pathways that currently serve as therapeutic targets function in normal cells as well as cancer cells. Thus, once you exceed the therapeutic window, you are hitting normal cells as well as cancer cells, resulting in on-target

"Outside of just diagnostics, I want to add that there is also a possibility of having orphan RNAs serve as therapeutic targets."

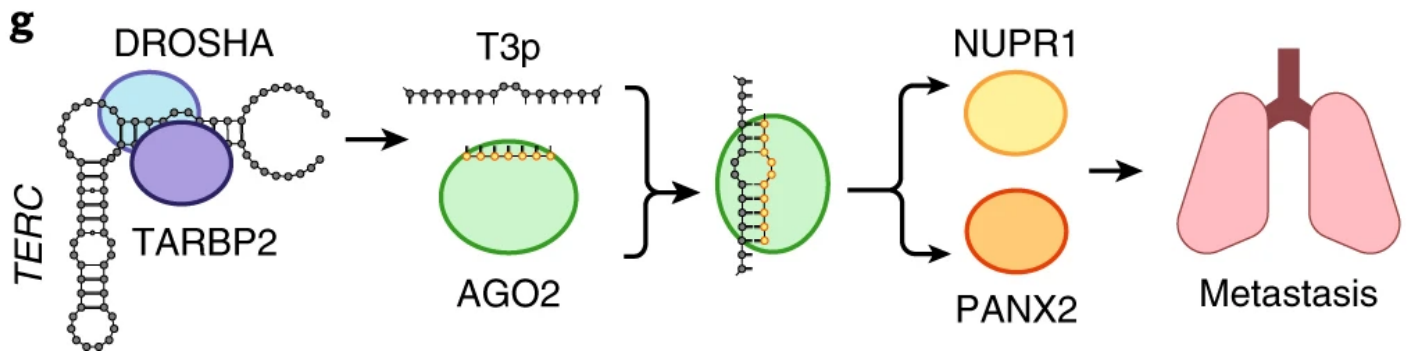


Figure 2: Model of the pathway through which T3p drives cancer metastasis. When expressed, T3p binds to Argonaute 2 (AGO2) in RISC complexes, preventing miR-10b-5p and miR-378c-5p from binding. These miRNAs are thus unable to silence expression of their downstream target genes, NUPR1 and PANX2. Elevated expression of these genes is associated with metastasis of breast cancer to the lungs.²

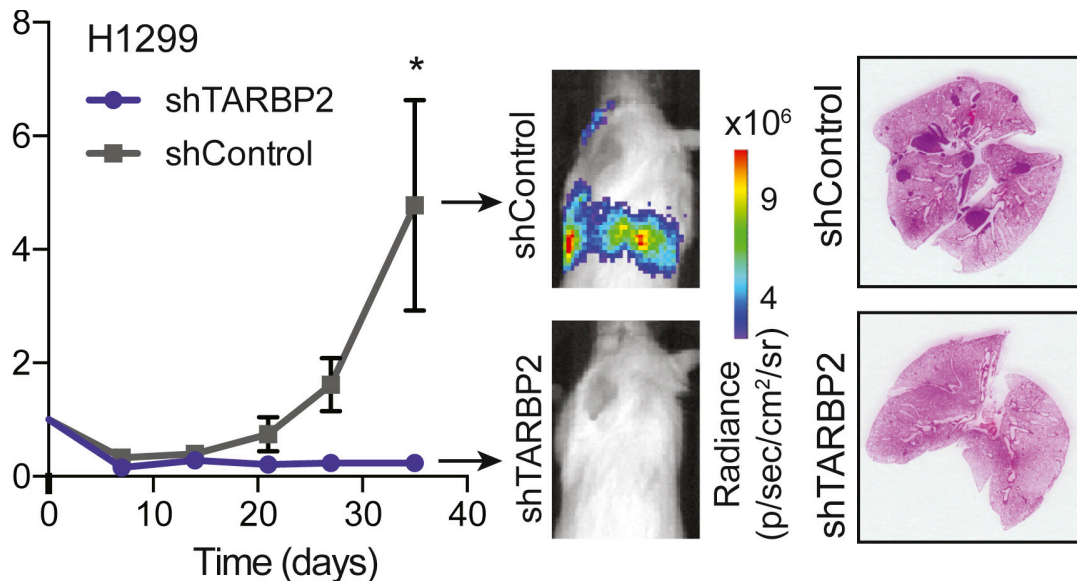


Figure 3: Graph of lung bioluminescence signal over time for mice injected with H1299 lung cancer cells that either express a control shRNA (in gray) or a shRNA targeting TARBP2 (in blue). Note the greater bioluminescence and presence of tumors in lungs of the control group.³

toxicity. However, targeting functional oncRNAs would not result in this toxicity, since they are not present in normal cells by definition.

BSJ: Another one of your papers deals with the RNA-binding protein TARBP2 and its oncogenic implications through its involvement in targeted intron retention. How did you initially come to hypothesize that TARBP2 was involved in this pathway?

HG: TARBP2 was actually one of the first genes I studied as an experimental cancer biologist. When I started as a postdoc, I was studying the changes in RNA stability we see when we compare poorly and highly metastatic breast cancer cells. I found a sizable regulon of genes that were changing the RNA stability in highly metastatic cells, but it was not clear as to why. At the time, most of what we knew about RNA stability had to do with microRNAs or some RNA binding proteins, but when I looked at those, none of them could explain the changes. That implied that there was an unknown mechanism through which the stability of these targets is being dysregulated in highly metastatic cells; landing on these kinds of problems is, in fact, my job. As a systems biologist, I try to build regulatory pathways from scratch, as opposed to relying on what is known. So, I took advantage of a kind of custom application of network biology: given a set of genes that are changing together, can you figure out an associated factor correlated with all these genes? In other words, can you identify a master regulator of genes of a regulon, where if that regulator changes, so will the targets?

To answer that, I essentially ran a lot of correlation analyses regarding gene expression. Through this network biological approach, I nominated three potential RNA-binding proteins to be regulators of RNA stability in this context. TARBP2 was one of them. I knocked down each one of them and measured changes in RNA stability, and TARBP2 was the one that was the right candidate. For the remainder of that paper, which came out in *Nature* back in 2014, we really focused on its function in metastasis. We showed through xenograft mouse models that if you change TARBP2 activity

and expression, you can modulate the metastatic capacity of cancer cells; however, it was not really clear what the actual mechanism was through which TARBP2 regulates RNA stability. This is where the second paper, which is basically a follow-up, comes into play. Basically, we were trying to find how TARBP2 is functioning to change the stability of its target regulon.

BSJ: How did you then narrow down how TARBP2 operates at a molecular level?

HG: We first made a couple of important observations. TARBP2 had a known function as part of the microRNA processing machinery, where it was thought to be a cytoplasmic RNA binding protein. However, our results localized its function to the nucleus; if you knock TARBP2 down, the stability of its target genes is changing inside the nucleus. This meant that we had this nuclear RNA stability pathway that was different from what was known.

We then used pull-down mass spectrometry to target TARBP2 and all the proteins that it could interact with in order to examine its function. This included the key components of the methyltransferase complex and also TPR, a component in nuclear pore-associated proteins involved in RNA surveillance and export. We next modulated the levels of these proteins and observed whether we could see a similar effect on the action of TARBP2 and its target regulon. We used this epistasis experiment to prove that TARBP2 is upstream of RNA methylation, which is upstream of our target regulon. In this manner, we revealed how TARBP2 binds to mostly intronic sequences as the RNA is transcribed. It recruits a methyltransferase complex, which methylates the RNA, and these methylation marks are then used as flags to regulate the rate of splicing.

Therefore, if TARBP2 is present, you get less efficient splicing and intron retention. In the nucleus, RNA that is not properly spliced is very quickly degraded. We ultimately think that at the same time TARBP2 is recruiting the methyltransferase complex and prohibiting

"I am broadly interested in gene expression control, and by understanding where it breaks, we learn how it works."

efficient splicing, through its interactions with TPR, it simultaneously brings this surveillance complex to these target transcripts, resulting in their degradation.

BSJ: What is the association between TARBP2 expression and cancer *in vivo*?

HG: In our first paper on the subject, we established that TARBP2 had a role in metastasis to the lung of breast cancer. In binding to its target RNA sequences, TARBP2 results in their degradation, promoting metastasis. On top of that, once we had the signature of TARBP2 and its targets, we looked broadly at where else this pathway could be functional. One of the places that we looked at was cancer gene expression data sets, and we identified breast cancer, which made sense. But even stronger than that, we saw a signal in lung cancer, which is why we started to go down that path and look at how modulations of TARBP2 will impact cancer growth.

BSJ: What are the implications of having a greater understanding of these post-transcriptional pathways in cancer cells?

HG: The models that we are using and creating are not just significant for the study of human disease, but they are useful in exploring normal cell physiology as well. As I mentioned earlier, most of the pathways that we find dysregulated in the context of cancer are performing their normal functions in normal cells. They are being perturbed in the context of cancer, but their identities are not changing. If, for instance, A regulates B in cancer cells, it very likely also regulates B in normal cells. In cancer cells, though, you may have more A than normal, and the hyperactivation of this pathway leads to phenotypic consequences. That is really my approach to science. I am broadly interested in gene expression control, and by understanding where it breaks, we learn how it works.

BSJ: Finally, as a member of the Innovative Genomics Institute (IGI), you are currently working on targeting RNA structural elements in SARS-CoV-2. Could you describe the project?

HG: Since the start of the pandemic, we have also wanted to contribute to the scientific effort to the extent that we could. We have a couple of projects focused on COVID, and this is one of them. Going back to the TARBP2 story, one of the key ways that we found TARBP2 was through first finding structural elements that TARBP2 binds to. I have had this long-standing interest in understanding how regulatory information is encoded not in the primary sequence, but the structures. Over the years, I have been involved in various projects to find these structural regulatory elements, and one of the projects that I was working on as a postdoc was in collaboration with Charles Rice, who recently won a Nobel

Prize. It focused on whether we found structural elements in viral genomes which had regulatory potential, and whether they were conserved.

In the initial analysis that I did, I had included coronaviruses among other families. We were never able to finish that project, but I had the understanding that came with it. I knew that coronaviruses were showing a lot of signal in terms of RNA structure, and so I decided to set up a proposal where we would start looking at the potential role of RNA secondary structure in COVID. Over the years, we have come up with this hybrid strategy of both experimental and computational probing of the secondary structure. Under this strategy, we proposed to use DMS-MaP seq to look at the secondary structure of the entire viral genome in pieces to see if we can find any docking sites or any interesting structural components called switches, where the same sequence can have multiple conformations dependent on what it is interacting with. We are currently building the library, but we will see how it goes.

REFERENCES

1. [Photograph of Hani Goodarzi]. UCSF Helen Diller Family Comprehensive Cancer Center. https://cancer.ucsf.edu/people/profiles/goodarzi_hani.7686
2. Fish, L., Zhang, S., Yu, J. X., Culbertson, B., Zhou, A. Y., Goga, A., & Goodarzi, H. (2018). Cancer cells exploit an orphan RNA to drive metastatic progression. *Nature Medicine*, 24(11), 1743–1751. <https://doi.org/10.1038/s41591-018-0230-4>
3. Fish, L., Navickas, A., Culbertson, B., Xu, Y., Nguyen, H., Zhang, S., Hochman, M., Okimoto, R., Dill, B. D., Molina, H., Najafabadi, H. S., Alarcón, C., Ruggero, D., & Goodarzi, H. (2019). Nuclear TARBP2 drives oncogenic dysregulation of RNA splicing and decay. *Molecular Cell*, 75(5), 967–981.e9. <https://doi.org/10.1016/j.molcel.2019.06.001>