RESEARCH NOTE

Triplex target sites of MEG3 RNA-chromatin interactions

[version 1; peer review: 1 approved, 2 not approved]

Ivan Antonov ^{1,2}, Yulia A. Medvedeva¹⁻³

¹Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russian Federation

²Department of Biological and Medical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation

³Vavilov institute of General Genetics, Russian Academy of Sciences, Moscow, Russian Federation

V1 First published: 17 Jan 2018, 7:76
<https://doi.org/10.12688/f1000research.13522.1>

Latest published: 09 May 2019, 7:76
<https://doi.org/10.12688/f1000research.13522.2>

Abstract

Many long noncoding RNAs are bound to chromatin. MEG3 binds to multiple different genomic locations, containing GA-rich motifs, and form RNA-DNA triplex structures. In this work, we test whether the MEG3 binding sites are specific enough to be regulated by a particular lncRNA. We show that at least in the case of MEG3, a subset of the triplex target sites (TTS) is able to hybridize with various different RNAs almost irrespectively of their sequences. Nowadays, the role of chromatin bound RNAs in the formation of 3D chromatin structure is actively discussed. We speculate that such universal TTSs may contribute to establishing long-distance chromosomal contacts.

Keywords

MEG3 lncRNA, triple helix, triplex target sites (TTS)

Open Peer Review

Approval Status

	1	2	3
version 2			
(revision)			
09 May 2019	view		view
version 1			
17 Jan 2018	view	view	view

1. **Ingrid Grummt**, German Cancer Research Center (DKFZ), Heidelberg, Germany

2. **Andrey A. Mironov**, Moscow State University, Moscow, Russian Federation
 RAS (Russian Academy of Sciences), Moscow, Russian Federation

3. **Hao Zhu**, Southern Medical University, Guangzhou, China

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Yulia A. Medvedeva (ju.medvedeva@gmail.com)

Author roles: **Antonov I:** Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Medvedeva YA:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Russian Science Foundation [grant 14-15-30002].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Antonov I and Medvedeva YA. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Antonov I and Medvedeva YA. **Triplex target sites of MEG3 RNA-chromatin interactions [version 1; peer review: 1 approved, 2 not approved]** F1000Research 2018, 7:76 <https://doi.org/10.12688/f1000research.13522.1>

First published: 17 Jan 2018, 7:76 <https://doi.org/10.12688/f1000research.13522.1>

Introduction

Many human long noncoding RNAs are localized in the nucleus and potentially can participate in chromatin formation and remodeling¹. Recently, technologies such as ChIRP², ChRIP³, ChOP⁴, CHART⁵, RAP⁶, MARGI⁷ and GRID⁸ have been developed to map the genomic interacting sites of various lncRNAs. Although these techniques determine location of RNA binding sites, they are unable to clarify the interaction mechanisms. lncRNAs are capable of binding chromatin proteins, nascent RNA, single-stranded or double-stranded DNA, forming R-loops or triple helices, respectively.

Maternally expressed gene 3 (MEG3) is one of the lncRNAs known to target chromatin. Genome-wide mapping of MEG3 with the chromatin oligo affinity precipitation followed by deep sequencing (ChOP-seq) method reveals that MEG3 binding sites are widespread, contain GA-rich motifs, and form RNA-DNA triplex structure⁴. Growing body of evidence shows that RNA-DNA triplex formation plays important role in RNA-chromatin interactions. Moreover, it has been shown earlier that triplex target sites (TTS) are frequently located near regulatory regions (including gene promoters) in the human genome⁹. In this work we investigate whether the DNA sites capable of triplex formation are specific enough to be regulated by a particular lncRNA.

Methods

After mapping 6837 ChOP-seq MEG3 peaks from hg19 to the hg38 using liftOver (the tool was downloaded from the UCSC Genome Browser on Nov 7, 2017 and ran as follows: liftOver MEG3.hg19_peaks.bed, hg19ToHg38.over.chain.gz MEG3.hg38_peaks.bed, unmapped.txt), 6694 peaks shorter than 1000 bp were used. Next, we selected 3kb regions (bins) centered at the peak middle positions using bedtools (version 2.25.0). The 3620 bins with overlapping genes (according to the GENCODE ver. 27 annotation¹⁰) were selected as true positives. Additionally, 3620 genomic regions of the same length and GC-content overlapping the GENCODE genes were randomly selected from the human genome as true negatives (the control bins). The validation set consisted of two subsets of bins without MEG3 peaks. Namely, another group of 3620 control bins were sampled and combined with the 3620 true negative regions from the test set.

We predicted triple helices using Triplexator¹¹ (version 1.3.2) with the settings recommended at the official website: `-l 15 -e 20 -c 2 -fr off`) with the following RNA queries:

- MEG3 (NR_002766: length = 1595 nt, GC-content = 57.55%),
- BE2L6 (NM_198183: 1620 nt, 57.59%),
- LILRA3 (NM_006865: 1608 nt, 57.71%),
- HMOX1 (NM_002133: 1590 nt, 57.80%).

The transcript sequences similar to the MEG3 in length and GC content were found using the RANN (version 2.5.1) R package¹². UBE2L6, LILRA3 and HMOX1 were used to verify the sequence

specificity of the MEG3-DNA hybridization. Additionally, the three random query sequences were obtained by mono-nucleotide shuffling the original MEG3 transcript using a custom Perl script. The Triplexator score for each genomic region was calculated as the sum of the scores of all the predictions between RNA and the genomic region.

Results

The Triplexator tool¹¹ was used to predict the RNA-DNA interactions between the MEG3 transcript and the 7240 genomic regions (bins) from the test set according to the Hoogsteen and reverse Hoogsteen base pairing rules. As anticipated, the triplex scores predicted for the MEG3 peak-containing bins were stronger than for the control bins – the median Triplexator SumScores were 48 and 25, respectively (p-value = 5.2e-100, see Figure 1a).

Strikingly, in all cases the SumScores produced by Triplexator for 3 other human RNAs (UBE2L6, LILRA3 and HMOX1) and the MEG3 peak-containing bins were also stronger than the scores for the control bins. Moreover, the statistical significances of the observed SumScore differences for two out of the three mRNAs were higher than for the MEG3 predictions (p-values = 0, 3.0e-24 and 1.4e-174, respectively – see Figure 1b). To find out whether it is a general property of the human transcripts or the identified MEG3-TTS have a tendency to form triplexes with any RNA in a nonspecific manner, the three random sequences were generated by the mono-nucleotide shuffling of the MEG3 transcript. Once again, the statistical significant differences between the two sets of bins were observed in all three cases (p-values = 1.0e-143, 5.8e-41 and 1.3e-33 – see Figure 1c).

To rule out a possibility of overprediction, we applied our computational approach to a 'validation set' consisting of the MEG3 peak-free bins only (see Methods). On the contrary to the test set, no significant difference between the two groups of control bins was found for any of the seven RNA sequences (all p-values > 0.05, Figures 1d–f).

Dataset 1. Coordinates of the original hg19-based and the converted hg38-based MEG3 peaks

<http://dx.doi.org/10.5256/f1000research.13522.d189750>

Dataset 2. Sequences of the seven queries as well as all the bins from the test and the validation sets

<http://dx.doi.org/10.5256/f1000research.13522.d189751>

Dataset 3. SumScores computed by running the Triplexator for each of the queries against the test or the validation set

<http://dx.doi.org/10.5256/f1000research.13522.d189752>

Dataset 4. Sequences of the bins containing putative 'universal' TTSs

<http://dx.doi.org/10.5256/f1000research.13522.d189753>

Figure 1. (a–c) The distributions of the Triplexator SumScores for the 3620 control regions without peaks and 3620 genomic regions with MEG3 peaks identified in the ChOP-seq experiment. (d–f) The distributions of the Triplexator SumScores for two sets of genomic regions without MEG3 peaks. The query transcript used in Triplexator run is indicated below each image.

Discussion

Our results suggest that at least in some cases, the formation of the RNA-DNA triplexes may be governed by the DNA sequence alone. If it is so, such 'universal' TTSs are able to hybridize with various different RNAs almost irrespectively of their sequences (however the length and nucleotide content are probably important). Indeed, 18 peak-containing bins were present in the top 5% of the predictions for all seven tested RNAs. In contrast, there was only one such bin among the control regions. Notably, some of the 18 identified universal bins were extremely GA-rich (for example, hg38:chr5:93580373-93583373). The presence of the universal TTSs among the MEG3 peaks may explain the phenomena observed in our computational analysis.

Therefore, some parts of the human genome can hybridize with a number of different RNAs (or different regions of the same long

RNA). It should be noted that one of the possible and actively discussed roles of the chromatin bound RNAs (including lncRNAs) is to bring different chromosomal parts together to enable the remote DNA-DNA interactions⁸. In the light of this biological role, the universal TTS can be viewed as the anchor point which can be bound by various nuclear RNAs to provide longdistance chromosomal contacts. Analysis of additional datasets is needed to further support this hypothesis.

Data availability

Dataset 1: Coordinates of the original hg19-based and the converted hg38-based MEG3 peaks. DOI, [10.5256/f1000research.13522.d189750](https://doi.org/10.5256/f1000research.13522.d189750)¹³

Dataset 2: Sequences of the seven queries as well as all the bins from the test and the validation sets. DOI, [10.5256/f1000research.13522.d189751](https://doi.org/10.5256/f1000research.13522.d189751)¹⁴

Dataset 3: SumScores computed by running the Triplexator for each of the queries against the test or the validation set. DOI, [10.5256/f1000research.13522.d189752](https://doi.org/10.5256/f1000research.13522.d189752)¹⁵

Dataset 4: [universal_TTS_bins.fna.gz]. Sequences of the bins containing putative 'universal' TTSs. DOI, [10.5256/f1000research.13522.d189753](https://doi.org/10.5256/f1000research.13522.d189753)¹⁶

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by the Russian Science Foundation [grant 14-15-30002].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are thankful to Dr. Chandrasekhar Kanduri (University of Gothenburg, Sweden) for providing the original coordinates of the ChOP-seq peaks for the lncRNA MEG3.

References

- Khalil AM, Guttman M, Huarte M, *et al.*: Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009; **106**(28): 11667–72. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chu C, Qu K, Zhong FL, *et al.*: Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011; **44**(4): 667–78. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pandey RR, Mondal T, Mohammad F, *et al.*: *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell*. 2008; **32**(2): 232–246. [PubMed Abstract](#) | [Publisher Full Text](#)
- Mondal T, Subhash S, Vaid R, *et al.*: MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA-DNA triplex structures. *Nat Commun*. 2015; **6**: 7743. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simon MD, Wang CI, Kharchenko PV, *et al.*: The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*. 2011; **108**(51): 20497–502. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Engreitz JM, Pandya-Jones A, McDonel P, *et al.*: The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013; **341**(6147): 1237973. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sridhar B, Rivas-Astroza M, Nguyen TC, *et al.*: Systematic Mapping of RNA-Chromatin Interactions *In Vivo*. *Curr Biol*. 2017; **27**(4): 602–609. [PubMed Abstract](#) | [Publisher Full Text](#)
- Li X, Zhou B, Chen L, *et al.*: GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017; **35**(10): 940–950. [PubMed Abstract](#) | [Publisher Full Text](#)
- Goñi JR, De La Cruz X, Orozco M: Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res*. 2004; **32**(1): 354–360. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harrow J, Frankish A, Gonzalez JM, *et al.*: GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012; **22**(9): 1760–1774. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buske FA, Bauer DC, Mattick JS, *et al.*: Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res*. 2012; **22**(7): 1372–81. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Arya S, Mount D, Kemp SE, *et al.*: Package 'RANN'. 2017. [Reference Source](#)
- Antonov I, Medvedeva Y: Dataset 1 in: Triplex target sites of MEG3 RNA-chromatin interactions. *F1000Research*. 2018. [Data Source](#)
- Antonov I, Medvedeva Y: Dataset 2 in: Triplex target sites of MEG3 RNA-chromatin interactions. *F1000Research*. 2018. [Data Source](#)
- Antonov I, Medvedeva Y: Dataset 3 in: Triplex target sites of MEG3 RNA-chromatin interactions. *F1000Research*. 2018. [Data Source](#)
- Antonov I, Medvedeva Y: Dataset 4 in: Triplex target sites of MEG3 RNA-chromatin interactions. *F1000Research*. 2018. [Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 05 February 2018

<https://doi.org/10.5256/f1000research.14683.r29953>

© 2018 Zhu H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hao Zhu

Bioinformatics Section, Southern Medical University, Guangzhou, China

Many lncRNAs can bind to DNA sequences by forming triplexes (the binding sites are often called TTS, triplex-targeting sites). Whether there are “universal TTS” as described here is interesting and unreported, and deserves a careful investigation. But I have a major concern about the work: the authors reach the conclusion upon too few examples. Also, why these lncRNAs (BE2L6, LILRA3, HMOX1) were chosen (randomly or selected for some reasons)?

A few others issues should also be addressed. First, what is the relationship between the “universal TTSS” and base-pairing rules is untouched. For example, do the universal TTSSs allow many lncRNAs to bind to them using the same rules? If very different rules are involved, what does this mean? To some extent, binding upon different rules indicates lncRNA specific TTSSs, instead of universal TTSSs. Second, I feel that the genomic regions used to sum scores are unreasonably long (3000 bp). Finally, it is said that “the median Triplexator SumScores were 48 and 25, respectively (p-value=5.2e-100)”. Statistically, the difference is significance, but biologically might not. I think 48 is not that large and 25 is not that small.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome analysis, lncRNA analysis, molecular evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 10 May 2019

Ivan Antonov, Russian Academy of Sciences, Moscow, Russian Federation

Many lncRNAs can bind to DNA sequences by forming triplexes (the binding sites are often called TTS, triplex-targeting sites). Whether there are “universal TTS” as described here is interesting and unreported, and deserves a careful investigation. But I have a major concern about the work: the authors reach the conclusion upon too few examples. Also, why these lncRNAs (BE2L6, LILRA3, HMOX1) were chosen (randomly or selected for some reasons)?

In the current version we have significantly increased the number of query RNAs (i.e. 153 expressed transcripts and 153 random sequences obtained by di-nucleotide shuffling of MEG3 lncRNA). The 153 transcripts that we use now were chosen so that their lengths and GC contents were similar to the MEG3 lncRNA.

1) A few others issues should also be addressed. First, what is the relationship between the “universal TTSs” and base-pairing rules is untouched. For example, do the universal TTSs allow many lncRNAs to bind to them using the same rules? If very different rules are involved, what does this mean? To some extent, binding upon different rules indicates lncRNA specific TTSs, instead of universal TTSs.

Our preliminary analysis indicated that different RNAs interacted with universal TTSs via mixed (G or U) motif a little bit more frequently than via the purine or pyrimidine motifs. Importantly, all the analyzed transcripts were predicted to form a lot of triple helices (using different RNA motifs) with the universal TTSs. This is what makes uTTS special rather than specific motifs .

2) Second, I feel that the genomic regions used to sum scores are unreasonably long (3000 bp).

In the current version we decreased the genomic region lengths by considering the exact ChOP-seq and Capture-seq peaks.

3) Finally, it is said that “the median Triplexator SumScores were 48 and 25, respectively (p-value=5.2e-100)”. Statistically, the difference is significance, but biologically might not. I think 48 is not that large and 25 is not that small.

We now use p-values instead of SumScore to estimate the statistical significance of

each RNA-DNA interaction.

Competing Interests: No competing interests were disclosed.

Reviewer Report 26 January 2018

<https://doi.org/10.5256/f1000research.14683.r29955>

© 2018 Mironov A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Andrey A. Mironov

¹ Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russian Federation

² Institute for Information Transmission Problems, RAS (Russian Academy of Sciences), Moscow, Russian Federation

The manuscript describes an application of the Triplexator software for search possible binding sites of the imprinting related MEG3 linc RNA on the human genome. The authors give a good example of the statistical analysis of the results. The main paper about this software has 62 references (google scholar data). Most of them have only reference to the software and only a few of them used the Triplexator. Only a few reports show a success story about the application of the Triplexator software and comparison the results with an experiment. In some papers, a significant enrichment of triplex targets on regions of interest was found. But they did not analyze the specificity of the predicted triplex formation. The current paper focused on a specificity of the Triplexator predictions. The authors got unexpected results that the Triplexator gives many non-specific hits for the case.

Comments:

1. Description of similar transcripts and the parameters of the Triplexator software should be rearranged because the appearance of some RNA names before definition sounds strange. The parameters of the Triplexator software contains a reference to the BE2L6 RNA while the description of the control RNA set has a reference on UBE2L6.
2. The di-nucleotide shuffling seems more adequate for RNA analysis.
3. In one paper¹ the Triplexator software also was used for analysis of MEG3 RNA-DNA contacts. The comparison of the obtained results with the results of given manuscript should be provided. Seems in current manuscript a more accurate analysis with good controls was provided.
4. It would be good to look at the practice of using the program on literature and make sure that the program has a sufficiently low specificity.

References

1. Mondal T, Subhash S, Vaid R, Enroth S, et al.: MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA-DNA triplex structures. *Nat Commun.* 2015; **6**: 7743
[PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 10 May 2019

Ivan Antonov, Russian Academy of Sciences, Moscow, Russian Federation

Dear reviewer,

We would like to apologise for a significant delay with the reply to all the comments. To implement the changes suggested by the reviewers we had to completely redesign our study and significantly increase the number of analyzed RNAs. Particularly, we developed a new method to estimate the statistical significance of the number of triplexes predicted for each RNA-DNA pair. Moreover, we analyzed the results of the recently published Capture-seq experiment that identified interactions of three different RNA oligos with "naked" DNA. We hope that all these analyses improved our study.

1) Description of similar transcripts and the parameters of the Triplexator software should be rearranged because the appearance of some RNA names before definition sounds

strange. The parameters of the Triplexator software contains a reference to the BE2L6 RNA while the description of the control RNA set has a reference on UBE2L6.

We have made the appropriate corrections in the text

2) The di-nucleotide shuffling seems more adequate for RNA analysis.

We now use the di-nucleotide shuffling to generate random RNA sequences.

3) In [one paper](#)¹ the Triplexator software also was used for analysis of MEG3 RNA-DNA contacts. The comparison of the obtained results with the results of given manuscript should be provided. Seems in current manuscript a more accurate analysis with good controls was provided.

In this original paper the authors focused on the triplex-based interactions of a single RNA (MEG3) with the chromatin. In the present study we are interested whether other transcripts may have a potential to interact with the same genomic regions. Taking into account the different aims (and approaches) of the studies we are not sure if it is reasonable to compare their results.

4) It would be good to look at the practice of using the program on literature and make sure that the program has a sufficiently low specificity.

In our recent benchmarking study [PMID:29697742] we showed that Triplexator was the most accurate tool as of 2018. This is why we used it in the present study.

Competing Interests: No competing interests were disclosed.

Reviewer Report 23 January 2018

<https://doi.org/10.5256/f1000research.14683.r29954>

© 2018 Grummt I. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ingrid Grummt

Division of Molecular Biology of the Cell II , DKFZ-ZMBH-Allianz, German Cancer Research Center (DKFZ), Heidelberg, Germany

Long noncoding RNAs (lncRNA) can regulate gene expression by targeting specific DNA sequences via Hoogsteen base pairing, forming RNA-DNA triple helical structures. Computational analyses have revealed that a large population of triplex-forming motifs is present across the genome, the majority of annotated human genes containing at least one unique and high-affinity triplex target site, preferentially in regulatory gene regions (Goni et al. 2004, Buske et al. 2012). Moreover, several studies have provided in vitro and in vivo evidence for the existence and biological relevance of RNA-DNA triplexes, including pRNA (Schmitz et al. 2010), Fendrr (Grote et al. 2013), Khps1 (Postepska-Igielska et al. 2015), PARTICLE (O'Leary et al. 2015), and MEG3 (Mondal et al.

2015). MEG3 has been shown to associate with AG-rich DNA motifs and facilitate recruitment of PRC2 to target sites. Considering the large number of purine-rich sequences in the genome, triplex-mediated targeting of lncRNAs and associated proteins to distinct genomic loci is very likely a commonly used mechanism of gene regulation.

Given the importance and emerging acceptance of the concept of triplex-dependent gene regulation, it is more than surprising, if not irritating, that the authors challenge this concept feeding the 'Triplexator' only with a few RNAs and a subset of MEG3-interacting regions rather than providing any experimental data and/or more global bioinformatic analysis.

Just some specific comments:

- In the abstract they claim *'these triplex interactions might contribute to establishing long-distance chromosomal contact'* without providing any information or bioinformatic analyses.
- They use the term 'hybridization' for the interaction between RNA and dsDNA. This is wrong as hybridization refers to Watson-Crick base-pairing between RNA and ssDNA and not to Hoogsteen bonding.
- They took MEG3-interacting DNA peaks shorter than 1000 bp, then selected 3000 bp bins centering these regions and used these bins for analysis. There is no rationale for this selection which of course determines the final outcome of the analysis. Accordingly, the majority of these bin regions did not coincide with regions determined by ChOP-seq. Probably, a shorter binning would be more reliable to analyze the available data.
- They focused on bins that overlap genes. Even if partial overlapping was accepted, they might have missed some promoters. Intergenic regions containing regulatory sequences (e.g. enhancers) were excluded.
- Why were only peaks overlapping with annotated genes considered to be significant (or "real")? Genomic regions that do not harbor annotated genes, such as enhancers, are important regulatory elements that are targeted by lncRNAs and as such are functional RNA-binding sites, highly relevant for this study. In addition, since it is known that the base composition of genic and intergenic regions is different, exclusion of intergenic regions introduces a considerable bias to the analysis.
- Selection of just three additional RNAs is certainly not adequate for the far-reaching conclusion: *'TTSs are able to hybridize with various different RNAs almost irrespectively of their sequence'*. It would be more convincing to show the results from scanning more RNAs, irrespectively of their length and GC-content. Also, there is no attention given to the expression profiles of selected MEG3-mimicking RNAs. This is important because transcription of MEG3 is highly tissue-specific.
- The sum scores from the Triplexator analysis are shown which does not mean that the same regions are involved in triplex formation. It would be much more convincing to show similarities (or differences) of triplex-forming RNAs for a given TTS in a given bin.
- The terms 'universal TTS' and 'universal bins' are not synonymous and interchangeable!! One bin (3000 bp) can contain many putative TTSs.
- If there are only 18 'universal bins' out of 3620 bins among seven RNA analyses, this small number is not sufficient for claiming that there is no specificity in RNA targeting.
- They hypothesize that *'the universal TTS can be viewed as the anchor point which can be bound by various nuclear RNAs to provide long-distance chromosomal contacts'*. Even if this might be true, without any supportive data this is pure speculation.

Altogether, the authors' claim that triplex formation occurs almost sequence-independent is not

justified but is based solely on *in silico* analyses. At least another available bioinformatics tool should have been used and standard *in vitro* assays (e.g. EMSA experiments) should have been performed to validate that the candidate RNAs are indeed capable to form triplexes. The authors do not even mention that the *in vivo* situation might be completely different than algorithm-based predictions and that there might be additional factors/constraints involved in triplex formation and stability.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Regulation of gene expression by noncoding RNA

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 10 May 2019

Ivan Antonov, Russian Academy of Sciences, Moscow, Russian Federation

We would like to thank Dr. Grummt for the extended comments to our work. They have helped us improve the design of our study and obtain additional results. We hope that they made our conclusions more reliable and reproducible.

Given the importance and emerging acceptance of the concept of triplex-dependent gene regulation, it is more than surprising, if not irritating, that the authors challenge this concept feeding the 'Triplexator' only with a few RNAs and a subset of MEG3-interacting regions rather than providing any experimental data and/or more global bioinformatic analysis.

We do not challenge the possibility of triplex-dependant regulation. We simply claim that RNA interactions with some genomic regions have low sequence specificity because many other RNAs may be able to bind the same loci. We investigated features of such regions and found them to be enriched in purine rich low complexity repeats. In the current version, we completely redesigned the study and incorporated the analysis of 306 RNA sequences to confirm our findings. We also modified the text so the main conclusions are clear and non-misleading.

Just some specific comments:

1) In the abstract they claim 'these triplex interactions might contribute to establishing long-distance chromosomal contact' without providing any information or bioinformatic analyses.

We believe it is reasonable to speculate about this possibility in the discussion for the following reasons. First, it has recently been shown that "RNAs originating from super-enhancers form triplexes at distant regions" [PMID: 30605520]. Second, we showed that the predicted Capture-seq universal TTSs were highly enriched in gene promoters (Supplementary Figure 10). Together these observations indicate that the same eRNA may be able to interact with several different universal TTSs and therefore contribute to the long-distance chromosomal (i.e. enhancer-promoter) contacts. However, additional experimental verification of this hypothesis is required.

We modified the text in the abstract as follows: "We speculated that such universal TTSs may contribute to establishing long-distance chromosomal contacts and may facilitate distal enhancer-promoter interactions."

2) They use the term 'hybridization' for the interaction between RNA and dsDNA. This is wrong as hybridization refers to Watson-Crick base-pairing between RNA and ssDNA and not to Hoogsteen bonding.

We have corrected the terminology used in the manuscript.

3) They took MEG3-interacting DNA peaks shorter than 1000 bp, then selected 3000 bp bins centering these regions and used these bins for analysis. There is no rationale for this selection which of course determines the final outcome of the analysis. Accordingly, the majority of these bin regions did not coincide with regions determined by ChOP-seq. Probably, a shorter binning would be more reliable to analyze the available data.

We now use the exact locations for all the ChOP-seq peaks. To compensate for the peak length variability we developed a method that estimates the statistical significance of the number of triplexes predicted for a RNA-DNA pair taking into account lengths of both sequences.

4) They focused on bins that overlap genes. Even if partial overlapping was accepted, they might have missed some promoters. Intergenic regions containing regulatory sequences (e.g. enhancers) were excluded.

We now analyze all the ChOP-seq peaks without considering their overlaps with the annotated genes.

5) Why were only peaks overlapping with annotated genes considered to be significant (or

"real")? Genomic regions that do not harbor annotated genes, such as enhancers, are important regulatory elements that are targeted by lncRNAs and as such are functional RNA-binding sites, highly relevant for this study. In addition, since it is known that the base composition of genic and intergenic regions is different, exclusion of intergenic regions introduces a considerable bias to the analysis.

We now analyze all the ChOP-seq peaks.

6) Selection of just three additional RNAs is certainly not adequate for the far-reaching conclusion: 'TTSs are able to hybridize with various different RNAs almost irrespectively of their sequence'. It would be more convincing to show the results from scanning more RNAs, irrespective of their length and GC-content. Also, there is no attention given to the expression profiles of selected MEG3-mimicking RNAs. This is important because transcription of MEG3 is highly tissue-specific.

We have increased the number of the considered query RNAs to 306 and used the expressed transcripts only.

7) The sum scores from the Triplexator analysis are shown which does not mean that the same regions are involved in triplex formation. It would be much more convincing to show similarities (or differences) of triplex-forming RNAs for a given TTS in a given bin.

We no longer use sum scores as a measure of triplex-based interaction. Instead, we estimate the statistical significance (p-value) of each RNA-DNA interaction based on the number of predicted triplexes.

Our work was focused on the properties of the DNA sequences that may allow them to interact with various different RNAs. We therefore analyzed the parts of the ChOP-seq/Capture-seq peaks universal TTSs that allowed them to interact with various different RNAs. Our analysis indicates that such triplex-forming hot-spots frequently coincide with the purine-rich low complexity genomic regions.

8) The terms 'universal TTS' and 'universal bins' are not synonymous and interchangeable!! One bin (3000 bp) can contain many putative TTSs.

We do not use the concept of bins and 'universal bins' in the current version of the manuscript. However, we kept the term 'universal TTS'.

9) If there are only 18 'universal bins' out of 3620 bins among seven RNA analyses, this small number is not sufficient for claiming that there is no specificity in RNA targeting.

We would like to emphasize that our paper don't question the concept of the sequence specific triplex-dependent gene regulation (moreover, we support this idea and conduct research in this direction). We claim that some genomic regions may have a potential to form triple helices with a variety of different long RNAs forming "universal" triplex target sites (TTSs). At the same time, we do not challenge the sequence specificity of the other triplex-based RNA-DNA interactions.

10) They hypothesize that 'the universal TTS can be viewed as the anchor point which can be bound by various nuclear RNAs to provide long-distance chromosomal contacts'. Even if this might be true, without any supportive data this is pure speculation.

We agree that at the moment our claim is a hypothesis/speculation. Nevertheless, the recent published results [PMID: 30605520] as well our own indicate the possibility of such

mechanism. By discussing it in the current manuscript we hope to attract attention of experimental biologists to further study this topic.

We modified the text in the paper to clarify the issue:

"One of the possible and actively discussed roles of the chromatin bound RNAs (including lncRNAs) is to bring different chromosomal parts together to enable the remote DNA-DNA contacts. Moreover, it has recently been shown that RNAs originating from super-enhancers form triplexes at distant regions. Therefore, it is possible that universal TTSs may facilitate distal enhancer-promoter interactions via engagement with the same enhancer RNA. In line with this hypothesis, we observed the statistical significant enrichment of the universal Capture-seq peaks near (< 1 kb) the transcription start sites (TSSs) of the annotated genes (Supplementary Figure 10C)."

11) Altogether, the authors' claim that triplex formation occurs almost sequence-independent is not justified but is based solely on in silico analyses. At least another available bioinformatics tool should have been used and standard in vitro assays (e.g. EMSA experiments) should have been performed to validate that the candidate RNAs are indeed capable to form triplexes. The authors do not even mention that the in vivo situation might be completely different than algorithm-based predictions and that there might be additional factors/constraints involved in triplex formation and stability.

We added analysis of the recent in vitro data obtained by the Capture-seq method.

According to this experimental data some genomic fragments can interact with all three RNA oligos used in the original study. This supports the observations obtained in the analysis of the ChOP-seq peaks. Our computational analysis classified almost 40% of these shared Capture-seq peaks as universal TTSs. Moreover, the ChOP-seq and the Capture-seq universal TTSs were similar in that they were enriched with the purine rich low complexity regions. We believe that these results indirectly support the existence of universal TTS. Yet, experimental validation of these results is beyond the scope of the current paper.

Still, at the end of the manuscript, we discuss the limitations of our computational approach and mention that the obtained results resemble mostly the situation with the naked DNA in vitro, than the interactions with the chromatin in vivo.

We added the following text to the discussion:

"Importantly, the current computational analysis has a number of limitations. Namely, the triplex-based interactions of the full length transcripts were predicted without taking their secondary structure into account. We are not aware of any bioinformatics tools that would be able to produce such predictions. Moreover, cellular localization of the 153 selected expressed transcripts as well as DNA binding proteins and chromatin compaction were not considered. Therefore, our simulations are more similar to the in vitro Capture-seq experiments with short oligos than to the interactions of long transcripts with the chromatin inside the nucleus."

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research