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
10010 N Torrey Pines Rd
La Jolla, CA 92037-1002

REPORT DATE: August 2020

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
Aug 2020	Annual	1 Aug 2019 – 31 Jul 2020

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
1) We were able to show that loss of TINCR microprotein (TINCR-MP) regulates cell growth and are now screening through several breast cancer cells to determine those lines that are most sensitive to this gene. 2) We decoupled TINCR-MP biological activity from that of the TINCR mRNA. The TINCR mRNA regulates keratinocyte differentiation while the microprotein does not; however, as we demonstrated in the first finding, the TINCRP-MP has unique activity on cell growth and proliferation. 3) We determined the pathways regulated by TINCR-MP using unbiased RNA-Seq and discovered that TINCR-MP is a potent regulator of genes linked to injury and wound healing pathways. As cancer is commonly referred to as the wound that doesn't heal, this connection will help us pinpoint the particular genes under the influence of TINCR-MP that regulate breast cancer cell growth and proliferation. 4) We have completed data collection for smORFs in HCC1954, MCF-7, and MD-MB-231 breast cancer cell lines. We are already preparing HCC1954, MCF-7, and MD-MB-231 cells with stably expressed Cas9 and have designed and ordered guide RNAs for our screens. We will complete the proposed studies by the end of the funding period and are planning to publish this work next year.

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:**

   - **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. (75% 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. (75% complete)
   - **What was accomplished under these goals?**

Specific Aim 1

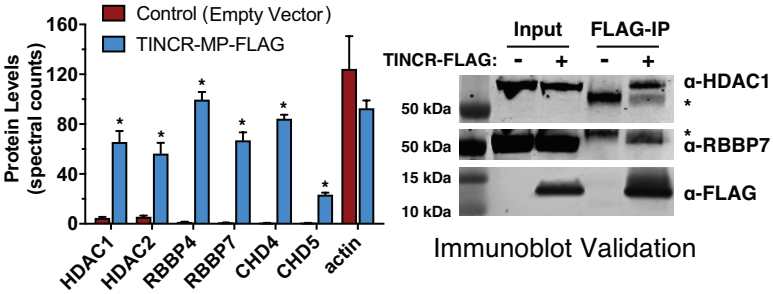

Figure 1. **(A)** Spectral Counts of NuRD complex proteins in control and TINCR-MP-FLAG IP-MS samples in HEK293T cells. *SAINT protein interaction score > 0.90 **(B)** TINCR-MP-FLAG interactions with HDAC1 and RBBP7 were validated by immunoblot. *Mouse IgG heavy chain.

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. 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 (data shown in 2019 progress report).

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 (**50% 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). We have developed a robust immunoprecipitation assay for the interaction between NuRD proteins and TINCR-MP (**Fig. 1**). Because TINCR is interacting with a protein complex involved in gene regulation, we have performed RNA-Seq experiments to determine if this TINCR can modulate gene expression. We find that TINCR regulates a series of genes that align with biological processes involved in wound healing and cancer (**Fig. 2**), consistent with our original hypothesis about this protein. Furthermore, these data support our proposed experiments as a means to delineate TINCR regulated phenotypes and biology.

Major Task 3: Knockdown of TINCR RNA and rescue experiments with wild type TINCR-MP and TINCR-MP mutants in MDA-MB-453 cells (**50% complete**). To establish whether TINCR-MP effects cancer proliferation, we measured the viability of LNCaP cells transfected with WT TINCR-MP or Mut TINCR-MP expression constructs by WST-1 assay (**Fig. 3**). TINCR-MP overexpression reduced proliferation relative to the translation impaired mutant and empty vector control transfection, consistent with the published study showing overexpression of TINCR RNA suppresses proliferation of LNCaP cells (9).

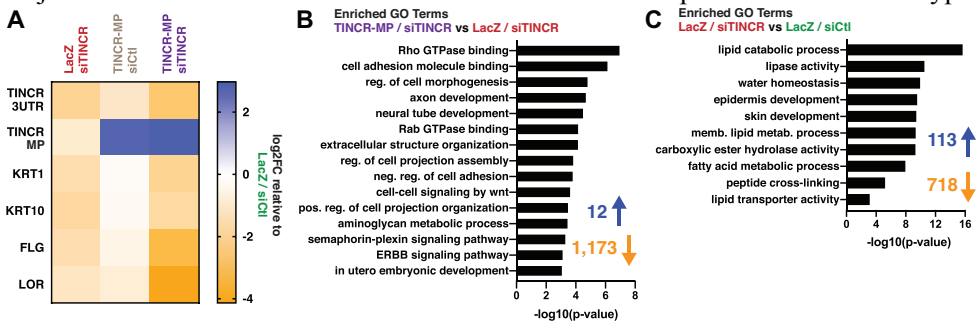

Figure 2. **(A)** Heatmap of log2 fold-changes (log2FC) in RNA expression of differentiation genes in 3D organotypic skin models relative to LacZ/siCtl control. TINCR-3UTR is representative of endogenous TINCR expression, while TINCR-MP measures change due to TINCR-MP retrovirus infection. **(B-C)** Significantly enriched gene ontology (GO) terms for TINCR-MP rescue **(B)** and TINCR-depletion **(C)** regulated genes. Number of significantly upregulated genes (blue) and downregulated genes (orange) are also shown ($|log2FC| > 1$, $p_{adj} < 0.05$). *TINCR-MP regulated genes suggest it is a regulator of wound healing processes.*

Major Task 4: Identify microprotein-protein interactions for newly discovered smORFs from the NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs (**50% complete**). We have started the immunoprecipitation experiments of these microproteins and will be complete in the next 6 months.

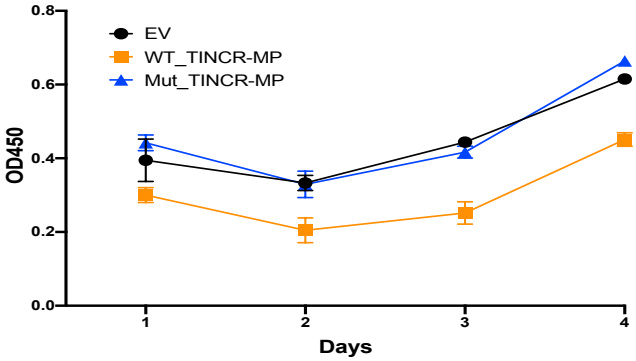

Figure 3. TINCR-MP suppresses LNCaP proliferation. Cells were transfected with empty vector (EV), WT TINCR-MP, or Mut TINCR-MP expression constructs. After 24 h, an equal number of transfected cells were seeded onto a 96-well plate (5,000 cells/well) and viability was measured at 1,2,3, and 4 days using WST-1 reagent.

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 (**100% complete**). We have completed the Ribo-Seq and RNA-Seq of the MCF-7, MDA-MB-231, and HCC1954 cells.

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 (**50% complete**). We have prepared our lentiviral guide RNA libraries for transduction of these cell lines and assays and will perform the CRISPR/Cas9 screens in the HCC1954 cells now, followed by shRNA screens in MCF-7 and MD-MBA-231 cells.

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

Though we had not explicitly proposed any training and professional development, Dr. Martinez is preparing for his independent career. He is currently awaiting a funding decision on a K01 award based on his preliminary results from TINCR and has applied for academic positions.

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

Nothing to Report as of now, but we anticipate publishing this work in 2021.

- **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) analyze the mass spectrometry data from our immunoprecipitation experiments to identify the region required for NuRD binding, 2) Use the predicted structure for TINCR (REF) (Fig. X). From this structure we can identify the amino acids on the TINCR surface in the region required for NuRD binding, and we will mutate these to amino acids to obtain key residues on TINCR structure responsible for NuRD binding. These experiments will identify TINCR mutants that are incapable of binding NuRD, and we will use these mutants to test whether the interaction between NuRD and TINCR is necessary for the biology of this project.

Plans for Major Task 3 in the next reporting period are to 1) validate for TINCR-MP presence MCF-7, MDA-MB-468, MDA-MB-231. All of the cell lines to be tested are included in the CCLE and show varying levels of TINCR RNA expression. 2) To separate TINCR-MP's role in regulating proliferation from TINCR RNA, we will exogenously express TINCR-MP in TINCR-RNA depleted cells and measure proliferation by WST-1 assay in these cells (Fig. 2). 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. These experiments will determine TINCR-MP's role in regulating proliferation and migration in breast cancer.

Plans for Major Task 4 in the next reporting period are to 1) analyze the proteomics data for NEAT1, LINC00346, ARRDC1-AS1, DLEU1, and LINC00958 ncRNAs and validate their interactions using Western blots.

Specific Aim 2
Plans for the Major Task 2 in the next reporting period are to 1) use CRISPR guide RNAs for all smORFs in HCC1954 and use a pooled CRISPR/Cas9 dropout screen to identify smORFs necessary for breast cancer cell growth and proliferation. 2) Generate shRNAs for smORFs in MCF-7 and MD-MB-231s for phenotypic invasion assays.

4. **IMPACT:**

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

We have learned that TINCR can regulate cellular proliferation (Fig. 3) and a transcriptional program related to wound healing and cancer (Fig. 2). We have also defined the smORFs in several cancer cell lines and will identify functions for these smORFs through screening to reveal new genes involved in breast cancer, and novel targets for therapeutic intervention.

- **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:**
    - **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:**
    - **Publications, conference papers, and presentations**.
        - **Journal publications.** Nothing to Report.
        - **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 will 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?**

  Because of COVID and the shift schedule at Salk we built the team up so that we could keep pace on this important project.

Name: Alan Saghatelian
Project Role: Principal Investigator
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 0.5
Contribution to project: Dr. Saghatelian oversees the project.
Funding Support: NIH, Clayton Foundation

Name: Meric Erikci Ertunc
Project Role: Postdoctoral Fellow
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 0.5
Contribution to project: Dr. Erikci Ertunc assisted with cell culture experiments.
Funding Support: NIH, Clayton Foundation

Name: Almudena Garcia Ruiz
Project Role: Postdoctoral Fellow
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 2
Contribution to project: Dr. Garcia Ruiz assisted with cell culture experiments and RNA-Seq data collection.
Funding Support: NIH, Clayton Foundation

Name: Raymond Mak
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 9
Contribution to project: Mr. Mak performed cell culture, assisted with the RiboSeq analysis, and quantified differences in smORFs.
Funding Support: No other funding.

Name: Lina Xie
Project Role: Postdoctoral Fellow
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 4
Contribution to project: Dr. Xie assisted with the design of guide RNA libraries.
Funding Support: Clayton Foundation

Name: Cynthia Donaldson
Project Role: Laboratory Coordinator
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: 2
Contribution to project: Ms. Donaldson assisted with the cell culture, plasmid generation, Ribo-Seq, and RNA-Seq experiments.
Funding Support: NIH, Clayton Foundation

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

Nothing to Report

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

Nothing to Report

8. **SPECIAL REPORTING REQUIREMENTS**
   - **COLLABORATIVE AWARDS**
   - **QUAD CHART**
9. **APPENDICES**