Submission ID #: 61564
Scriptwriter Name: Bridget Colvin
Project Page Link: https://www.jove.com/account/file-uploader?src=18777068

Title: Mapping the Structure-Function Relationships of Disordered Oncogenic Transcription Factors Using Transcriptomic Analysis

Authors and Affiliations: Iftekhar A. Showpnil[1,2], Kyle R. Miller[1], Cenny Taslim[1], Kathleen I. Pishas[1], Stephen L. Lessnick[1,3], and Emily R. Theisen[1,4]

[1]Center for Childhood Cancer and Blood Diseases, Abigail Wexner Research Institute at Nationwide Children's Hospital
[2]Molecular, Cellular, and Developmental Biology Program, The Ohio State University
[3]Division of Pediatric Hematology/Oncology/Blood & Marrow Transplant, The Ohio State University
[4]Department of Pediatrics, The Ohio State University

Corresponding Author:
Emily R. Theisen
emily.theisen@nationwidechildrens.org

Co-authors:
iftekhar.showpnil@nationwidechildrens.org
kyle.miller2@nationwidechildrens.org
cenny.taslim@nationwidechildrens.org
kathleen.pishas@nationwidechildrens.org
stephen.lessnick@nationwidechildrens.org

Author Questionnaire

1. Microscopy: Does your protocol involve video microscopy, such as filming a complex dissection or microinjection technique? **N**

2. Software: Does the part of your protocol being filmed demonstrate software usage? **Y**

3. Filming location: Will the filming need to take place in multiple locations (greater than walking distance)? **N**

Protocol Length
Number of Shots: **49**

Introduction

1. Introductory Interview Statements

REQUIRED:

1.1. **Iftekhar A. Showpnil**: Performing structure-function analysis on repetitive and disordered transcription factors is difficult. By coupling transcriptomics with the right cellular context, this approach better uncovers important structure-function relationships **[1]**.

1.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera

REQUIRED:

1.2. **Iftekhar A. Showpnil**: Using RNA-sequencing as a functional output allows the effective assessment of all of the genes regulated by a single protein in one experiment, making partial function detection more likely **[1]**.

1.2.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.3. **Emily R. Theisen**: Partial function detection is particularly important for oncogenic fusion transcription factors, as we don't know how these proteins function and their sequencing can lead to better therapies in fusion-driven cancers **[1]**.

1.3.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera

OPTIONAL:

1.4. **Cenny Taslim**: Although we will focus on the disordered EWS domain in EWS/FLI, EWS is involved in other fusions that contain disordered domains with poorly defined functions **[1]**.

1.4.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera *Videographer: Can cut for time*

Protocol

2. In Vitro Construct Panel Setup

2.1. For cDNA (C-D-N-A) expression construct transduction, first quickly thaw the frozen virus with cDNA constructs in a 37-degree Celsius water bath **[1-TXT]** and gently mix 2.5 microliters of 8 milligrams/milliliter polybrene into each vial **[2]**.

2.1.1. WIDE: Talent placing vials into water bath **TEXT: cDNA: coding DNA**

2.1.2. Talent adding polybrene to vial, with polybrene container visible in frame

2.2. Next, remove the medium from one 50-70%-confluent 10-centimeter cell culture plate per viral construct **[1]** and gently pipette the entire 2-milliliter volume of construct down the side of the plate **[2]**.

2.2.1. Medium being removed

2.2.2. Construct being added to side of plate

2.3. Rock the plate to spread the virus evenly across the cells **[1]** and place the plate in a 37-degree Celsius tissue culture incubator for 2 hours **[2]**, rocking the plate every 30 minutes to prevent any areas of the plate from drying out **[3]**.

2.3.1. Plate being rocked

2.3.2. Talent placing plate into incubator

2.3.3. Talent rocking plate

2.4. At the end of the incubation, add 5 milliliters of medium supplemented with fetal bovine serum, antibiotics, sodium pyruvate, and polybrene **[1-TXT]**.

2.4.1. Talent adding medium to plate, with medium container visible in frame **TEXT: See text for all medium preparation details**

2.5. After overnight incubation in the cell culture incubator, replace the supernatant with selection medium **[1]** and return the cells to the cell culture incubator for an additional 7-10 days to allow for selection and cDNA construct expression **[2]**.

2.5.1. Talent removing medium, with medium container visible in frame

2.5.2. Talent placing plate into cell culture incubator

3. Cell Collection

3.1. At the end of the selection period, collect the cells into a 15-milliliter conical tube for counting **[1]** and aliquot 5-10 x 10^5 cells into a new tube for RNA-sequencing and 2 x 10^6 cells into another new tube for protein extraction **[2]**.

3.1.1. WIDE: Talent adding cells to tube, with culture plate and hemocytometer visible in frame

3.1.2. Talent adding cells to second tube, with both tubes visible in frame

3.2. Sediment the cells by centrifugation **[1-TXT]** and resuspend the pellets in 1 milliliter of cold PBS for a second centrifugation **[2]**.

3.2.1. Talent adding tube(s) to centrifuge **TEXT: 5 min, 1000 x g, 4 °C**

3.2.2. Shot of pellet if visible, then PBS being added to tube, with PBS container visible in frame

3.3. Then flash freeze both pellets in liquid nitrogen **[1]** and store the cells at minus 80 degrees Celsius **[2]**.

3.3.1. Talent adding tube to LN2

3.3.2. Talent placing tube(s) at -80 °C

4. Construct Expression Validation and Correlative Phenotypic Assay Setup

4.1. To validate the knockdown of proteins of interest and the expression of the panel of constructs, blot the protein lysate samples **[1]** with the appropriate primary and secondary antibodies according to standard western blot analysis protocols **[2]**.

4.1.1. WIDE: Talent placing blot onto gel *Videographer: Important step*

4.1.2. Talent adding antibody to blot, with antibody containers visible in frame *Videographer: Important step*

4.2. To assess the RNA quality and quantity, use the lysis buffer from a silica spin-column based extraction kit **[1]** to lyse the RNA-sequencing cell samples **[2]** and apply the lysates to a genomic DNA removal column at greater than 13,000 revolutions per minute for 30-60 seconds **[3]**.

4.2.1. Talent opening kit and/or removing buffer

4.2.2. Talent adding buffer to tube, with buffer container visible in frame

4.2.3. Talent adding lysate to column, with centrifuge visible in frame

4.3. Next, proceed with silica spin-column purification and wash the RNA on the column according to the kit instructions **[2]**.

~~4.3.1. Talent adding digestion buffer to column, with digestion buffer container visible in frame *Videographer: Important step*~~

4.3.2. Talent adding wash buffer to column, with wash buffer container visible in frame *Videographer: Important step*

4.4. Then elute the RNA in 30 microliters of elution buffer **[1]** and analyze at least 2.5 micrograms of RNA on a spectrophotometer at a 260-to-280-nanometer ratio to assess the RNA quantity and sample quality **[2]**.

4.4.1. Talent adding elution buffer to tube, with elution buffer container visible in frame *Videographer: Important step*

4.4.2. Talent loading sample onto spectrophotometer *Videographer: Important step*

5. Alignment and Transcript Counting Pipeline

5.1. For FAST-Q file analysis, use PuTTY (putty) to open a terminal to the high-performance computing environment **[1]** and create an analysis directory called **project [2]**.

5.1.1. WIDE: Talent opening terminal, with monitor visible in frame

5.1.2. SCREEN: screenshot_1: 00:04-00:12

5.2. Navigate to the **path_to-project** directory and create a directory for the compressed raw fastq.gz files called **fast-q** and a second directory called **trimmed [1]**.

5.2.1. SCREEN: screenshot_1: 00:13-00:27 *Video Editor: can speed up*

5.3. Use an appropriate secure file transfer program to transfer the compressed raw fast-q.gz files from local storage to the **path_to-project-fast-q** directory **[1]** and check that there is a "R1" and an "R2" file for each sample **[2]**.

5.3.1. SCREEN: screenshot_1: 00:31-00:51 *Video Editor: please speed up*

5.3.2. SCREEN: screenshot_2: 00:05-00:24 *Video Editor: please speed up*

5.4. Navigate to path_to-project-fast-q and use the command in TrimGalore as indicated to trim the low-quality reads from the fastq.gz files **[1]**.

5.4.1. SCREEN: screenshot_2: 00:25-01:05 *Video Editor: please speed up*

5.5. Navigate to the path_to-project-directory and create a new directory called **STAR** (star)**_output [1]**.

5.5.1. SCREEN: screenshot_3: 00:01-00:09

5.6. Navigate to the path_to-project-trimmed directory and use the command as indicated to run STAR to align the trimmed fast-q.gz files **[1]**.

5.6.1. SCREEN: screenshot_3: 00:10-01:45 *Video Editor: please speed up*

5.7. Locate the required output for the next steps, which contain the counts per transcript, at the indicated location **[1-TXT]** and use the command to read in each ReadsPerGene.out.tab file **[2]**.

5.7.1. SCREEN: screenshot_4 00:04-00:54 *Video Editor: please speed up* **TEXT: path_to/project/STAR_output/sampleN_ReadsPerGene.out.tab**

5.7.2. SCREEN: screenshot_5: 00:59-01:19

5.8. For the first column, use only the characters before the period in the **Ensembl gene ID** column for the ease of downstream processing **[1]**.

5.8.1. SCREEN: screenshot_5: 01:20-02:00 *Video Editor: please speed up*

5.9. Then use the command to compile the counts from all of the samples into a dataframe called **totcts** (tote-counts) and save this new table of raw count data as a tab delimited .txt file **[2]**.

5.9.1. SCREEN: screenshot_5: 02:01-03:19

6. Differential Expression

6.1. To define the differential expression profile for each construct using DESeq2 (D-E-seek-two), input the experimental DESeq2 design **[1]** and use the **DESeq Data Set From Matrix** function to construct a DESeq DataSet, to estimate the size factors, and to run DESEq2 **[2-TXT]**.

6.1.1. WIDE: Talent inputting design, with monitor visible in frame

6.1.2. SCREEN: screenshot_6: 00:02-00:49 *Video Editor: please speed up* **TEXT: Enter column data for condition in same order as count matrix column**

6.2. To evaluate the quality of the analysis, use DESeq2 to extract the regularized log-normalized counts **[1]**.

6.2.1. SCREEN: screenshot_7: 00:02-00:34 *Video Editor: please speed up*

6.3. When extracting the results for each transcriptional profile from the DESeq2results, perform pairwise comparisons in reference to either the knockdown condition or the baseline empty vector **[1]**.

6.3.1. SCREEN: screenshot_8: 00:02-00:36

6.4. Further amend these results with the HGNC (H-G-N-C) gene symbols and extract the data from the DESeq2 data as a single file with the Ensembl gene ID, HGNC symbol, base mean expression, and differential expression data for all of the constructs with log-2 fold change and raw and adjusted p-values **[1]**.

6.4.1. SCREEN: screenshot_9: 00:05-00:59 *Video Editor: please speed up*

6.5. Assess the successful batch normalization and intra-sample similarity and use the code to use the regularized log normalized counts to check the sample clustering with principal component analysis and sample-to-sample distance plots **[1-TXT]**.

6.5.1. SCREEN: screenshot_10: 00:02-00:59 *Video Editor: please speed up* **TEXT: Control for batch-to-batch variability within samples**

6.6. Use the regularized log normalized counts to extract the 1000 most variable genes into a matrix **[1]** and use a heatmap to perform an unsupervised hierarchical clustering of the samples based on these genes **[2]**.

6.6.1. SCREEN: screenshot_11: 00:09-00:17

6.6.2. SCREEN: screenshot_11: 00:17-00:33 *Video Editor: please speed up*

6.7. To extract the clusters of interest from the dendrogram, decide at what level of the dendrogram clusters of interest appear and set **k** equal to the number of clusters at that level **[1]**.

6.7.1. SCREEN: screenshot_11: 00:34-00:51 *Video Editor: please speed up*

6.8. To determine which clusters are of interest, replot the heatmap ordered by cluster **[1]** and export the list of genes associated with each cluster in a Table **[2]**.

6.8.1. SCREEN: screenshot_11: 00:52-01:19 *Video Editor: please speed up*

6.8.2. SCREEN: screenshot_11: 01:39-01:45 *Video Editor: please speed up*

6.9. Then use an appropriate bioinformatic tool to identify the biological roles for the different clusters of genes identified and compared between the classes **[1]**.

6.9.1. SCREEN: screenshot_1: 02:33-03:02

Protocol Script Questions

A. Which steps from the protocol are the most important for viewers to see?

4.1., 4.3., 4.4.

B. What is the single most difficult aspect of this procedure and what do you do to ensure success?

6.5. Ensuring adequate controlling for batch-to-batch variability in samples is really important and can be really subtle. How to do this batch controlling is in the text about performing batch normalization, but I couldn't figure out a way to show it in the video due to the space limitations and they way I wrote out the steps of code in detail in the text. It's only entering a line of code, hitting enter, and waiting for the code to run. Checking the batch normalization with a figure is more amenable to video, so I included that.

Results

7. Results: Representative Disordered Oncogenic Transcription Factor Structure-Function Relationship Mapping

7.1. In this representative analysis **[1]**, an effective knockdown **[2]** and rescue with the positive **[3]** and negative constructs can be observed **[4]**.

7.1.1. LAB MEDIA: Figures 2B
7.1.2. LAB MEDIA: Figure 2B *Video editor: please emphasize the weaker (as compared to lane 1) FLI band in lanes 2-5 in the top blot*
7.1.3. LAB MEDIA: Figure 2B *Video Editor: please emphasize 3XFLAG-delta 22 band in middle lane of top and middle blot*
7.1.4. LAB MEDIA: Figure 2B *Video Editor: please emphasize 3XFLAG-wtEF band in lane 4 of top and middle blot*

7.2. Note that DAF (D-A-F)-rescued cells fail to form colonies **[1]**, suggesting impaired oncogenic transformation **[2]**.

7.2.1. LAB MEDIA: Figure 2E *Video Editor: please emphasize DAF data bar*
7.2.2. LAB MEDIA: Figure 2E

7.3. Following completion of the replicate validation, phenotypic assays, and initial RNA-sequence data processing, gene counts can be obtained for all of the samples for batch normalization and analysis **[1]**.

7.3.1. LAB MEDIA: Figure 3

7.4. DESeq2 without batch normalization can result in confounding batch effects **[1]**, likely due to biological variability introduced by the passage of cells in culture and differences in the processing of each batch **[2]**.

7.4.1. LAB MEDIA: Figure 4A *Video Editor: please add/emphasize red arrows*
7.4.2. LAB MEDIA: Figure 4A

7.5. Following batch normalization, DESeq2 can be used to generate transcriptional profiles for the constructs of interest relative to the baseline **[1]**.

7.5.1. LAB MEDIA: Figure 4B

7.6. Principal component analysis for these data **[1]** suggests that the transcriptional profile of DAF is intermediate **[2]** between wild-type EWS/FL1 (E-W-S-fly) **[3-TXT]** and delta 22, confirming partial function **[4]**.

7.6.1. LAB MEDIA: Figure 5A
7.6.2. LAB MEDIA: Figures 5B *Video Editor: please emphasize DAF graph*
7.6.3. LAB MEDIA: Figures 5B *Video Editor: please emphasize WT-EF graph* **TEXT: EWS/FL1: Ewing Sarcoma/Friend leukemia integration 1**
7.6.4. LAB MEDIA: Figures 5B *Video Editor: please emphasize delta22 graph*

7.7. Moreover, hierarchical clustering of the 1000 most variable genes across samples **[1]** shows that DAF fails to repress EWS/FL1 target genes **[2]** and only partially retains gene activation activity **[3]**.

7.7.1. LAB MEDIA: Figure 6A
7.7.2. LAB MEDIA: Figure 6A *Video Editor: please emphasize red bars of iEF lanes*
7.7.3. LAB MEDIA: Figure 6A *Video Editor: please emphasize blue bars of iEF lanes*

7.8. ToppGene analysis **[1]** suggests that the classes of genes that DAF activates are functionally distinct **[2]** from those EWS/FL1-activated targets where DAF is non-functional **[3]**.

7.8.1. LAB MEDIA: Figure 6B
7.8.2. LAB MEDIA: Figure 6B *Video Editor: please emphasize right graph*
7.8.3. LAB MEDIA: Figure 6B *Video Editor: please emphasize left graph*

7.9. Interestingly, DAF is most able to rescue **[1]** GGAA (G-G-A-A)-microsatellite activated genes **[2]**, but unable to rescue activated genes near a high affinity site **[3]**.

7.9.1. LAB MEDIA: Figures 7B and 7C
7.9.2. LAB MEDIA: Figures 7B and 7C *Video Editor: please emphasize GGAA DAF graphs*
7.9.3. LAB MEDIA: Figures 7B and 7C *Video Editor: please emphasize High affinity DAF graphs*

Conclusion

8. Conclusion Interview Statements

8.1. **Emily R. Theisen**: Pairing the transcriptomic output with the relevant phenotypic assays completes the structure-function analysis. Researchers can also use other techniques to study the mechanistic drivers of different transcriptional functions **[1]**.

8.1.1. INTERVIEW: Named talent says the statement above in an interview-style shot, looking slightly off-camera