AWARD NUMBER: W81XWH-18-1-0421

TITLE: Discovery and Characterization of Functional Breast Cancer Microproteins

PRINCIPAL INVESTIGATOR: Alan Saghatelilan

CONTRACTING ORGANIZATION: The Salk Institute for Biological Studies
La Jolla, CA 92037

REPORT DATE: August 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
August 2019	Annual	1 Aug 2018 - 31 Jul 2019

4. TITLE AND SUBTITLE
Discovery and Characterization of Functional Breast Cancer Microproteins

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0421

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Alan Saghatelian

E-Mail: asaghatelian@salk.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Salk Institute for Biological Studies
10010 N Torrey Pines Rd
La Jolla, CA 92037-1002

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

There were two significant findings during the research period. First, using antibodies we raised against the TINCR microprotein (TINCR-MP), we validated the translation of TINCR-MP into a stable microprotein. Though we had detected a small open reading frame (smORF) on TINCR this was the first confirmation of the TINCR microprotein. We also used this antibody to show that changes in TINCR microprotein levels readily be detected, which is necessary for our biological studies looking at the consequences of TINCR-MP on breast cancer cells. Second, we detected thousands of smORFs in MCF-7 and MD-MB-231 breast cancer cell lines, with approximately one hundred smORFs showing a 3-fold or more difference between these cell lines. Identifying these smORFs will help determine whether any of these genes contribute to the cancer phenotypes of these cell lines. These findings further validate our methodology and we are in an ideal position to complete the project in the allotted time.

15. SUBJECT TERMS
breast cancer, disease genes, small open reading frames (smORFs), microproteins, smORF oncogene, smORF tumor suppressor, CRISPR, TINCR, NuRD complex

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	9	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. **INTRODUCTION:** The goal of this proposal is to identify novel genetic drivers of breast cancer. Previous genome-wide screens for genes that drive breast cancer did not include a set of human genes that contain small open reading frames (smORFs). smORFs encode peptides and small proteins (microproteins) that are less than 100 amino acids in length. The human genome project missed smORFs because gene-finding algorithms utilized a length cutoff to try limit false positives, but in the process missed thousands of protein-coding smORFs. Our hypothesis is that there are smORFs with critical roles in breast cancer such as smORF oncogenes, tumor suppressors, and synthetic lethal genes. Here, we endeavor to identify these and characterize cancer-driving smORFs using two approaches. The first approach focuses on an RNA called TINCR that has been linked to breast cancer. TINCR is thought to be 'non-coding' but we discovered that TINCR contains a smORF that encodes an 87-amino acid microprotein. We will test whether this microprotein is responsible for the deleterious role of TINCR reported in breast cancer. Second, we will take an unbiased screening approach to identify smORFs with roles in breast cancer cell survival and invasion. This strategy begins by using our smORF-discovery platform to identify smORFs in HCC1954, HMEC, MCR-7, and MDA-231 cell lines (note: the MCF-7 and MDA-MB-231 lines were added based on a referee comments). We will then perturb these smORF-encoding genes to identify those smORFs with roles in driving the phenotypes associated with breast cancer, with the goal of finding oncogenes or tumor suppressors. Our recent discovery of thousands of novel protein-coding genes has revealed an unexpected blind spot in gene annotation methods and in this proposal we examine whether any of these new genes have roles in breast cancer.

2. **KEYWORDS:** breast cancer, disease genes, small open reading frames (smORFs), microproteins, smORF oncogene, smORF tumor suppressor, CRISPR, TINCR, NuRD complex.

3. **ACCOMPLISHMENTS:** *The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction.*

   - **What were the major goals of the project?**
   - Specific Aim 1. Breast cancer microproteins from smORFs on non-coding RNAs that have already been linked to breast cancer. (33% complete)
   - Specific Aim 2. Use Ribo-Seq to define the smORFeome of HCC1954 breast cancer cells and human mammary epithelial cells (HMECs).
   - UPDATED Specific Aim 2. Use Ribo-Seq to define the smORFome of HCC1954, HMEC, MCF-7, and MDA-MB-231 breast cancer cells. We now include MDA-MB-231 (as requested by a referee) and the MCF-7 cell line, which is the comparator for MDA-MB-231 line. (33% complete)
   - **What was accomplished under these goals?**

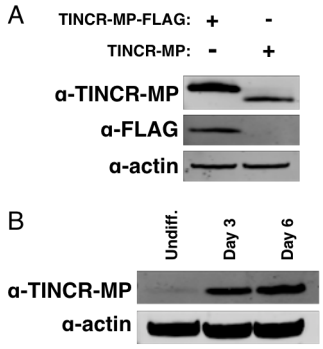

Figure 1. Validation of TINCR-MP antibodies. A) anti-TINCR-MP antibodies recognize overexpressed TINCR-MP-FLAG and TINCR-MP. B) Endogenous TINCR-MP expression in differentiating keratinocytes demonstrate that the antibody can measure TINCR-MP in a biologically relevant setting.

Specific Aim 1

Major task 1: Generation of Anti-TINCR-MP antibodies Specific (**100% complete**). The investigation of microproteins requires antibodies to accurately assess endogenous levels, or overexpression or knockdown of the microprotein. No anti-TINCR microprotein (anti-TINCR-MP) antibodies exist so we generated our own antibodies using methods developed over the last 40 years at the Salk institute. Our methods rely on monitoring antibody titer and boosting over long enough periods (6-9 months) to obtain high affinity antibodies. By contrast, commercial antibody production will only go out 3 months because this is the only way to be cost effective. As a result, we routinely generate antibodies that are far superior to any such products provided by commercial entities and this allowed us to obtain remarkably effective antibodies against TINCR. In fact, we have polyclonal antibodies against the TINCR N-terminus and a second polyclonal antibodies that recognize the TINCR C-terminus. We raised two sets of antibodies to ensure that we would obtain antibodies of sufficient quality for biological studies. We have validated these antibodies in cell lines overexpressing TINCR (Fig. 1A). Furthermore, we also tested these antibodies in keratinocyte differentiation because

TINCR expression naturally increases under these conditions, which allows these antibodies to be validated for their ability to detect changes in the endogenous expression of TINCR-MP. As expected, we see TINCR-MP expression increase during keratinocyte differentiation, validating these antibodies for use in measuring TINCR-MP in a biologically relevant situation (Fig. 1B).

Major Task 2: Identify the NuRD-binding domain within TINCR-MP by testing TINCR-MP mutants for their functional impact on the immunoprecipitation of NuRD proteins (HDAC1, RBPP7) from MDA-MB-453 cells (**20% complete**). We have prepared the TINCR-MP deletion constructs (i.e. TINCR-MP-FLAG deletion from 2-10, TINCR-MP-FLAG deletion from 11-20… TINCR-MP-FLAG deletion 80-87 (a total of 9 constructs).

Major Task 3: Knockdown of TINCR RNA and rescue experiments with wild type TINCR-MP and TINCR-MP mutants in MDA-MB-453 cells (**0% complete**).

Major Task 4: Identify microprotein-protein interactions for newly discovered smORFs from the NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs (**0% complete**).

Specific Aim 2

Major Task 1: Ribo-Seq characterization of the smORFome of HCC1954 and HMECs. **Updated Major Task 1**: Ribo-Seq characterization of the smORFome of HCC1954, HMECs, MCF-7, and MDA-MB-231 cells (**50% complete**). We have completed the Ribo-Seq and RNA-Seq of the MCF-7 and MDA-MB-231 cells. These cells grow much faster than HCC1954s, so we decided to collect this data first, and are currently collecting data form HCC1954s and HMECs. The data from these MCF-7 and MDA-MB-231 cells validated our hypothesis that there are many smORFs that are differentially regulated between cancer cells with different phenotypes (**see Figure 2**). MCF-7 and MDA-MB-231 cells are breast cancer cell lines that differ in their invasiveness—an in vitro measure of matrix migration that is thought to represent the ability of a cell to bore through the extracellular matrix during growth and metathesis. MDA-MB-231 are far more invasive than MCF-7 and the comparison of these cells in the past has identified genes that contribute to cancer growth in vivo. In our analysis, the Ribo-Seq serves as a method to identify mRNAs that contain smORFs, and then we use RNA-Seq to quantify mRNAs between samples.

In this case we quantify smORF-containing mRNAs between MCF-7 and MDA-MB-231 cell lines. We detected a total of 3,732 smORFs between the MCF-7 and MDA-MB-231 cell lines. If we focus on the "non-coding" RNAs that contain smORFs, we find a total 363 smORFs between the two cell lines, with 110 smORFs that have fold changes greater than or equal to three. If we plot those smORF mRNAs that are changing between 3-10-fold, we see a number of smORFs that are up in MDA-MB-231 cells, which could potentially ***promote*** invasiveness (Fig. 2, blue dots). Likewise, there are a similar number of smORFs that are increased in MCF-7 cells that might ***inhibit*** invasiveness (Fig. 2, red dots).

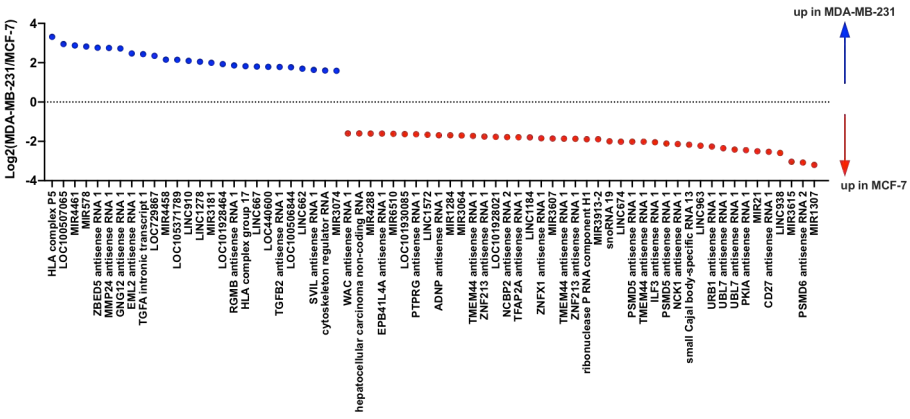

Fig. 2. Novel smORFs from non-coding RNAs that are increased 3-fold or more in MDA-MB-231 and MCF-7 cells. MDA-MB-231 cells are invasive and smORFs that are elevated in these cells (blue) may promote invasiveness. smORFs that are elevated in MCF-7 cells may suppress invasion. We will screen these smORFs for activity in invasion assays.

These smORFs are found on antisense RNAs, long non-coding RNAs (LINCs), microRNAs (MIRs). In the upcoming year (next 6 months), we will complete these Ribo-Seq and RNA-Seq experiments with HCC1954 cells and HMECs, completing our smORFome analysis of MCF-7, MDA-MB-231, HCC1954, and HMEC.

Major Task 2: Use CRISPR/Cas9 to screen for smORFs that mediate breast cancer cell proliferation by targeting smORFs that are specific to HCC1954 cells. **Updated Major Task 2:** Use CRISPR/Cas9 to screen for smORFs that mediate breast cancer cell proliferation by targeting smORFs that are specific to HCC1954 cells, and for proliferation and invasion in MDA-MB-231 cells. We now include the MDA-MB-231 cells for screening for smORFs that regulate proliferation. MDA-MB-231cells also have an invasiveness phenotype in an in vitro assay that can be used to rapidly screen for smORFs that prohibit invasion. The comparator cell line, MCF-7, is non-invasive and will be used to identify smORFs that differ between the cell lines and, therefore, might regulate invasion (**0% complete**).

- **What opportunities for training and professional development has the project provided?**

Though we had not explicitly proposed any training and professional development, Dr. Martinez has attended several conferences and spoken about his work in the area of Ribo-Seq profiling to discover smORFs. In addition, we have set up a collaboration with a mentor at UCSD, Dr. George Sen, to help with culture studies to validate our antibodies for TINCR-MP. These opportunities will behoove Dr. Martinez when as he heads out to set up his academic lab in the near future.

- **How were the results disseminated to communities of interest?**

Nothing to Report

- **What do you plan to do during the next reporting period to accomplish the goals?**

Specific Aim 1
Plans for Major Task 2 in the next reporting period are to 1) transfect MDA-MB-453 cells with the TINCR-MP deletion constructs and the expression of each the truncated TINCR-MPs using Western Blot (1 month); 2) identify the region of TINCR-MP that contains the NuRD-biding domain by performing immunoprecipitation experiments followed by Western blots against HDAC1 and RBPP7 (2 months); and 3) identify specific amino acids within TINCR-MP responsible for binding to the NuRD-biding domain by immunoprecipitation followed by Western blots against HDAC1 and RBPP7 using the alanine scanning library (2 months). Potential problems and alternative approaches: we believe TINCR is a folded microprotein and removal of any region might disrupt structure and confound interaction studies with the deletion constructs. If so, we will simply skip ahead and perform alanine scanning experiments across the entire TINCR sequence (all 87 amino acids).

Plans for Major Task 3 in the next reporting period are to 1) validate the expression of TINCR-MP in MDA-MB-453 cells with our new antibody and use CRISPR to knockdown TINCR in these cell lines and measure a loss of TINCR-MP. Then, we will 2) test whether TINCR-MP is responsible for biological activity ascribed to TINCR ncRNA, and 3) determine whether TINCR-MP activity correlates with the ability of WT or deletion mutants binding to the NuRD complex to provide a mechanistic rationale.

Plans for Major Task 4 in the next reporting period are to 1) generate constructs with N- and C-terminal tagged smORFs from NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs and 2) test for the stable expression of the resultant microproteins (3 months). For those smORFs from NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs that produce stable microproteins, we will carry out microprotein-protein interactions to determine the biochemical and potential biological functions of these smORFs.

Specific Aim 2
Plans for Major Task 1 in the next reporting period are to obtain Ribo-Seq and RNA-Seq data from HCC1954 and HMECs to identify smORFs for screening in Major Task 2.

Plans for the Major Task 2 in the next reporting period are to 1) prepare the CRISPR and overexpression plasmids to test the MCF-7 and MDA-MB-231 smORFs for activity in proliferation and invasion assays. At the end of this period we will have finished the Ribo-Seq profiling of HCC1954 and HMECs, and 2) we will begin prepare HCC1954 and HMEC smORF shRNAs to test whether any of the smORFs regulate proliferation and viability assays in HCC1594 cells (which will likely happen in the last reporting period).

4. **IMPACT:** *Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:*

    - **What was the impact on the development of the principal discipline(s) of the project?**

Though early, there are some clearly novel contributions from our work. First, we have demonstrated that TINCR-MP is a translated microprotein in cells and tissues, and that this microprotein is regulated. Since there are many reported non-coding RNAs that are biologically active, like TINCR, our discovery will open up new research into whether other non-coding RNAs are producing microproteins to mediate biology. Second, we demonstrate that our Ribo-Seq, RNA-Seq platform can be extended to any cell lines, in this case breast cancer cell lines, and that we can use this technology to detect thousands of smORFs. Third, this is the first time that anyone has quantified smORF to reveal those smORFs that correlate, and possibly contribute, to a particular phenotype, in this case breast cancer growth and invasiveness. In the next reporting period, we hope to extend this and be able to claim that our work has identified the first smORFs with a biological role in breast cancer.

    - **What was the impact on other disciplines?**

Nothing to Report

    - **What was the impact on technology transfer?**

Nothing to Report

    - **What was the impact on society beyond science and technology?**

Nothing to Report

5. **CHANGES/PROBLEMS:** *The Project Director/Principal Investigator (PD/PI) is reminded that the recipient organization is required to obtain prior written approval from the awarding agency Grants Officer whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:*

    - **Changes in approach and reasons for change**

Nothing to Report

    - **Actual or anticipated problems or delays and actions or plans to resolve them**

Nothing to Report

    - **Changes that had a significant impact on expenditures**

Nothing to Report

    - **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to Report

    - **Significant changes in use or care of human subjects**

Nothing to Report

    - **Significant changes in use or care of vertebrate animals.**

Nothing to Report

- **Significant changes in use of biohazards and/or select agents**

Nothing to Report

6. **PRODUCTS:** *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."*

   - **Publications, conference papers, and presentations**
     Report only the major publication(s) resulting from the work under this award.

     - **Journal publications.** Nothing to Report. One paper in revision.
     - **Books or other non-periodical, one-time publications.** Nothing to Report.
     - **Other publications, conference papers, and presentations.** Presented this work at several national meetings including the American Chemical Society National Meeting, and a Gordon Research Conference on Translation.

   - **Website(s) or other Internet site(s)**

Lab website: https://saghatelian.salk.edu

   - **Technologies or techniques**

The technologies we use are based on next-generation sequencing technologies and they will be disseminated in our published manuscript. In addition, we will have reagents such as antibodies and plasmids that we'll share freely with any scientist that wants to use them.

   - **Inventions, patent applications, and/or licenses**

Nothing to Report

   - **Other Products**

Our databases of smORFs will be provided in our published manuscript and we also submit all the raw data to the appropriate national databases for others to use as they wish. As mentioned, we will also have antibodies, and DNA constructs for smORF overexpression, as well as shRNA and CRISPR libraries that we will share.

7. **PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

   - **What individuals have worked on the project?**

Name:	*Alan Saghatelian*
Project Role:	*PI, No change*
Researcher Identifier (e.g. ORCID ID):	https://orcid.org/0000-0002-0427-563X
Nearest person month worked:	0.5
Contribution to Project:	*Dr. Saghatelian has overseen the project.*
Funding Support:	*Frederik Paulsen Chair/NIH/Clayton*

Name:	*Thomas Martinez*
Project Role:	*Postdoctoral Fellow*
Researcher Identifier (e.g. ORCID ID):	

Nearest person month worked:	6
Contribution to Project:	*Dr. Martinez handles all of the TINCR work and oversees all of the data analysis for the smORF discovery.*
Funding Support:	*NIH/Ferring Foundation*

Name:	*Cynthia Donaldson*
Project Role:	*Lab Coordinator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	7
Contribution to Project:	*Mrs. Donaldson handles all of the cell culture, plasmid generation, Ribo-Seq, and RNA-Seq experiments*
Funding Support:	*NIH/Clayton*

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Yes.

New/renewal: NIH 2P30CA014195-46 (Shaw), Cancer Center Support Grant, 02/01/2019-01/31/2024, $11,360 annual direct. The major goal of this project is to develop new diagnostic tools and define cancer subtypes so that clinicians will be able to select the most effective therapeutic approaches, while also identifying new molecular targets. No overlap.

New: Clayton Medical Research Foundation (Saghatelian), Novel Biologically Active Peptides and Small Proteins; Bioactive Lipids, Lipid Metabolism, and Synthetic Chemistry, $432,000 annual direct. The major goal of Project 2, Novel Biologically Active Peptides and Small Proteins, is to identify biologically essential members within newly discovered human protein-coding genes and to use that knowledge to develop new peptide and small protein therapeutics. The major goal of Project 4, Bioactive Lipids, Lipid Metabolism, and Synthetic Chemistry, is to identify new gene/protein targets (GPR40 and GPR120) for slowing multiple sclerosis and new small molecules (PAHSA analogs) that can provide the basis for novel anti-MS drugs. No overlap.

- **What other organizations were involved as partners?**

Nothing to Report

8. **SPECIAL REPORTING REQUIREMENTS**
   - **COLLABORATIVE AWARDS:** *For collaborative awards, independent reports are required from **BOTH** the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to https://ers.amedd.army.mil for each unique award.*
   - **QUAD CHARTS:** *If applicable, the Quad Chart (available on https://www.usamraa.army.mil) should be updated and submitted with attachments.*
9. **APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc. Reminder: Pages shall be consecutively numbered throughout the report. **DO NOT RENUMBER PAGES IN THE APPENDICES.***