Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61564R1
Full Title:	Mapping the Structure-Function Relationships of Disordered Oncogenic Transcription Factors Using Transcriptomic Analysis
Section/Category:	JoVE Cancer Research
Keywords:	Transcriptomics, RNA-seq, Structure-function, Transcription factors, intrinsically disordered domains, EWS/FLI, Ewing sarcoma, gene regulation
Corresponding Author:	Emily R Theisen Research Institute at Nationwide Childrens Hospital Columbus, OH UNITED STATES
Corresponding Author's Institution:	Research Institute at Nationwide Childrens Hospital
Corresponding Author E-Mail:	emily.theisen@nationwidechildrens.org
Order of Authors:	Iftekhar A Showpnil
	Kyle R Miller
	Cenny Taslim
	Kathleen I Pishas
	Stephen L Lessnick
	Emily R Theisen
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Columbus, OH, USA

Center for Childhood Cancer and Blood Diseases
The Abigail Wexner Research Institute
Nationwide Children's Hospital
700 Children's Drive
Columbus, OH 43205

15 April 2020

Dear Editor,

It is my pleasure to submit our manuscript, 'Mappin the Structure-Function Relationships of Disordered Oncogenic Transcription Factors Using Transcriptomic Analysis,' for consideration in the *Journal of Visualized Experiments*. This study was performed at the Abigail Wexner Research Institute at Nationwide Children's Hospital with individual contributions from, Iftekhar Showpnil, Kyle Miller, Cenny Taslim, Kathleen Pishas, Stephen Lessnick, and Emily Theisen.

Many cancers with are driven by a pathognomonic gene fusion, the expression of which produces and oncogenic transcription factor. These are particularly important proteins in malignancies of children and young adults. While many such fusions have been identified to date, and while they present attractive therapeutic targets, much of their function remains poorly understood. A deeper basic understanding of the biochemical mechanisms used by the proteins to dysregulate gene expression is important to better understand the diseases they cause and how to treat them.

Many fusions incorporate an intrinsically disordered domain, such as the N-terminal portion of the EWSR1 protein. One of the best studied EWS fusion proteins is EWS/FLI. This protein both activates and represses thousands of genes to cause Ewing sarcoma and the EWS domain plays critical roles in recruitment of co-regulators and DNA-binding. Efforts to understand which structural features of the disordered EWS domain contribute to different functions of EWS/FLI (i.e. activation vs. repression, co-regulator recruitment vs. DNA-binding) have been hindered by the repetitive nature of the domain. Efforts using deletion mapping have failed to find any distinct structural elements, and have primarily focused on targeted reporter assays looking at single genes. Here we, present a new approach to address the functional contributions of different EWS structural features to EWS/FLI function using transcriptomic mapping. This approach successfully identified a mutant of EWS/FLI, called "DAF," that represents the first report of a partially functional mutant (Theisen, et al. Genes & Cancer, 2019) and we use some of that data as an example in this report of the method.

This study establishes a new paradigm for understanding the structure-function relationships of the oncogenic fusion transcription factors that drive mutationally quiet cancers. Only by employing genomic methods with a unique set of mutant constructs were we able to fully capture the effects of distinct structural alterations to the EWS domain on the multi-faceted transcriptional activity of EWS/FLI. Studies illuminating new complexity in EWS/FLI-mediated transcriptional regulation genome-wide are of significant interest to the field, and this approach is broadly applicable to other oncogenic transcription factors. We are confident this study will be of great interest as other researchers work to uncover the function of other fusions that are less characterized that EWS/FLI, such as PAX3-FOXO1.

There are two conflicts-of-interest to disclose: Stephen Lessnick is on the scientific advisory board for Salarius Pharmaceuticals; currently undertaking clinical development of LSD1 inhibitors for Ewing sarcoma. Stephen Lessnick is also listed as an inventor on United States Patent No. US 7,939,253 B2, "Methods and compositions for the diagnosis and treatment of Ewing's Sarcoma" and United States Patent No. US 8,557,532, "Diagnosis and treatment of drug-resistant Ewing's sarcoma." These are not affected by the data reported in this report and not impact our adherence to *JoVE* policies on sharing data and materials.

We are excited about submitting out manuscript to *JoVE* and we look forward to hearing from you at your earliest convenience.

Sincerely,
Emily R. Theisen, PhD

1 **TITLE:**
2 Mapping the Structure-Function Relationships of Disordered Oncogenic Transcription Factors
3 Using Transcriptomic Analysis
4

5 **AUTHORS AND AFFILIATIONS:**
6 Iftekhar A. Showpnil[1,2], Kyle R. Miller[1], Cenny Taslim[1], Kathleen I. Pishas[1], Stephen L. Lessnick[1,3],
7 Emily R. Theisen[1,4]
8

9 [1]Center for Childhood Cancer and Blood Diseases, Abigail Wexner Research Institute at
10 Nationwide Children's Hospital, Columbus, OH, USA
11 [2]Molecular, Cellular, and Developmental Biology Program, The Ohio State University, Columbus,
12 OH, USA
13 [3]Division of Pediatric Hematology/Oncology/Blood & Marrow Transplant, The Ohio State
14 University, Columbus, OH, USA
15 [4]Department of Pediatrics, The Ohio State University, Columbus, OH, USA
16

17 Corresponding Author:
18 Emily R. Theisen   (emily.theisen@nationwidechildrens.org)
19

20 E-mail addresses of co-authors:
21 Iftekhar A. Showpnil   (iftekhar.showpnil@nationwidechildrens.org)
22 Kyle R. Miller   (kyle.miller2@nationwidechildrens.org)
23 Cenny Taslim   (cenny.taslim@nationwidechildrens.org)
24 Kathleen I. Pishas   (kathleen.pishas@nationwidechildrens.org)
25 Stephen L. Lessnick   (stephen.lessnick@nationwidechildrens.org)
26

27 **KEYWORDS:**
28 transcriptomics, RNA-seq, structure-function, transcription factors, intrinsically disordered
29 domains, EWS/FLI, Ewing sarcoma, gene regulation
30

31 **SUMMARY:**
32 Intrinsically disordered domains are important for oncogenic fusion transcription factor function.
33 To therapeutically target these proteins, a more detailed understanding of the regulatory
34 mechanisms employed by these domains is required. Here, we use transcriptomics to map
35 important structural features of the intrinsically disordered EWS domain in Ewing sarcoma.
36

37 **ABSTRACT:**
38 Many cancers are characterized by chromosomal translocations which result in the expression of
39 oncogenic fusion transcription factors. Typically, these proteins contain an intrinsically
40 disordered domain (IDD) fused with the DNA-binding domain (DBD) of another protein and
41 orchestrate widespread transcriptional changes to promote malignancy. These fusions are often
42 the sole recurring genomic aberration in the cancers they cause, making them attractive
43 therapeutic targets. However, targeting oncogenic transcription factors requires a better
44 understanding of the mechanistic role that low-complexity, IDDs play in their function. The N-

45　terminal domain of EWSR1 is an IDD involved in a variety of oncogenic fusion transcription
46　factors, including EWS/FLI, EWS/ATF, and EWS/WT1. Here, we use RNA-sequencing to investigate
47　the structural features of the EWS domain important for transcriptional function of EWS/FLI in
48　Ewing sarcoma. First shRNA-mediated depletion of the endogenous fusion from Ewing sarcoma
49　cells paired with ectopic expression of a variety of EWS-mutant constructs is performed. Then
50　RNA-sequencing is used to analyze the transcriptomes of cells expressing these constructs to
51　characterize the functional deficits associated with mutations in the EWS domain. By integrating
52　the transcriptomic analyses with previously published information about EWS/FLI DNA binding
53　motifs, and genomic localization, as well as functional assays for transforming ability, we were
54　able to identify structural features of EWS/FLI important for oncogenesis and define a novel set
55　of EWS/FLI target genes critical for Ewing sarcoma. This paper demonstrates the use of RNA-
56　sequencing as a method to map the structure-function relationship of the intrinsically disordered
57　domain of oncogenic transcription factors.

58

59　**INTRODUCTION:**
60　A subset of cancers, including many malignancies of childhood and adolescence, are
61　characterized by chromosomal translocations which generate novel fusion oncogenes[1–6]. The
62　resulting fusion proteins frequently function as oncogenic transcription factors, orchestrating
63　widespread changes in transcriptional regulation to promote tumorigenesis[7,8]. Cancers with
64　these translocations commonly possess an otherwise quiet mutational landscape, with few
65　recurring genomic aberrations aside from the pathognomonic fusion[4,9]. As such, directly
66　targeting the fusion protein is an attractive therapeutic strategy in these diseases. However,
67　these oncogenic transcription factors commonly consist of a low-complexity, intrinsically
68　disordered, transcriptionally activating domain fused with a DNA-binding domain (DBD)[10–14].
69　Both the intrinsically disordered domains (IDDs) and DBDs of these proteins have proven difficult
70　to target with conventional pharmacological approaches. Development of novel therapeutic
71　approaches, therefore, requires a more detailed molecular understanding of the mechanisms
72　employed by these fusions to aberrantly regulate gene expression.

73

74　The N-terminal IDD portion of EWSR1 is commonly fused to a DBD in cancer, including EWS/FLI
75　in Ewing sarcoma, EWS/WT1 in diffuse small round cell tumor, and EWS/ATF1 in clear cell
76　sarcoma of soft parts[10]. The mechanistic role of the EWS IDD in each of these fusions is
77　incompletely understood. The EWS/ETS family of fusions, specifically EWS/FLI, is the most
78　functionally characterized to date. EWS/FLI coordinates genome-wide epigenetic and
79　transcriptional changes leading to the activation and repression of thousands of genes[7,11,15,16].
80　Studies have shown that the IDD is important for the recruitment of both transcriptional co-
81　activators (such as p300, WDR5, and the BAF complex), as well as co-repressors (such as the NuRD
82　complex)[11,15,17]. The fusion of the EWS IDD to the C-terminal portion of FLI1 confers novel DNA-
83　binding specificity to the ETS DBD of FLI1, such that the fusion oncoprotein (EWS/FLI) binds to
84　repetitive GGAA-microsatellite regions of the genome in addition to the consensus ETS motif[18–
85　20]. Combined with the co-activator recruitment function, this emergent DNA-binding activity of
86　EWS/FLI promotes de novo enhancer formation at GGAA-microsatellites distal to transcription
87　start sites (TSS) ("enhancer-like" microsatellites) and recruits RNA polymerase II to promote
88　transcription at GGAA-microsatellites proximal to TSS ("promoter-like" microsatellites)[11,15,16,21].

89

90   Taken together, these data led us to hypothesize that discrete elements within the EWS domain
91   contribute to the recruitment of distinct co-regulators to different types of EWS/FLI binding sites.
92   However, discerning these elements within the EWS portion of EWS/FLI, and how they function,
93   has been hindered by the highly repetitive and disordered nature of the domain. Here we utilize
94   a previously published knockdown-rescue system in Ewing sarcoma cells to functionally map
95   these elements in the EWS IDD. In this system EWS/FLI is depleted using an shRNA targeting the
96   3'UTR of the *FLI1* gene, and expression is rescued with varying EWS/FLI mutant cDNA constructs
97   lacking the 3'UTR[7,17,22]. These experiments focused on constructs with various deletions to map
98   the structure-function relationship between the EWS IDD and important oncogenic phenotypes,
99   including activation of a GGAA-microsatellite reporter construct, colony formation assays, and
100   targeted validation of EWS/FLI-activated and -repressed genes[7,17,22]. However, these studies
101   failed to find discrete sub-domains within the EWS IDD in EWS/FLI that are uniquely important
102   for either activation or repression. All tested constructs were either able to both activate and
103   repress specific target genes, leading to efficient colony formation, or unable to regulate any of
104   the EWS/FLI target genes, leading to loss of colony formation[7,17,22].

105

106   Transcriptomic analyses enabled by the widespread adoption of the next generation sequencing
107   are commonly used to compare gene expression signatures in two conditions, frequently in the
108   context of screening or descriptive studies. We instead wanted to leverage the ability to capture
109   genome-wide expression data using RNA-sequencing (RNA-seq) to characterize the contributions
110   of IDDs to transcription factor function. In this case RNA-seq is paired with the knockdown-rescue
111   system to explore the structure-function relationship of the EWS domain. This approach is
112   applicable to other fusion transcription factors, including other EWS fusions or wildtype
113   transcription factors with poorly understood function, and has multiple advantages over the
114   other assays used for functional mapping studies, such as reporter assays or targeted qRT-PCR.
115   These include testing structural determinants of function in the relevant chromatin context, the
116   ability to test multiple types of response elements in one assay (i.e., activated and repressed,
117   GGAA-microsatellite and non-microsatellite, etc.), and the resulting ability to better detect partial
118   function.

119

120   Successful implementation of this approach depends on a cell-based system that captures the
121   phenotypes of interest (in this case A673 cells with shRNA-mediated EWS/FLI depletion), and a
122   panel of mutant constructs in an expression vector appropriate for the cell-based system (in this
123   case, pMSCV-hygro with various 3x-FLAG-tagged EWS/FLI mutants to be delivered by retroviral
124   transduction). Viral transduction of either CRISPR-based depletion constructs, shRNA-based
125   depletion constructs, and cDNA expression constructs with appropriate selection to generate
126   stable cell lines is recommended over transient transfection. The downstream interpretation of
127   results is strengthened when the transcriptomic data can be paired with other data related to
128   localization of the transcription factor and other phenotypic readouts where available.

129

130   In this paper, we apply this approach to characterize the activity of the DAF mutant of EWS/FLI[14].
131   The DAF mutant has 17 tyrosine to alanine mutations in the repetitive regions of the EWS IDD of
132   EWS/FLI[14]. This particular EWS mutant had been previously reported and is unable to activate

133   reporter gene expression when fused to the ATF1 DBD[14]. However, preliminary qRT-PCR data
134   suggested that this mutant was able to activate transcription of the EWS/FLI target *NR0B1*[23]. The
135   transcriptomic approach described here enabled successful detection of partial function of the
136   DAF mutant. By pairing these transcriptomic data with information about EWS/FLI binding and
137   recognition motifs we further show that the DAF mutant retains function at GGAA-microsatellite
138   repeats. These results identify DAF as the first partially functional EWS/FLI mutant and highlight
139   function at non-microsatellite genes as important for oncogenesis (as reported[23]). This
140   demonstrates the power of this transcriptomic structure-function mapping approach to provide
141   insight into the function of oncogenic transcription factors.

142

143   **PROTOCOL:**

144

145   **1. Set up in vitro panel of constructs**

146

147   NOTE: This step will vary depending on the specific protein to be analyzed.

148

149   1.1. Prepare aliquots of virus for depletion and expression constructs as necessary.

150

151   1.1.1. Seed a 10 cm tissue culture dish with 3-5 x 10^6 HEK293-EBNA or HEK293T cells for each
152   construct needed for viral transduction. Let cells adhere overnight in Dulbecco's Modified Eagle
153   Media   (DMEM)   supplemented   with   10%   fetal   bovine   serum   (FBS),
154   penicillin/streptomycin/glutamine (P/S/Q), and 0.3 mg/mL G418.

155

156   NOTE: HEK293-EBNA and HEK293T cells are recommended for viral production because they are
157   easy to grow, have high transfection efficiency, and efficiently express recombinant proteins
158   from episomal plasmids. The cells should be between 50-70% confluent the day of transfection.

159

160   1.1.2. Prepare a transfection mix for each viral transduction construct. Combine 2 mL of reduced
161   serum media with 90 µL of transfection reagent.

162

163   NOTE: Pre-warming reduced serum media is recommended.

164

165   1.1.3. Add 10 µg each of a viral packaging plasmid (e.g., gag-pol), viral envelope plasmid (e.g.,
166   VSV-G), and one of either CRISPR-based depletion, shRNA-based depletion, or cDNA expression
167   construct (e.g., pMKO or pMSCV) to the transfection mix. Mix well by gentle pipetting.

168

169   1.1.4. Let the transfection mix sit for 20 min at room temperature. Remove HEK293-EBNA growth
170   media from tissue culture dishes and add 3 mL DMEM supplemented with 10% FBS, P/S/Q, and
171   10 mM sodium pyruvate. To each dish, add 2 mL of transfection mix dropwise. Let cells sit in
172   transfection media overnight in an incubator at 37 ˚C and 5% CO_2.

173

174   1.1.5. The following morning add 20 mL of DMEM media with 10% FBS, P/S/Q supplementation,
175   and 10 mM sodium pyruvate. Incubate the cells in it at 37 ˚C and 5% CO_2 for overnight.

176

177   1.1.6. The next morning, replace media with 5 mL viral collection media (VCM) (DMEM
178   supplemented with 10% heat inactivated FBS, P/S/Q, and 20 mM HEPES).

179
180   1.1.7. After 4 h, collect VCM from plates and store in a 50 mL conical tube on ice at 4 °C. Replace
181   with 5 mL of fresh VCM.

182
183   1.1.8. After 4 h, collect VCM from plates in same 50 mL conical tube and store on ice at 4 °C.
184   Replace with 8 mL of fresh VCM for overnight collection.

185
186   1.1.9. In the morning collect VCM from plates and store in the 50 mL conical tube on ice at 4 °C.
187   Replace with 5 mL of fresh VCM.

188
189   1.1.10. After 4 h, collect VCM from plates and store in the 50 mL conical tube on ice at 4°C.
190   Replace with 5 mL of fresh VCM. After 4 h, collect VCM from plates and add to the 50 mL conical
191   tube.

192
193   1.1.11. Aliquot collections from 50 mL tube into cryotubes (2 mL per aliquot) after filteration
194   through a 0.45 µm filter. Store viral aliquots at -80 °C until use.

195
196   NOTE: The protocol can be paused here, and the viral aliquots can be stored until ready for use.

197
198   1.2. Seed cells at the appropriate density in a 10 cm tissue culture dish. Target 50% confluence.
199   Let cells adhere overnight by placing in the incubator at 37 ˚C containing 5% CO_2.

200
201   NOTE: For A673 cells this is 5 x 10^6 cells in 10 mL of DMEM media with 10% FBS, P/S/Q
202   supplementation, and 10 mM sodium pyruvate. These conditions may vary depending on the
203   growth rate of the cells used.

204
205   1.3. Deplete endogenous factor of interest. If cells do not need to have the endogenous protein
206   of interest depleted, skip ahead to step 1.4.

207
208   1.3.1. Thaw viral aliquot for transduction of shRNA or CRISPR construct targeting the protein of
209   interest. Thaw frozen aliquots quickly in a 37 °C water bath.

210
211   1.3.2. Add 2.5 µL of 8 mg/mL polybrene to each viral aliquot and mix by gentle pipetting. Remove
212   media from plates of cells and gently add viral aliquot to 10 cm plate by pipetting along the side
213   of the plate. Rock the plate to spread the 2 mL of viral aliquot.

214
215   1.3.3. Incubate at 37 °C in the tissue culture incubator for 2 h. Rock the plate every 30 min to
216   prevent any areas of the plate drying out.

217
218   1.3.4. Add 5 mL of DMEM media with 10% FBS, P/S/Q supplementation, and 10 mM sodium
219   pyruvate, with 5 µL of 8 mg/mL polybrene. Let cells incubate overnight.

220

221 1.3.5. In the morning remove media from cells and passage cells into media supplemented with
222 a selection reagent. When passaging cells, seed them in a manner to allow them to grow for 48-
223 72 h and reach 50% confluency.

225 NOTE: For A673 cells with pSRP-iEF-2, cells are seeded in a 1:5 split and selected for 72 h with 2
226 μg/mL puromycin.

228 1.4. Transduce cDNA expression constructs.

230 1.4.1. Check cells to confirm 50-70% confluency.

232 1.4.2. Thaw viral aliquot(s) for transduction of cDNA construct(s) of interest. Thaw frozen aliquots
233 quickly in a 37 °C water bath. Add 2.5 μL of 8 mg/mL polybrene to each viral aliquot and mix by
234 gently pipetting.

236 1.4.3. Remove media from plated cells and gently add viral aliquot to 10 cm plate by pipetting
237 along the side of the plate. Rock the plate to spread the 2 mL of viral aliquot.

239 1.4.4. Incubate at 37 °C in the tissue culture incubator for 2 h. Rock the plate every 30 min to
240 prevent any areas of the plate drying out.

242 1.4.5. Add 5 mL of DMEM media with 10% FBS, P/S/Q supplementation, and 10 mM sodium
243 pyruvate, with 5 μL of 8 mg/mL polybrene. Let cells incubate overnight.

245 1.4.6. In the morning remove media from cells and passage cells into double selection media.
246 Grow and passage cells as needed for 7-10 days to allow for double selection and expression of
247 the cDNA construct.

249 NOTE: This split of this passage may require optimization for different cell lines. For A673 cells
250 with pSRP-iEF-2 and a pMSCV-hygro construct, cells are passed without splitting into 2 μg/mL
251 puromycin and 100 μg/mL hygromycin.

253 **2. Collect cells, validate expression of constructs, and set up correlative phenotypic assays**

255 2.1.   After 7-10 days of double selection collect cells in a 15 mL conical tube. Count collected
256 cells with a hemocytometer. Aliquot collected cells for RNA-sequencing and to validate
257 expression of cDNA constructs.

259 NOTE: Set up any correlative phenotypic assays required by the research question under
260 investigation. Colony forming assays are an example of a correlative phenotypic assay that are
261 used here.

263 2.1.1.   Collect between 5 x 10^5 and 1 x 10^6 cells for RNA-sequencing and 2 x 10^6 cells for protein
264 extraction. Pellet cells by centrifugation at 1,000 $x\,g$ at 4 °C for 5 min and remove the supernatant.

265

266   2.1.2.   Wash the pellet with 1 mL cold PBS. Pellet by centrifugation at 1,000 $x\,g$ at 4 °C for 5 min
267   and remove supernatant. Flash freeze pellets in liquid nitrogen and store at -80 °C.

268

269   2.1.3.   Set up any correlative assays with the remaining cells.

270

271   NOTE: The protocol can be paused here with collected samples stored in the -80 °C freezer.

272

273   2.2.   Validate the knockdown of protein of interest (if used) and expression of the panel of
274   constructs.

275

276   2.2.1.   Thaw cell pellets for protein extraction on ice. Resuspend cells in ice cold 500 µL nuclear
277   extraction buffer (20 mM HEPES pH 7.9, 140 mM NaCl, 10% glycerol, 1.5 mM $MgCl_2$, 1 mM EDTA,
278   1 mM DTT, 1% IGEPAL) with protease inhibitor. Let it sit for 5 min on ice.

279

280   2.2.2.   Pellet nuclei by centrifugation at 1,000 $x\,g$ at 4 °C for 5 min and remove supernatant.
281   Wash nuclei in 500 µL ice cold nuclear extraction buffer (20 mM HEPES pH 7.9, 140 mM NaCl,
282   10% glycerol, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1% IGEPAL) with protease inhibitor.

283

284   2.2.3.   Pellet nuclei by centrifugation at 1,000 $x\,g$ at 4 °C for 5 min and remove the supernatant.
285   Resuspend nuclei in 200 µL cold RIPA buffer with protease inhibitor (adjust the volume of RIPA
286   buffer according to pellet size.) Let it sit on ice for 45-60 min with vigorous vortexing every 15
287   min.

288

289   2.2.4.   Pellet cell debris by centrifugation at 16,000 $x\,g$ at 4 °C for 45-60 min. Keep the
290   supernatant and transfer to a fresh cold tube

291

292   2.2.5.   Prepare samples for SDS-PAGE electrophoresis by boiling 5-10 µg of protein with 1x
293   loading buffer for 5 min. Run an SDS-PAGE gel as required for the protein of interest.

294

295   2.2.6.   Transfer to a nitrocellulose or PVDF membrane as needed for the protein of interest.
296   Block, and blot with the appropriate primary and secondary antibodies to confirm the knockdown
297   of the endogenous protein (if used) and ectopic expression of the cDNA construct.

298

299   NOTE: The protocol can be paused here.

300

301   2.3.   Extract RNA. Assess RNA quality and quantity.

302

303   2.3.1.   Thaw cell pellets on ice. Extract total RNA using a silica spin-column based extraction kit
304   according to the manufacturer's instructions.

305

306   2.3.2.   Briefly, lyse the cells using the lysis buffer from the kit. Either apply the lysate to a silica
307   spin-column with a brief spin at >13000 rpm for 30-60 seconds or remove gDNA by applying the
308   lysate to a gDNA removal column with a brief spin at >13000 rpm for 30-60 seconds.

309

310

311 2.3.3.   Perform an on-column DNA digestion if lysate was directly applied to a silica spin-column.
312 If using a gDNA removal column, apply the eluate to a silica-spin column with a brief spin at
313 >13000 rpm for 30-60 s.

314

315 2.3.4.   Wash RNA on the column per the manufacturer's instructions. Elute RNA in 30 µL of
316 elution buffer.

317

318 2.3.5.   Assess RNA quality and quantity using a fluorometer, or any other comparable
319 instrument. Make sure the 260/280 ratio is close to 2 and that there are at least 2.5 µg of RNA to
320 submit for sequencing.

321

322 NOTE: As replicates are gathered, each replicate must be processed with the same RNA
323 extraction protocol.

324

325 2.3.6.   Use a small aliquot of RNA to confirm the stable knockdown of the protein of interest, if
326 required, by qRT-PCR. Store the remaining RNA sample at -80 °C.

327

328 2.3.7.   Collect biological replicates by repeating steps 1-2 until 3-4 complete sets of RNA have
329 been collected. Ensure that each replicate displays adequate expression of cDNA constructs and
330 stable knockdown of the endogenous protein (if used).

331

332 **3.   Next-Generation Sequencing**

333

334 3.1.   Submit extracted RNA to be sequenced using a next generation sequencing platform with
335 a target of 50 million 150 base pair (bp) paired end reads. Follow the instructions of the facility
336 processing the samples. Select for poly-adenylated RNAs and strand-specific sequencing.

337

338 **4.   Alignment and Transcript Counting Pipeline**

339

340 NOTE: This protocol assumes that following sample submission and processing, a set of paired
341 FASTQ files are returned for each sample. These files are frequently compressed with a suffix of
342 "fastq.gz". Further analysis of these FASTQ files will require access to a high-performance
343 computing (HPC) facility running a Linux operating system.

344

345 4.1.   Transfer files

346

347 4.1.1.   Open a terminal to the HPC environment with PuTTY. Make a directory for the analysis
348 called "project".

349

350 4.1.2.   Navigate to the "path_to/project" directory and make a new directory for the
351 compressed raw fastq.gz files called "fastq". Also make a directory called "trimmed". This is
352 shown in **Figure S1A-C**.

353

354 4.1.3. Transfer the compressed raw fastq.gz files from local storage to the
355 "path_to/project/fastq/" directory using WinSCP or a similar program. Check that there is a "R1"
356 and an "R2" file for each sample as shown in **Figure S1B**.

357

358 4.1.4. Optional: If required, install TrimGalore. Set the directory containing the trim_galore
359 executable file in the PATH environment variable in Linux.

360

361 NOTE: Low quality reads and adapters are trimmed with TrimGalore. TrimGalore is available at
362 https://github.com/FelixKrueger/TrimGalore.

363

364 4.1.5. Optional: Navigate to the directory for downloaded software packages (i.e
365 "path_to/software"). Download the latest TrimGalore package using the command "curl -fsSL
366 https://github.com/FelixKrueger/TrimGalore/archive/[version].tar.gz   -o   trim_galore-
367 [version].tar.gz."

368

369 4.1.6. Optional: Unpack the tar.gz file. Use the command "tar -xvzf trim_galore-
370 [version_number].tar.gz".

371

372 4.1.7. Optional: Make TrimGalore executable. Use the command "chmod a+x
373 path_to/software/TrimGalore-[version]/trim_galore". Make sure this new directory is in the
374 PATH. Use the command "export PATH=path_to/software/TrimGalore-[version]:$PATH".

375

376 4.1.8. Navigate to path_to/project/fastq/. Use TrimGalore to trim the low quality reads from
377 the fastq.gz files using the command shown in **Figure S1C**.

378

379 NOTE: Additional flags for this command may be relevant and can be found here:
380 https://github.com/FelixKrueger/TrimGalore/blob/master/Docs/Trim_Galore_User_Guide.md

381

382 4.1.9. Check for the trimmed fastq.gz files in the path_to/project/trimmed directory. Ensure
383 they are called sample1_R1_val_1.fq.gz and sample1_R2_val_2.fq.gz

384

385 4.2.   Align trimmed FASTQ files with STAR and generate transcript counts.

386

387 NOTE: STAR is available at https://github.com/alexdobin/STAR)

388

389 4.2.1. Optional: Install STAR version 2.6 or later. Set the STAR executable in the path.

390

391 4.2.2. Optional: Navigate to the directory for downloaded software packages (i.e
392 "path_to/software").

393

394 4.2.3. Optional: Download the STAR package using the command "curl -SLO
395 https://github.com/alexdobin/STAR/archive/[version].tar.gz". Unpack the tar.gz file.

396

397   4.2.4. Optional: Use the command "tar -xzf [version].tar.gz". Make STAR executable. Use the
398   command "chmod a+x path_to/software/STAR-[version]/bin".

399

400   4.2.5. Optional: Make sure this new directory is in the path. Use the command "export
401   PATH=path_to/software/STAR-[version_number]/bin/linux_x86_64_static:$PATH".

402

403   NOTE: The   STAR   manual   is   available   at:
404   (https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf).

405

406   4.2.6. Ensure there is genome index to use with STAR. Place this in a directory separate from
407   the path_to/project/ directory. If an index was previously generated for prior experiments use
408   that.   Alternatively   use   an   appropriate   pre-generated   index   if   available   here:
409   http://refgenomes.databio.org/. Otherwise, construct a new index using the "STAR--runMode
410   genomeGenerate" command using the instructions in the STAR manual.

411

412   NOTE: For the remainder of this protocol the path to the STAR index will be referred to as
413   "path_to/STAR_index".

414

415   4.2.7. Navigate to the path_to/project/ directory. Make a new directory called "STAR_output"
416   as shown in **Figure S1D**.

417

418   4.2.8. Navigate to the path_to/project/trimmed/ directory. Use the command shown in **Figure
419   S1D** to run STAR to align the trimmed fastq.gz files.

420

421   NOTE: This step is the most computationally intensive and it is recommended to perform this on
422   a HPC cluster with multiple threads (i.e. >16) designated for the task of alignment. Depending on
423   the number of samples and available computational resources this step may take many hours to
424   days.

425

426   4.2.9. Find the required output for the next steps which contain the counts per transcript at the
427   following location: path_to/project/STAR_output/sampleN_ReadsPerGene.out.tab.

428

429   NOTE: In the ReadsPerGene.out.tab file column 1 holds information about the feature being
430   counted. Column 2 holds the unstranded read counts, column 3 holds the forward stranded read
431   counts, and column 4 holds the reverse stranded read counts. The first four rows of this file will
432   have information about the aligned reads that did not align to a single gene. This protocol
433   requires the unstranded read counts.

434

435   4.2.10. Use RStudio (preferable) or R in the HPC environment to compile the data from row 5 and
436   below for columns 1 and 2 for each sample. Set the working directory to "project" in R.

437

438   4.2.11. Read in each ReadsPerGene.out.tab file using the command in **Figure S2A**. For the first
439   column, take only the characters before the "." in the "Ensembl gene ID" column for the ease of
440   downstream processing.

441

442   4.2.12. <mark>Compile counts from all samples into a dataframe called "totcts" using the commands in</mark>
443   <mark>**Figure S2B**. Save this new table of raw count data as a tab delimited .txt file, i.e.</mark>
444   <mark>sample_counts.txt, if desired, using the "write.table" command</mark>**.**

445

446   NOTE: The order of the Ensembl gene ID is the same for every ReadsPerGene.out.tab file across
447   samples.

448

449   **5.   <mark>Differential expression</mark> and downstream analysis**

450

451   5.1. Normalize for batch effects between samples with ComBat.

452

453   NOTE: There are two possible variables that explain changes in gene expression, the first being
454   the construct used (i.e. the sample) and the second being external factors associated with the
455   passage of cells through time (i.e. the batch). A step to normalize samples for batch-to-batch
456   variation with the R-package ComBat is recommended.

457

458   5.1.1. Install if needed and load the libraries for sva, DESeq2, AnnotationDBI, org.Hs.eg.db,
459   pheatmap, RColorBrewer, genefilter, Cairo, ggplot2, ggbiplot, rgl, and reshape2 as shown in
460   **Figure S2C**. For installation, use the "install.packages" command or Bioconductor per the
461   documentation for each package.

462

463   5.1.2. First filter the data to only those genes that have at least one count per read. Save this
464   new table to denote filtering as seen in **Figure S2D**.

465

466   NOTE: Frequently, many genes will have very low or no read counts.

467

468   5.1.3. Prepare a second table for batch normalization called "vars" as shown in **Figure S2E**. Set
469   the row names to the unique names of each sample. Set the column names to "sample", "batch",
470   and "construct".

471

472   5.1.4. Assign all samples a unique number in the "sample" column from 1 to n, with n being the
473   number of samples. Assign batch numbers to all samples in the "batch" column such that
474   condition-a_1 and condition-b_1 are both assigned 1, and condition-a_2 and condition-b_2 are
475   both assigned 2. Assign all condition designations to all samples in the "construct" column such
476   that condition-a samples are all "A" and condition-b samples are all "B".

477

478   5.1.5. Define the batch variable as well, and a specific null model matrix for ComBat as shown
479   in **Figure S2F**. Run ComBat with the command defined in **Figure S2F**.

480

481   5.2.   Further curate the data by rounding to the nearest integer. Also remove genes with a
482   negative value. Use the commands shown in **Figure S3A**.

483

484   NOTE: The output of batch normalization will have non-integer read counts and some genes with

485 negative values. This step is required because the downstream differential expression analysis
486 doesn't support negative read counts.
487
488 5.3.   Define the differential expression profile for each construct using DESeq2.
489
490 5.3.1. Input the experiment design for DESeq2 as shown in **Figure S3B**. Construct a
491 DESeqDataSet (dds) using the DESeqDataSetFromMatrix function, estimate the size factors, and
492 run DESEq2, as shown in **Figure S3B**.
493
494 NOTE: It is imperative that the column data entered for "condition" is in the same order as the
495 column in the count matrix.
496
497 5.3.2. In order to evaluate the quality of the analysis, extract the rlog-normalized counts used
498 by DESeq2 as shown in **Figure S3B**.
499
500 NOTE: During analysis, DESeq2 transforms counts with a "regularized log," rlog, transformation
501 to shrink the sample-to-sample differences for genes with low counts (low information) in order
502 to preserve differences in genes with higher counts across samples (high information).
503
504 5.3.3. When extracting the results for each transcriptional profile from the results of DESeq2,
505 perform pairwise comparisons in reference to either the knockdown condition or baseline empty
506 vector as shown in **Figure S3C**. Further amend these results with the HGNC gene symbols as
507 shown in **Figure S3D**.
508
509 5.3.4. As seen in **Figure S3E**, extract data from DESeq2 results. Export as a single file with the
510 Ensembl gene ID, HGNC symbol, base mean expression, and differential expression data for all
511 constructs with log2FoldChange and raw and adjusted p-values.
512
513 NOTE: Using an adjusted p-value < 0.05 is the recommended cutoff for differential expression.
514
515 5.3.5. Assess successful batch normalization and intra-sample similarity. Check sample
516 clustering with PCA and sample-to-sample distance plots using the rlog normalized counts using
517 athe code shown in **Figures S4A-B**.
518
519 5.4.   Use the differential expression profiles to generate volcano plots using the code in **Figure
520 S4C.** Evaluate changes in gene expression across constructs.
521
522 5.5.   Use the rlog normalized counts and hierarchical clustering to identify gene signatures
523 unique to the different constructs. Use the code shown in **Figure S4D**.
524
525 5.5.1. Extract the 1000 most variable genes across all constructs in a matrix. Use pheatmap to
526 perform unsupervised hierarchical clustering of your samples based on these genes.
527
528 5.5.2. Extract the clusters of interest from the dendrogram by deciding at what level of the

529 dendrogram clusters of interest appear. Set "k" equal to the number of clusters at that level.
530 Replot the heatmap ordered by cluster to determine which clusters are of interest as shown in
531 **Figure S5**.

532

533 5.5.3.   Export the list of genes associated with each cluster as demonstrated in **Table S1**. Use this
534 information to determine the genes in clusters of interest.

535

536 5.6.   Identify the biological roles for different clusters of genes identified and compare
537 between the classes. This can be performed using a variety of bioinformatics tools. ToppGene[24]
538 is used here and is freely available online.

539

540 NOTE: There are many free tools which require just a list of genes to copy and paste into a field
541 on a website. Choose the analytical tools most appropriate for the research questions under
542 investigation.

543

544 5.7.   Optionally, if there are data available about genomic binding that drives transcriptional
545 output for transcription factor of interest, compare the transcriptional response at genes
546 associated with different binding elements to further evaluate mutant function.

547

548 **6. Comparison with Relevant Phenotypes**

549

550 6.1   Compare the correlative phenotypes with the transcriptomic profile data generated and
551 interpret as appropriate.

552

553 **REPRESENTATIVE RESULTS:**
554 Preliminary qRT-PCR data suggested that an EWS/FLI mutant called DAF, with specific tyrosine to
555 alanine mutations in the repetitive and disordered region of EWS, maintained the ability to
556 activate EWS/FLI target genes, but failed to repress critical target genes[23]. In order to better
557 understand the relationship between these residues in the EWS domain and EWS/FLI function,
558 the protocol described above and outlined in **Figure 1** was used. A673 Ewing sarcoma cells were
559 virally transduced with an shRNA targeting the 3'UTR of *FLI1*, resulting in the depletion of
560 endogenous EWS/FLI. After four days of selection, EWS/FLI function was rescued with viral
561 transduction of different 3XFLAG-tagged EWS/FLI mutant constructs, with empty vector as a
562 control for no rescue. A non-functional mutant lacking the EWS domain, called Δ22, was used as
563 a negative control and wild-type EWS/FLI, called wtEF, was used as a positive control **(Figure 2A)**.
564 DAF was used as the test construct, though more than one test construct can be used if desired.
565 Cells were selected for an additional 10 days to allow construct expression to stabilize and then
566 collected for RNA (with a gDNA removal step), protein and colony forming assays. Four replicates
567 were collected and representative qRT-PCR and western blots showing effective knockdown and
568 rescue are shown in **Figure 2B-D**. It should be noted that DAF-rescued cells failed to form colonies
569 as shown in **Figure 2E**, suggesting impaired oncogenic transformation.

570

571 Following completion of the replicate validation and phenotypic assays, RNA was submitted to
572 the Institute for Genomic Medicine at Nationwide Children's Hospital for library preparation and

573 next generation sequencing with ~50 million 150-bp paired-end reads collected. The data was
574 returned as fastq.gz files. Low-quality reads were trimmed from these files with TrimGalore and
575 STAR was used to align reads to the human genome hg19 and count the reads per gene. hg19
576 was used for purposes of compatibility with the other curated datasets for EWS/FLI used in
577 downstream analysis. These read counts were combined into a single count matrix for all
578 samples, the first 6 rows of which are shown in **Figure 3**.
579
580 Counts were initially run through DESeq2 without batch normalization, however, visual
581 inspection of the sample-to-sample distance showed potential confounding batch effects as
582 shown highlighted with red arrows in **Figure 4A**. This likely arose due to biological variability
583 introduced by the passage of cells in culture and differences in the processing of each batch.
584 Normalization for batch effects was performed with ComBat and is generally recommended. The
585 sample-to-sample distances of the batch-normalized data are shown in **Figure 4B**. Following
586 batch normalization, DESeq2 was used to generate transcriptional profiles for the three
587 constructs (wtEF, Δ22, and DAF) relative to the baseline. Note that while "parental" A673 cells
588 (mock knockdown and mock rescue, called "iLuc" here) were included in the differential analysis,
589 the reference for this experiment are the cells with EWS/FLI-depleted, called iEF cells. The
590 transcriptional profile can be generated for the endogenous protein here by comparing the iLuc
591 sample to iEF, and this may be of interest in understanding how the rescue system works, but
592 that is not the goal of this particular analysis. The transcriptional profiles generated for the
593 mutants include positive (wtEF) and negative (Δ22) controls, with respect to iEF, such that these
594 should function as the benchmarks for other mutants. This is important, as the positive control
595 in this example did not completely recapitulate the function of endogenous EWS/FLI as discussed
596 elsewhere[7,23].
597
598 The principal component analysis (PCA) in **Figure 5** suggests that the transcriptional profile of
599 DAF is intermediate between wtEF and Δ22, confirming partial function. Moreover, hierarchical
600 clustering of the 1000 most variable genes across samples showed that DAF failed to repress
601 EWS/FLI target genes, and only partially retained gene activation activity as shown in **Figure 6A**
602 and **Figure S5**. ToppGene analysis suggested that the classes of genes that DAF activates are
603 functionally distinct from those EWS/FLI-activated targets where DAF is non-functional (**Figure**
604 **6B**). Interestingly, the function of activated genes rescued by wtEF, but not by DAF, appear to be
605 related to transcriptional control and chromatin regulation. Based on the results of the colony
606 formation assays, the genes from this core gene signature should be further analyzed for their
607 role in EWS/FLI-mediated oncogenesis. The importance of EWS/FLI-mediated gene repression
608 has been previously described[17].
609
610 It is known that EWS/FLI possesses a unique binding affinity for GGAA-microsatellite repeat
611 elements[19,22], and that binding at these elements drives downstream gene regulation[11,15,18,20,22].
612 These microsatellites have been characterized as either associated with activation or repression,
613 and either proximal to (< 5 kb) TSS or distal to (> 5 kb) TSS[25]. In addition, there are EWS/FLI-
614 regulated genes with high affinity (HA) ETS motifs proximal to TSS[23]. In order to further analyze
615 the characteristics of DAF function and what types of EWS/FLI-activated genes DAF was able to
616 rescue, differential expression of genes associated with these different classes was analyzed.

617 Interestingly, DAF was most able to rescue GGAA-microsatellite activated genes, but unable to
618 rescue activated genes near an HA site as seen in **Figure 7**. As seen with hierarchical clustering,
619 DAF fails to rescue EWS/FLI-mediated repression across motif classes. These data suggest that
620 DAF retains sufficient structural features of EWS to bind to and activate from GGAA-
621 microsatellites, both proximal and distal to TSS. This likely arises from the intact SYGQ domain
622 thought to be important for EWS/FLI activity at GGAA repeats[11]. These data also suggest that the
623 specific tyrosines mutated in DAF play important, but poorly understood, roles in EWS/FLI-
624 mediated gene regulation from HA sites, as well as in gene repression, highlighting an important
625 area of further investigation.
626
627 **FIGURE AND TABLE LEGENDS:**
628 **Figure 1: Workflow**. Depiction of the step-by-step procedure to perform structure-function
629 mapping by transcriptomics. Cells were first prepared to express the suite of constructs required
630 for structure-function mapping. Following expression, cells were harvested for RNA and protein
631 and assayed for correlative phenotypes. Expression of the constructs was validated, and this
632 process was repeated 3-4 times to gather independent biological replicates. RNA was then
633 submitted for next-generation sequencing (NGS). When data was received, data was trimmed for
634 quality, aligned, and counts per transcript were calculated. Batch effects were controlled for and
635 transcriptomic signatures and differential expression were determined using DESeq2.
636 Hierarchical clustering and downstream analysis integrating other -omics datasets and different
637 pathway or functional analysis can be incorporated.
638
639 **Figure 2: Validation of construct expression and correlative assays.** (**A**) Schematic depicting the
640 constructs tested in this example. (**B**) Validation of knockdown of endogenous EWS/FLI and
641 expression of 3X-FLAG-tagged constructs by immunoblot. (**C,D**) Validation of construct activity at
642 an EWS/FLI (**C**) activated target gene, *NR0B1*, and (**D**) repressed target gene, *TGFBR2*, by qRT-
643 PCR. Data are presented as mean +/- standard deviation. P-values were calculated with a Tukey's
644 honest significance test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ (E) Colony counts from soft-agar
645 assays performed to assess transforming activity of constructs. P-values were calculated with a
646 Tukey's honest significance test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$. This figure is adapted from
647 Theisen, et al.[23]
648
649 **Figure 3: Final collated count data for analysis.** Screenshot of the first 6 rows of the count file
650 with gene counts for all the samples to be batch normalized and analyzed.
651
652 **Figure 4: Sample-to-sample distance heatmaps.** (**A**) Sample-to-sample distance plot showing the
653 sample clustering of the raw count data. Samples which are clustering both by batch and by
654 sample are denoted with red arrows. (**B**) Sample-to-sample distance plot following batch
655 normalization with ComBat. Here, samples from all replicates cluster together, independent of
656 batch.
657
658 **Figure 5: Results of differential expression analysis.** (**A**) Principle component analysis (PCA) plot
659 of the transcriptomic signatures generated for all the samples show strong intra-sample
660 clustering and demonstrate that DAF is intermediated between the positive (wtEF) and negative

661   (Δ22) controls. (**B**) Volcano plots showing the -log(p-value) plotted against the log2FoldChange
662   for genes in each construct. Genes with an adjusted p-value < 0.05 and a |log2(FoldChange)| > 1
663   are considered significant and are shown in red. Panel 5B is adapted from Theisen, et al.[23]
664
665   **Figure 6: Hierarchical clustering to identify gene classes.** (**A**) Hierarchical clustering of the top
666   1000 most variable genes across all constructs and the baseline, iEF, shows DAF partially rescues
667   EWS/FLI-mediated gene activation. (**B**) Gene ontology (molecular function) results from
668   ToppGene showing the functional enrichment of EWS/FLI-activated genes that are either rescued
669   or not rescued by DAF. Panel 6B is adapted from Theisen, et al.[23]
670
671   **Figure 7: Detailed analysis of different transcription factor response elements to different**
672   **constructs:** (**A**) Schematic depicting the data processing used to generate panels (B) and (C) by
673   incorporating other available datasets with the transcriptomic profiles here. (**B,C**) Compilation
674   showing the rescue of different classes of direct EWS/FLI- (**B**) activated and (**C**) repressed targets.
675   Genes included were only those genes with detectable differential expression by endogenous
676   EWS/FLI. In each pie chart, gray depicts the portion of genes which are not rescued by the
677   construct. Red depicts the portion of genes that are differentially activated, and blue depicts the
678   portion of genes that are differentially repressed. This figure is adapted from Theisen, et al.[23]
679
680   **Figure S1: Loading the fastq.gz files to the HPC environment, trimming and alignment.**
681
682   **Figure S2: Collating read counts across samples and running batch normalization with ComBat.**
683
684   **Figure S3: Running DESeq2 and extracting results of differential expression analysis.**
685
686   **Figure S4: Analyzing output.**
687
688   **Figure S5: Hierarchical clustering to identify gene classes:** Hierarchical clustering of the top 1000
689   most variable genes across all constructs and the baseline, iEF, sorted into *k* clusters. In this
690   instance *k*=7, but this parameter is set by the user as shown in **Figure S4D**.
691
692   **Table S1: List of genes (Ensembl gene ID) with cluster annotation**
693
694   **DISCUSSION:**
695   Studying the biochemical mechanisms of oncogenic transcription factors is critically important to
696   understand the diseases they cause and to design new therapeutic strategies. This is especially
697   true in malignancies characterized by chromosomal translocations resulting in fusion
698   transcription factors. The domains included in these chimeric proteins may lack meaningful
699   interactions with regulatory domains present in the wild-type proteins, complicating the ability
700   to interpret structure-function information in the context of the fusion[26–28]. Moreover, many of
701   these oncogenic fusions are characterized by low-complexity intrinsically disordered
702   domains[10,13,29,30].
703
704   The EWS domain is an example of such an intrinsically disordered domain that is involved in a

705   variety of oncogenic fusions[10]. The intrinsically disordered and repetitive nature has hindered
706   efforts to understand the molecular mechanisms employed by the EWS domain. Prior efforts to
707   investigate the structure-function have largely resorted to using different mutants in the context
708   of reporter gene assays or in cell backgrounds that fail to recapitulate the relevant cellular
709   context, or lack any structural variations which produce meaningful partial function[11,17,25]. The
710   method presented here addresses these issues. Structure-function mapping is performed in a
711   disease-relevant cell context and next generation sequencing enables transcriptomic profiling to
712   evaluate transcription factor function in the setting of native chromatin. In the specific case of
713   the DAF mutant of EWS/FLI, DAF was reported to show little activity in reporter assays using
714   isolated response elements, but to show activity in the context of the full gene promoter, either
715   in a reporter assay or in native chromatin, suggesting an interesting phenotype[23]. Using the
716   method described here more directly resolves the question of which type of regulatory elements
717   across the genome are most responsive in the disease setting. By testing all candidate target
718   genes in their native chromatin context simultaneously, a transcriptomic approach is more likely
719   to identify constructs with partial function.
720
721   The inherent strength of using a disease-relevant cell background is perhaps the biggest
722   limitation of this technique. One of the most important factors is choosing the appropriate cell
723   system for these experiments. Many cell lines derived from malignancies with pathognomonic
724   transcription factors do not readily tolerate knockdown of that transcription factor, and in many
725   instances, particularly for pediatric cancers, the true cell of origin remains controversial and the
726   expression of the oncogene in other cell backgrounds is prohibitively toxic[31,32]. In these cases, it
727   may be helpful to perform experiments in a different cell background, so long as the researcher
728   exercises caution in the interpretation of results and appropriately validates any relevant findings
729   in a more disease-relevant cell type.
730
731   It is critically important to carefully validate the stability and phenotypic consequences of
732   expression of the oncogene and to only submit samples for sequencing that meet strict criteria.
733   Here, this included western blot to confirm knockdown and rescue, and qRT-PCR of a small
734   number of known target genes to validate the positive control (**Figure 2**). It is likewise crucial to
735   decrease as much batch variability as possible by carefully performing the cell and RNA
736   preparations as similarly as possible through each batch.
737
738   The method described here becomes especially powerful when paired with other types of
739   genomic data that speak to the genome-wide function of the transcription factor under study.
740   Future directions for this type of structure-function analysis would expand to include ChIP-seq
741   and ATAC-seq to determine the binding of the transcription factor and any induced changes in
742   chromatin accessibility. As a suite, this type of data can point to where different structural
743   components of an oncogenic transcription factor contribute to different aspects of function (i.e.
744   DNA binding vs. chromatin modification vs. co-regulator recruitment). Overall, using NGS-based
745   approaches to map the structure-function relationships of fusion transcription factors can reveal
746   new insights in the biochemical determinants of the oncogenic function of these proteins. This is
747   important to further our understanding of the diseases they cause and in enabling the
748   development of new therapeutic strategies.

757

758 **DISCLOSURES:**
759 SLL declares a conflict of interest as a member of the advisory board for and an equity holder of
760 Salarius Pharmaceuticals. SLL is also a listed inventor on United States Patents No. US 7,393,253
761 B2, "Methods and compositions for the diagnosis and treatment of Ewing's Sarcoma," and US
762 8,557,532, "Diagnosis and treatment of drug-resistant Ewing's sarcoma." This does not alter our
763 adherence to JoVE policies on sharing data and materials.

764

765 **REFERENCES:**
766 1.   Miettinen, M. et al. New fusion sarcomas: histopathology and clinical significance of selected
767 entities. *Human Pathology.* **86**, 57–65 (2019).
768 2.   Knott, M. M. L. et al. Targeting the undruggable: exploiting neomorphic features of fusion
769 oncoproteins in childhood sarcomas for innovative therapies. *Cancer and Metastasis Reviews.*
770 **38**, 625–642 (2019).
771 3.   Yoshihara, K. et al. The landscape and therapeutic relevance of cancer-associated transcript
772 fusions. *Oncogene.* **34**, 4845–4854 (2015).
773 4.   Duesberg, P. H. Cancer genes generated by rare chromosomal rearrangements rather than
774 activation of oncogenes. *Medical Oncology and Tumor Pharmacotherapy.* **4**, 163–175 (1987).
775 5.   Dupain, C., Harttrampf, A. C., Urbinati, G., Geoerger, B., Massaad-Massade, L. Relevance of
776 Fusion Genes in Pediatric Cancers: Toward Precision Medicine. *Molecular Therapy - Nucleic Acids.*
777 **6**, 315–326 (2017).
778 6.   Mitelman, F., Johansson, B., Mertens, F. The impact of translocations and gene fusions on
779 cancer causation. *Nature Reviews Cancer.* **7**, 233–245 (2007).
780 7.   Smith, R. et al. Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in
781 Ewing's sarcoma. *Cancer Cell.* **9**, 405–416 (2006).
782 8.   Davicioni, E. et al. Identification of a PAX-FKHR gene expression signature that defines
783 molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. *Cancer
784 Research.* **66**, 6936–6946 (2006).
785 9.   Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature.*
786 **555**, 321–327 (2018).
787 10. Kim, J., Pelletier, J. Molecular genetics of chromosome translocations involving EWS and
788 related family members. *Physiological Genomics.* **1**, 127–138 (1999).
789 11. Boulay, G. et al. Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell.*
790 **171**, 163-178 (2017).
791 12. Lessnick, S. L., Braun, B. S., Denny, C. T., May, W. A. Multiple domains mediate transformation
792 by the Ewing's sarcoma EWS/FLI-1 fusion gene. *Oncogene.* **10**, 423–431 (1995).

793   13. Leach, B. I. et al. Leukemia fusion target AF9 is an intrinsically disordered transcriptional
794   regulator that recruits multiple partners via coupled folding and binding. *Structure.* **21**, 176–183
795   (2013).
796   14. Ng, K. P. et al. Multiple aromatic side chains within a disordered structure are critical for
797   transcription and transforming activity of EWS family oncoproteins. *Proceedings of the National
798   Academy of Sciences U.S.A.* **104**, 479–484 (2007).
799   15. Riggi, N. et al. EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly
800   Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell.* **26**, 668–681 (2014).
801   16. Tomazou, E. M. et al. Epigenome Mapping Reveals Distinct Modes of Gene Regulation and
802   Widespread Enhancer Reprogramming by the Oncogenic Fusion Protein EWS-FLI1. *Cell Reports.*
803   **10**, 1082–1095 (2015).
804   17. Sankar, S. et al. Mechanism and relevance of EWS/FLI-mediated transcriptional repression in
805   Ewing sarcoma. *Oncogene.* **32**, 5089–5100 (2013).
806   18. Gangwal, K. et al. Microsatellites as EWS/FLI response elements in Ewing's sarcoma.
807   *Proceedings of the National Academy of Sciences U.S.A.* **105**, 10149–10154 (2008).
808   19. Gangwal, K., Close, D., Enriquez, C. A., Hill, C. P., Lessnick, S. L. Emergent Properties of
809   EWS/FLI Regulation via GGAA Microsatellites in Ewing's Sarcoma. *Genes & Cancer.* **1**, 177–187
810   (2010).
811   20. Guillon, N. et al. The Oncogenic EWS-FLI1 Protein Binds In Vivo GGAA Microsatellite
812   Sequences with Potential Transcriptional Activation Function. *PLoS One.* **4**, e4932 (2009).
813   21. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that
814   control gene transcription. *Science.* **361**, eaar2555 (2018).
815   22. Johnson, K. M. et al. Role for the EWS domain of EWS/FLI in binding GGAA-microsatellites
816   required for Ewing sarcoma anchorage independent growth. *Proceedings of the National
817   Academy of Sciences U.S.A.* **114**, 9870–9875 (2017).
818   23. Theisen, E. R. et al. Transcriptomic analysis functionally maps the intrinsically disordered
819   domain of EWS/FLI and reveals novel transcriptional dependencies for oncogenesis. *Genes &
820   Cancer.* **10**, 21–38 (2019).
821   24. Chen, J., Bardes, E. E., Aronow, B. J., Jegga, A. G. ToppGene Suite for gene list enrichment
822   analysis and candidate gene prioritization. *Nucleic Acids Research.* **37**, W305–W311 (2009).
823   25. Johnson, K. M., Taslim, C., Saund, R. S., Lessnick, S. L. Identification of two types of GGAA-
824   microsatellites and their roles in EWS/FLI binding and gene regulation in Ewing sarcoma. *PLOS
825   One.* **12**, e0186275 (2017).
826   26. Kim, P., Ballester, L. Y., Zhao, Z. Domain retention in transcription factor fusion genes and its
827   biological and clinical implications: a pan-cancer study. *Oncotarget.* **8**, 110103–110117 (2017).
828   27. Mendíbil, I. O. de, Vizmanos, J. L., Novo, F. J. Signatures of Selection in Fusion Transcripts
829   Resulting from Chromosomal Translocations in Human Cancer. *PLOS One.* **4**, e4805 (2009).
830   28. Frenkel-Morgenstern, M., Valencia, A. Novel domain combinations in proteins encoded by
831   chimeric transcripts. *Bioinformatics.* **28**, i67-74 (2012).
832   29. Hegyi, H., Buday, L., Tompa, P. Intrinsic Structural Disorder Confers Cellular Viability on
833   Oncogenic Fusion Proteins. *PLoS Computational Biology.* **5**, e1000552 (2009).
834   30. Latysheva, N. S., Babu, M. M. Discovering and understanding oncogenic gene fusions through
835   data intensive computational approaches. *Nucleic Acids Research.* **44**, 4487–4503 (2016).
836   31. Deneen, B., Denny, C. T. Loss of p16 pathways stabilizes EWS/FLI1 expression and

837   complements EWS/FLI1 mediated transformation. *Oncogene.* **20**, 6731–6741 (2001).

838   32. Kendall, G. C. et al. PAX3-FOXO1 transgenic zebrafish models identify HES3 as a mediator of

839   rhabdomyosarcoma tumorigenesis. *eLife.* **7**, e33800 (2018).
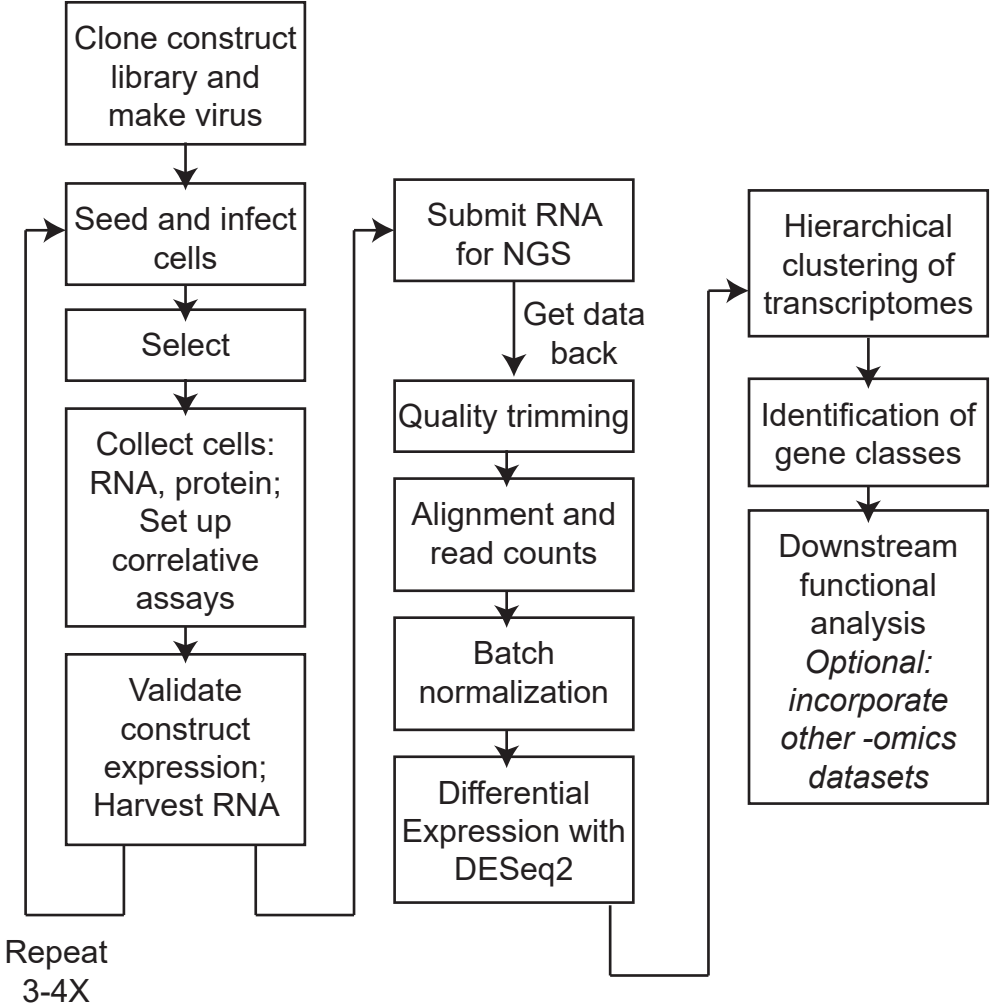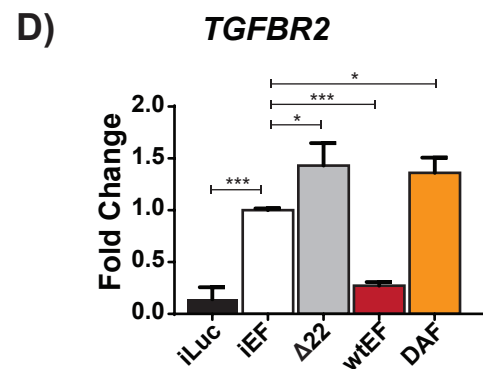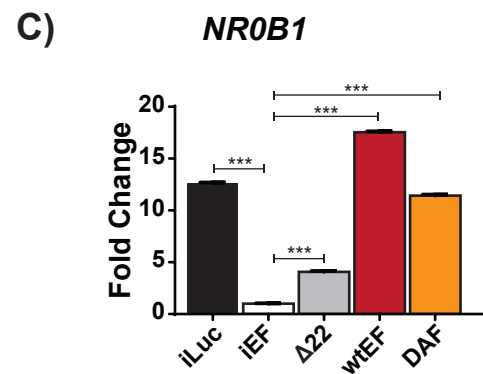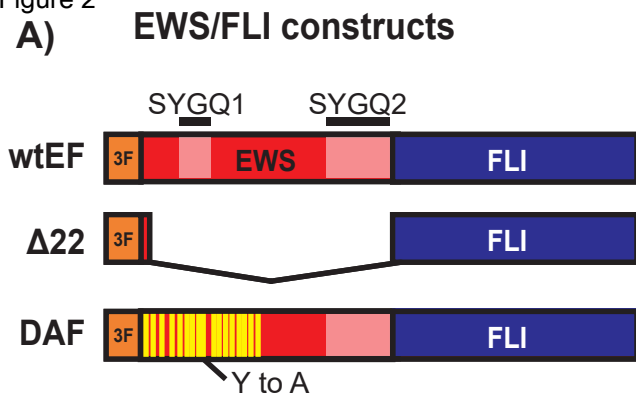
840

Figure 1

Figure 1

Figure 2

A) EWS/FLI constructs

B)

C) NR0B1

D) TGFBR2

E)

Figure 3

```
> head(totcts)
             iluc1 iluc2 iluc3 ief1 ief2 ief3 wtef1 wtef2 wtef3 del221 del222 del223 daf1 daf2 daf3
ENSG00000000003 2527  3118  3003 2807 2676 2626  2217  3182  2312   2275   2444   2035 1558 2116 1566
ENSG00000000419 3029  3072  3367 3554 2580 3518  2974  3548  3550   3295   3459   3023 2481 3493 3309
ENSG00000000457  456   520   519  501  461  514   443   588   536    425    554    413  337  465  359
ENSG00000000460 1129  1329  1212 1253  889 1139  1123  1455  1227    972   1032    566  701  962  587
ENSG00000000938    2     2     0    0    3    2     0     2     2      0      1      3    0    0    0
ENSG00000000971   81   135   161   31   64   70    18    78    35     19     79    204   61  226  400
```

Figure 4

A) Sample to Sample Distance - Raw Counts

B) Sample to Sample Distance - ComBat Counts

Figure 5

Figure 5

Figure 7   Click here to access/download;Figure;Figure 7.pdf ⬇

A)

KD/Rescue: RNA-Seq Differential Expression

Filter by EWS/FLI response element

Genes activated HA sites

Genes repressed HA sites

Genes activated distal [GGAA]n "Enhancer-like"

Genes repressed distal [GGAA]n "Enhancer-like"

Genes activated proximal [GGAA]n "Promoter-like"

Genes repressed proximal [GGAA]n "Promoter-like"

Bin gene by
1) Rescue construct
2) DE behavior

Δ22   DAF   WT-EF

🟥 Activated
🟦 Repressed
⬜ No significant (knockdown or rescue)

Pie Chart: Bin proportions per construct

B) **EWS/FLI-Activated**

	Enhancer-like [GGAA]n	Promoter-like [GGAA]n	High affinity ETS site
WT-EF			
DAF			
Δ22			
	n=404 genes	n=106 genes	n=1446 genes

C) **EWS/FLI-Repressed**

	Enhancer-like [GGAA]n	Promoter-like [GGAA]n	High affinity ETS site
WT-EF			
DAF			
Δ22			
	n=347 genes	n=48 genes	n=1314 genes

Table of Materials

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Wet Lab Reagents			
anti-FLI rabbit pAb	Abcam	ab15289	1:500
anti-lamin B1 rabbit pAb	Abcam	ab16048	1:2000
Cell-based system for introduction of mutant constructs			Determined by cell system used
Cryotubes	Corning		For viral aliquots
DMEM	Cellgro	10-013-CV	For viral production
Fetal bovine serum	Gibco	16000-044	For viral production
G418	ThermoFisher	10131027	For viral production
HEK293-EBNAs	ATCC	CRL-10852	For viral production
HEPES	Gibco	15630106	
Hygromycin B	ThermoFisher	10687010	
M2 anti-FLAG mouse mAb	Sigma	F3165	1:2000
Near IR-secondary antibodies	Li-Cor		
Optimem	Gibco	31985062	For viral production
Penicillin/Streptomycin/Glutamine	Gibco	10378-016	For viral production
Polybrene	Sigma	TR-1003-G	For viral transduction
Puromycin	Sigma	P8833	Stored at 2 mg/mL stock
RNeasy Plus kit	Qiagen	74136	Has gDNA removal columns
Selection reagents			As dictated by cell system used
Sodium Pyruvate	Gibco	11360-070	For viral production
Tissue culture media			Determined by cell system used
TransIT-LT1	Mirus	MIR 2304	For viral production
Software			
Access to HPC environment			
AnnotationDbi		1.38.2	

Cairo	1.5-10
DESeq2	1.16.1
genefilter	1.58.1
ggbiplot	0.55
ggplot2	3.1.1
org.Hs.eg.db	3.4.1
pheatmap	1.0.12
PuTTY	
R	3.4.0
RColorBrewer	1.1-2
reshape2	1.4.3
rgl	0.100.19
R-studio	
STAR	Version 2.6 or later
sva	3.24.4
TrimGalore!	
WinSCP	

Center for Childhood Cancer and Blood Diseases
The Abigail Wexner Research Institute
Nationwide Children's Hospital
700 Children's Drive
Columbus, OH 43205

25 May 2020

Dear Dr. Bajaj,

Thank you for the opportunity to revise our manuscript, 'Mapping the Structure-Function Relationships of Disordered Oncogenic Transcription Factors Using Transcriptomic Analysis.' The editorial and peer review comments were helpful and incorporating them has improved the quality of our submission, particularly the Protocol section. Find below our point-by-point responses to those comments in italics. Find included in our resubmission a version of the document with all of the changes tracked.

We are excited about resubmitting out manuscript to *JoVE* and we look forward to hearing from you at your earliest convenience.

Sincerely,
Emily R. Theisen, PhD

Editorial comments:

Changes to be made by the Author(s):

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

Response: We have thoroughly proofread the manuscript text, figure legends, and code represented in the supplementary figures.

2. Please format the manuscript as: paragraph Indentation: 0 for both left and right and special: none, Line spacings: single. Please include a single line space between each step, substep and note in the protocol section. Please use Calibri 12 points

Response: We have made these formatting changes throughout the protocol section.

3. Please define all abbreviations during the first-time use.

Response: As part of our proofreading we have checked that all abbreviations are defined at first use.

4. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.
For example: TransIT-LT1, Illumina, STAR, TrimGalore, ComBat, DESeq2,

Response: We have checked the manuscript and removed commercial language, including, but not limited to "TransIT-LT1" and "Illumina." For the program names (STAR, TrimGalore, ComBat, and DESeq2), these are not trademark symbols or registered symbols, are open-source code, freely available, and maintained by not-for-profit entities. These commands outlined in the protocol are specific to these programs, and as they do not constitute commercial language, we have kept their names in the manuscript.

5. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., "Do this," "Ensure that," etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as "could be," "should be," and "would be" throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a "Note."

Response: This change has been made. Any information that did not easily fit into an imperative tense has been moved to a "Note" below the relevant step or substep.

6. The Protocol should contain only action items that direct the reader to do something in complete sentences.

Response: This change has been made where needed.

7. Please ensure you answer the "how" question, i.e., how is the step performed?

Response: This change was made as needed throughout the protocol.

8. Please ensure that individual steps of the protocol should only contain 2-3 actions sentences per step.

Response: All steps in the protocol are now 2-3 sentences.

9. Lines 145-152: some of the details can be moved to the introduction.

Response: This information has been moved to the fifth paragraph of the introduction.

10. Significance of using HEK-EBNA cells?

Response: A "Note" regarding the use of HEK293-EBNA has been added after step 1.1.

11. There is a 10-page limit for the Protocol, but there is a 2.75-page limit for filmable content. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.

Response: We have tried to reduce the amount of highlighted text to within the 2.75-page limit. We do note, however, that several steps that we have highlighted are relatively verbose because they describe a section of code, but take only seconds to run on the computer. The amount of time to perform these tasks is much shorter than for a similar length of text related to wet lab steps. We included the minimal steps we thought were necessary to get to the end of the protocol and are happy to revisit our suggested highlighted sections should the Editorial staff feel this is necessary.

12. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."

Response: Genes & Cancer applies the Creative Commons Attribution 3.0 License (CC BY 3.0) to all works published such that so long as the original sources are cited no permission is required from either the authors or publishers for the re-use of published material. This policy can be found at the following URL: http://www.impactjournals.com/Genes&Cancer/index.php?forauthors. Accordingly, when material has been adapted from Theisen, et al. Genes & Cancer 2019 this has been noted appropriately.

13. As we are a methods journal, please ensure that the Discussion explicitly cover the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol
b) Any modifications and troubleshooting of the technique
c) Any limitations of the technique
d) The significance with respect to existing methods
e) Any future applications of the technique

Response: We have proofed the Discussion to ensure these topics have been explicitly addressed.

14. Please do not abbreviate the journal titles in the references section.

Response: We have entered the full journal title throughout the references.

Reviewers' comments:
Reviewer #1:
In this work the authors detail methods used in a recently published work (ref23) that uses RNA-seq to explore the structure-function relationship between the EWS domain of EWS-FLI1 and activation or repression function of the fusion transcription factor using the EWS-FLI1 DAF mutant.

The detail of the experimental methods is good particularly the sections on analysis of the sequencing data.

2. The authors give the appropriate cell density of plating for A673 cells. How is the correct density determined for other cell lines?

Response: We have added more information regarding determining the appropriate density at Step 2 (with a Note), Step 3.5 (with a Note), and Step 4.1. The goal is to have 50-70% confluency so cells are plentiful, but still dividing during viral transduction. The information provided should help users determine the appropriate conditions for their cells.

Figure 7. I thought that the meaning of the colors used in the pie charts were confusing. I found the legend in ref 23 more clear.

Response: The reviewer makes a good point here, and we realize we may have oversimplified the legend. We have re-incorporated information that was present in reference 23 to aid in interpretation of this figure.

Reviewer #2:

Manuscript Summary:

The manuscript by Showpnil and colleagues describes a technique to study the contribution of individual domains to the activity of the fusion transcription factor EWS-FLI1 (structure-function analysis). The procedure combines silencing of endogenous EWS-FLI1 in Ewing sarcoma cells with an shRNA directed against the 3`-UTR of the corresponding mRNA with overexpression of a RNAi-resistant, exogenous EWS-FLI1. For this, A673 Ewing sarcoma cells are first transduced with a shRNA plasmid, followed by selection with puromycin for 72h. Selected cells are then transduced with the cDNA expression plasmid and selected with hygromycin for 7-10 days. Effects of deletions of parts of the protein are then determined at the level of target gene expression by RNA seq. The protocol that is described in detail refers to bioinformatical analysis of the RNAseq dataset.

Major Concerns:

Many Figures are copy-pasted from a previous publication. It is unclear to me, whether the Journal's policy allows this.

Response: Genes & Cancer applies the Creative Commons Attribution 3.0 License (CC BY 3.0) to all works published such that so long as the original sources are cited no permission is required from either the authors or publishers for the re-use of published material. This policy can be found at the following URL: http://www.impactjournals.com/Genes&Cancer/index.php?forauthors. Accordingly, when material has been adapted from Theisen, et al. Genes & Cancer 2019 this has been noted appropriately.

Further, the lists of genes that are specific for EWS-FLI versus DAF mutant rescue should be included.

Response: We have included the table that is exported when following the code in Figure S4. This information is in Table S1 and contains only the Ensembl Gene ID and the cluster. In order to illustrate how to determine which cluster is of interest, we have included a new supplementary Figure S5. This contains the k-means clustered heatmap grouped by cluster order. We hope this is helpful.

Minor Concerns:

Figure 2F is stated in the text, but the data refers to Figure 2E.

Response: We have corrected this reference.
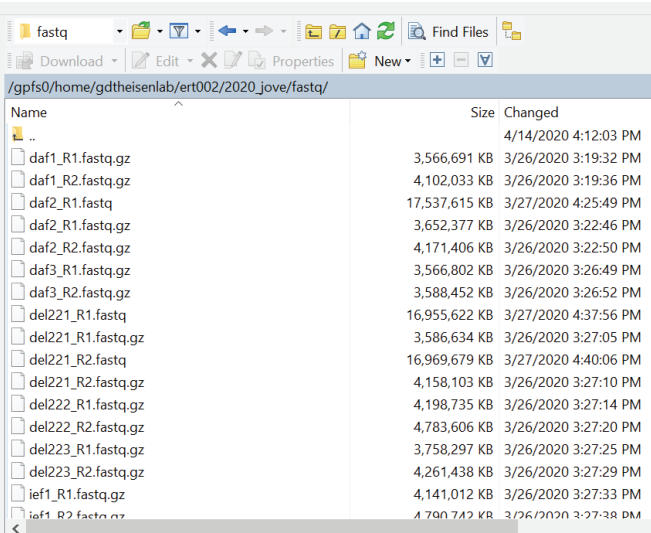
Figure 4A: Different batches could be colored to visualize the batch effect mentioned in more clearer way.

Response: This is a great suggestion. We have included the batch information in Figure 4A per the reviewer's request.

Figure S1

A
```
cd project
mkdir fastq
```

B

C
```
mkdir trimmed
cd fastq
trim_galore -o path_to/project/trimmed --paired sample1_R1.fastq.gz sample1_R2.fastq.gz
```

D
```
cd ..
mkdir STAR_output
cd trimmed
STAR --runThreadN [N] --genomeDir path_to/STAR_index --readFilesIn sample1_R1_val_1.fq.gz sample1_R2_val_2.fq.gz
--readFilesCommand zcat --outFileNamePrefix path_to/project/STAR_output/sample1_ --quantMode GeneCounts
--twopassMode Basic --outSAMunmapped Within
```

Figure S2

A
```
###Read in count data###
daf1_cts <- read.table("STAR_output/daf1_ReadsPerGene.out.tab", header = FALSE)
daf1 <- daf1_cts[-c(1:4),]
daf1$V1 <- gsub("\\..*","", daf1$V1)

daf2_cts <- read.table("STAR_output/daf2_ReadsPerGene.out.tab", header = FALSE)
daf2 <- daf2_cts[-c(1:4),]
daf2$V1 <- gsub("\\..*","", daf2$V1)
```

B
```
###Make dataframe with count data for all samples###
totcts <- data.frame(iluc1=iluc1$V2, iluc2=iluc2$V2, iluc3=iluc3$V2,
                     ief1=ief1$V2, ief2=ief2$V2, ief3=ief3$V2,
                     wtef1=wtef1$V2, wtef2=wtef2$V2, wtef3=wtef3$V2,
                     del221=del221$V2, del222=del222$V2, del223=del223$V2,
                     daf1=daf1$V2, daf2=daf2$V2, daf3=daf3$V2)

rownames(totcts) <- iluc1$V1
```

C
```
library("sva")
library(DESeq2)
library("AnnotationDbi")
library("org.Hs.eg.db")
library("pheatmap")
library("RColorBrewer")
library("genefilter")
library(Cairo)
library("ggplot2")
library("ggbiplot")
library("rgl")
library("reshape2")
```

D
```
###Filter out genes with low read counts###
n <- as.number(ncol(totcts))
keep <- (rowSums(totcts)>=n)
totcts_flt <- totcts[keep,]
```

E
```
###Define your experimental variables for batch normalization###
rows <- c("iluc1", "iluc2", "iluc3", "ief1", "ief2", "ief3", "wtef1", "wtef2", "wtef3",
          "dell221", "del222", "del223", "daf1", "daf2", "daf3")
smpl <- 1:15 ; smpl
btch <- c(1,2,3,1,2,3,1,2,3,1,2,3,1,2,3)
cond <- c("iluc", "iluc", "iluc", "ief", "ief", "ief", "wtef", "wtef", "wtef",
          "del22", "del22", "del22", "daf", "daf", "daf")

vars <- data.frame(sample = smpl, batch = btch, construct = cond)
rownames(vars) <- rows
```

F
```
###Run ComBat###
b <- vars$batch
modComBat <- model.matrix(~1, data = vars)
combat_cts <- ComBat(dat = totcts_flt, batch = b, mod = modComBat, par.prior = FALSE, prior.plots = TRUE)
```

Figure S3

A
```
###Filter out genes with negative ComBat normalized counts an round to integer###
combat_cts <- combat_cts[-which(combat_cts < 0, arr.ind=TRUE)[,1],]

combat_countData <- round(as.matrix(combat_cts,sep="\t"))
```

B
```
###Define your experimental design for DESeq2###
combat_totDesign = data.frame(
  row.names = colnames( combat_countData ),
  condition = c( "iluc", "iluc", "iluc", "ief", "ief", "ief", "wtef", "wtef", "wtef",
                "del22", "del22", "del22", "daf", "daf", "daf"),
  batch = c( "A", "B", "C", "A", "B", "C", "A", "B", "C", "A", "B", "C", "A", "B", "C") )

combat_totDesign

###Create a DESeq DataSet with count data and design###
combat_dds <- DESeqDataSetFromMatrix( countData = combat_countData,
                                      colData = combat_totDesign,
                                      design = ~ batch + condition)

###Estimate size factors and run DESeq2###
combat_dds <- estimateSizeFactors(combat_dds)
combat_dds.analysis <- DESeq(combat_dds)

###Extract rlog-normalized counts###
combat_dds.rlog <- rlog(combat_dds, blind = FALSE)
head( assay( combat_dds.rlog ), 5 )
```

C
```
###Extract pairwise comparison results###
results.1 <- results(combat_dds.analysis, contrast = c("condition", "ief", "iluc"))
summary(results.1)
results.2 <- results(combat_dds.analysis, contrast = c("condition", "ief", "wtef"))
summary(results.2)
results.3 <- results(combat_dds.analysis, contrast = c("condition", "ief", "del22"))
summary(results.3)
results.4 <- results(combat_dds.analysis, contrast = c("condition", "ief", "daf"))
summary(results.4)
```

D
```
###Add gene symbols to data###
results.1$symbol <- mapIds(org.Hs.eg.db,
                           keys=row.names(results.1),
                           column="SYMBOL",
                           keytype="ENSEMBL",
                           multiVals="first")
head(results.1)
results.1.DF <- as.data.frame(results.1)
```

E
```
###Export differenital expression data###

de_data <- data.frame(ENSID = rownames(results.1.DF), baseMean = results.1.DF$baseMean, symbol = results.1.DF$symbol,
                      iluc_log2FC = results.1.DF$log2FoldChange, iluc_pvalue = results.1.DF$pvalue, iluc_padj = results.1.DF$padj,
                      wtef_log2FC = results.2.DF$log2FoldChange, wtef_pvalue = results.2.DF$pvalue, wtef_padj = results.2.DF$padj,
                      del22_log2FC = results.3.DF$log2FoldChange, del22_pvalue = results.3.DF$pvalue, del22_padj = results.3.DF$padj,
                      daf_log2FC = results.4.DF$log2FoldChange, daf_pvalue = results.4.DF$pvalue, daf_padj = results.4.DF$padj)

write.table(de_data, "differential_expression.txt", sep = "\t", quote = FALSE, row.names = FALSE)
```

Figure S4

A
```
###Evaluate PCA for constructs###
mutde <- as.matrix(assay(combat_dds.rlog))
mutde <- as.data.frame( t(mutde))
construct <- c("iluc", "iluc", "iluc", "ief", "ief", "ief", "wtef", "wtef", "wtef", "del22", "del22", "del22", "daf", "daf", "daf")
construct <- factor(construct, level = c("iluc", "ief", "wtef", "del22", "daf"))
psg <- c("A", "B", "C", "A", "B", "C", "A", "B", "C", "A", "B", "C", "A", "B", "C")
psg <- factor(psg, level = c("A", "B", "C"))
mutde <- mutde[ , which(apply(mutde,2,var)!=0)]
pca <- prcomp(mutde, scale = TRUE, center = TRUE)
summary(pca)
plot(pca,type="lines")
scores <- data.frame(construct, psg, pca$x[,1:6])
pc1.2 <- qplot(x=PC1, y=PC2, data=scores, color=factor(construct), shape=factor(psg), size=3, xlab = "PC1", ylab = "PC2")
print(pc1.2)
```

B
```
###Evaluate Sample-to-Sample Distance
sampleDists <- dist( t( assay(combat_dds.rlog) ) )
sampleDistMatrix <- as.matrix ( sampleDists )
rownames(sampleDistMatrix) <- combat_dds.rlog$condition
colnames(sampleDistMatrix) <- NULL
colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
pheatmap(sampleDistMatrix,
        clustering_distance_rows=sampleDists,
        clustering_distance_cols=sampleDists,
        col=colors,main="Heatmap of sample-to-sample distances")
```

C
```
###Generate Volcano Plots###
with(results.2, plot(-log2FoldChange, -log10(pvalue), yaxt = "n", xaxt = "n", bty ="n", cex.lab = 1.5, pch = 20, xlim=c(-6, 6), ylim=c(0,175)))
with(subset(results.2, padj<0.05 & abs(log2FoldChange)>1), points(-log2FoldChange, -log10(pvalue), pch=20, col = "red"))
axis(side=1, lwd = 2, cex.axis = 1.75)
axis(side=2, lwd = 2, cex.axis = 1.75)

with(results.3, plot(-log2FoldChange, -log10(pvalue), yaxt = "n", xaxt = "n", bty ="n", cex.lab = 1.5, pch = 20, xlim=c(-6, 6), ylim=c(0,175)))
with(subset(results.3, padj<0.05 & abs(log2FoldChange)>1), points(-log2FoldChange, -log10(pvalue), pch=20, col = "red"))
axis(side=1, lwd = 2, cex.axis = 1.75)
axis(side=2, lwd = 2, cex.axis = 1.75)

with(results.4, plot(-log2FoldChange, -log10(pvalue), yaxt = "n", xaxt = "n", bty ="n", cex.lab = 1.5, pch = 20, xlim=c(-6, 6), ylim=c(0,175)))
with(subset(results.4, padj<0.05 & abs(log2FoldChange)>1), points(-log2FoldChange, -log10(pvalue), pch=20, col = "red"))
axis(side=1, lwd = 2, cex.axis = 1.75)
axis(side=2, lwd = 2, cex.axis = 1.75)
```

D
```
###Perform hierarchical clustering to identify gene groups in constructs###
construct.rlog <- (assay(combat_dds.rlog))[,4:15]
topVarGenes <- head(order(rowVars(construct.rlog), decreasing = TRUE), 1000)
mat <- construct.rlog[topVarGenes,]
mat  <- mat - rowMeans(mat)
anno <- as.data.frame(colData(combat_dds.rlog)[4:15, c("condition","batch")])
pheatmap(mat, annotation_col = anno, show_rownames = FALSE)

###Extract cluster information and export###
res <- pheatmap(mat, scale = "row", kmeans_k = 7)
cluster.df <- as.data.frame(factor(res$kmeans$cluster))
colnames(cluster.df) <- "Cluster"
OrderByCluster <- mat[order[(cluster.df$Cluster),]
write.table(cluster.df, "clusters.txt", quote = F, sep = '\t')
pheatmap(OrderByCluster, scale = "row", annotation_row = cluster.df, show_rownames = FALSE, cluster_rows = FALSE)
```

Ensembl Gene	Cluster
ENSG00000	3
ENSG00000	6
ENSG00000	4
ENSG00000	3
ENSG00000	4
ENSG00000	6
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	4
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	2
ENSG00000	7
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	3
ENSG00000	6
ENSG00000	6
ENSG00000	4
ENSG00000	7
ENSG00000	1
ENSG00000	4
ENSG00000	7
ENSG00000	1
ENSG00000	4
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	3
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	7
ENSG00000	7
ENSG00000	2
ENSG00000	3
ENSG00000	3
ENSG00000	4
ENSG00000	6

ENSG00000	2
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	2
ENSG00000	2
ENSG00000	7
ENSG00000	7
ENSG00000	2
ENSG00000	6
ENSG00000	7
ENSG00000	4
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	7
ENSG00000	6
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	7
ENSG00000	5
ENSG00000	7
ENSG00000	7
ENSG00000	4
ENSG00000	3
ENSG00000	5

ENSG00000	7
ENSG00000	2
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	2
ENSG00000	6
ENSG00000	4
ENSG00000	3
ENSG00000	7
ENSG00000	5
ENSG00000	3
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	4
ENSG00000	3
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	3
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	2
ENSG00000	7
ENSG00000	4

ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	6
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	4
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	7
ENSG00000	4
ENSG00000	2
ENSG00000	2
ENSG00000	3
ENSG00000	5
ENSG00000	7
ENSG00000	7
ENSG00000	2
ENSG00000	1
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	1
ENSG00000	1
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	7
ENSG00000	2

ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	3
ENSG00000	1
ENSG00000	7
ENSG00000	5
ENSG00000	4
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	7
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	6
ENSG00000	3
ENSG00000	5
ENSG00000	1
ENSG00000	7
ENSG00000	1
ENSG00000	6
ENSG00000	7
ENSG00000	5
ENSG00000	3
ENSG00000	3
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	4
ENSG00000	6
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	7

ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	6
ENSG00000	4
ENSG00000	5
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	3
ENSG00000	5
ENSG00000	6
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	7
ENSG00000	2
ENSG00000	2
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	7
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	3
ENSG00000	5

ENSG00000	4
ENSG00000	4
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	3
ENSG00000	4
ENSG00000	7
ENSG00000	2
ENSG00000	7
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	4
ENSG00000	4
ENSG00000	6
ENSG00000	1
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	6
ENSG00000	7
ENSG00000	6
ENSG00000	4

ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	4
ENSG00000	1
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	6
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	4
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	1
ENSG00000	4
ENSG00000	6
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	2
ENSG00000	7
ENSG00000	4
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	6

ENSG00000	3
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	6
ENSG00000	1
ENSG00000	2
ENSG00000	3
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	6
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	7
ENSG00000	6
ENSG00000	4
ENSG00000	5
ENSG00000	3
ENSG00000	2
ENSG00000	3
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	4
ENSG00000	4
ENSG00000	7
ENSG00000	6
ENSG00000	4
ENSG00000	1
ENSG00000	1
ENSG00000	5
ENSG00000	7
ENSG00000	7

ENSG00000	6
ENSG00000	2
ENSG00000	6
ENSG00000	6
ENSG00000	7
ENSG00000	7
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	6
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	5
ENSG00000	3
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	7
ENSG00000	4
ENSG00000	3
ENSG00000	2
ENSG00000	4
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	3
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	7
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	5

ENSG00000	7
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	1
ENSG00000	4
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	5
ENSG00000	7
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	6
ENSG00000	7
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	3
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	3
ENSG00000	3
ENSG00000	5
ENSG00000	5

ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	7
ENSG00000	7
ENSG00000	4
ENSG00000	1
ENSG00000	6
ENSG00000	1
ENSG00000	7
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	7
ENSG00000	1
ENSG00000	3
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	4
ENSG00000	3
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	2

ENSG00000	6
ENSG00000	7
ENSG00000	6
ENSG00000	6
ENSG00000	5
ENSG00000	2
ENSG00000	7
ENSG00000	7
ENSG00000	2
ENSG00000	4
ENSG00000	4
ENSG00000	3
ENSG00000	5
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	3
ENSG00000	2
ENSG00000	5
ENSG00000	3
ENSG00000	5
ENSG00000	4
ENSG00000	1
ENSG00000	7
ENSG00000	3
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	3
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	7
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	5

ENSG00000	5
ENSG00000	1
ENSG00000	4
ENSG00000	5
ENSG00000	1
ENSG00000	3
ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	3
ENSG00000	4
ENSG00000	6
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	3
ENSG00000	1
ENSG00000	1
ENSG00000	2
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	7
ENSG00000	1
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	5

ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	7
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	4
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	7
ENSG00000	3
ENSG00000	1
ENSG00000	5
ENSG00000	7
ENSG00000	6
ENSG00000	7
ENSG00000	7
ENSG00000	1
ENSG00000	6
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	2

ENSG00000	2
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	6
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	4
ENSG00000	7
ENSG00000	2
ENSG00000	7
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	2
ENSG00000	4
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	6
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	7
ENSG00000	2
ENSG00000	2
ENSG00000	5

ENSG00000	3
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	6
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	7
ENSG00000	1
ENSG00000	4
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	3
ENSG00000	3
ENSG00000	2
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	4
ENSG00000	5
ENSG00000	1
ENSG00000	5
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	1
ENSG00000	7
ENSG00000	3
ENSG00000	1

ENSG00000	6
ENSG00000	6
ENSG00000	7
ENSG00000	2
ENSG00000	1
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	7
ENSG00000	3
ENSG00000	7
ENSG00000	1
ENSG00000	6
ENSG00000	1
ENSG00000	7
ENSG00000	5
ENSG00000	3
ENSG00000	3
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	6
ENSG00000	4
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	3
ENSG00000	1
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	1
ENSG00000	4
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	7
ENSG00000	1

ENSG00000	5
ENSG00000	6
ENSG00000	2
ENSG00000	1
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	4
ENSG00000	4
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	2
ENSG00000	7
ENSG00000	3
ENSG00000	3
ENSG00000	5
ENSG00000	2
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	3
ENSG00000	2
ENSG00000	6
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	4
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	2

ENSG00000	2
ENSG00000	2
ENSG00000	3
ENSG00000	7
ENSG00000	1
ENSG00000	5
ENSG00000	6
ENSG00000	6
ENSG00000	6
ENSG00000	2
ENSG00000	4
ENSG00000	5
ENSG00000	3
ENSG00000	1
ENSG00000	7
ENSG00000	4
ENSG00000	3
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	1
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	6
ENSG00000	6
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	3
ENSG00000	2
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	7
ENSG00000	1
ENSG00000	7
ENSG00000	1
ENSG00000	1
ENSG00000	7
ENSG00000	5
ENSG00000	6

ENSG00000	1
ENSG00000	4
ENSG00000	5
ENSG00000	4
ENSG00000	5
ENSG00000	7
ENSG00000	5
ENSG00000	7
ENSG00000	4
ENSG00000	5
ENSG00000	1
ENSG00000	2
ENSG00000	2
ENSG00000	2
ENSG00000	6
ENSG00000	2
ENSG00000	3
ENSG00000	7
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	7
ENSG00000	1
ENSG00000	5
ENSG00000	7
ENSG00000	2
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	7
ENSG00000	1
ENSG00000	5
ENSG00000	1
ENSG00000	3
ENSG00000	5
ENSG00000	2
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	2
ENSG00000	3

ENSG00000	4
ENSG00000	1
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	7
ENSG00000	6
ENSG00000	7
ENSG00000	2
ENSG00000	5
ENSG00000	5
ENSG00000	5
ENSG00000	1
ENSG00000	7

Copyright and License Policies

Open-Access License
No Permission Required

Click here to access/download
Supplemental Coding Files
processing.R