

TT-seq captures enhancer landscapes immediately after T-cell stimulation

Margaux Michel, Miss Carina Demel, Benedikt Zacher, Björn Schwalb, Stefan Krebs, Helmut Blum, Julien Gagneur and Patrick Cramer

*Corresponding authors: Julien Gagneur, Technische Universität München;
Patrick Cramer, Max Planck Institute for Biophysical Chemistry*

Review timeline:

Submission date:	16 December 2016
Editorial Decision:	23 January 2017
Revision received:	13 February 2017
Accepted:	15 February 2017

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision

23 January 2017

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who agreed to evaluate your study. As you will see below, the reviewers raise some concerns, which we would ask you to address in a revision.

The reviewers' recommendations are quite clear so I think that there is no need to repeat the points listed below. Please let me know in case you would like to further discuss any of the issues raised.

REFeree REPORTS

Reviewer #1:

This manuscript reports the application of TT-seq, a powerful method recently developed by the authors, to study rapid transcriptional responses of enhancers and promoters. It is demonstrated here that 4sU labeling pulses as short as 5 minutes can be applied to detect rapid up- and down-regulation of transcriptional activity genome-wide after application of particular stimuli to cells. Because TT-seq appears to be a relatively simple method (compared to some other run-on transcription mapping methods) these results are important to a broad readership and worthy of publication in MSB. The manuscript is clear and the data look solid.

Specific comments:

1. From the main text (incl Methods) it is unclear how many independent (?) biological replicates were done, and how these data were combined. How reproducible are the results between replicates? Scatterplots and correlation coefficients or similar metrics should be included (as Supp figures), so that readers get a clear sense of the reproducibility of the key results.
2. At some points the text is quite dry, as it is stuffed with counts of transcripts and other numbers, e.g. "TT-seq uncovered 247 up- and 109 down-regulated transcripts after 5 min, 1,369 up- and 581 down-regulated transcripts after 10 min, and 2,447 up- and 1,297 down-regulated transcripts after 15 min following stimulation", followed by "Out of a total of 3,744 transcripts that showed significantly changed synthesis 15 min after stimulation, 638 were mRNAs, and 2,986 were ncRNAs, including 120 lncRNAs. Of the mRNAs, 311 were up-regulated and 327 were down-regulated". Perhaps some of this detailed information can be summarized more globally and/or moved to figures or tables.
3. Please provide a reviewer link to the GEO accession, which is currently private.
4. p6. Please clarify how half-life of mRNAs was calculated based on these TT-seq data. It seems to me that a pulse-chase design is needed, but here only a pulse is given without the chase. How reliable/reproducible are these estimates? In Fig 3C half life is shown on an "arbitrary scale"; why not % per time unit?
5. Fig 3D. Y-axis reads "percentage ..." The maximum value is about 0.5 -- which would be very low as a percentage (about 1 in 200). Or should this be "fraction"? Furthermore, the text states "the putative eRNAs were flanked by a region of high DNase hypersensitivity" but ignores that ~50% (or 99.5%?) are not. A more balanced discussion is warranted.
6. I had hoped for a more critical attitude towards the simplistic assumption that "insulated neighborhoods" can be captured by taking CTCF-delineated ChIA-PET loops. In fact, it seems to me that the new TT-seq data offer an interesting opportunity to test how good this assumption is, compared to an alternative model in which simply distance between enhancer and promoter is the relevant parameter. Does the loop model indeed perform better in terms of correlations between enhancer and promoter activation? Such an analysis would be very informative for the "CTCF loop" community.

Reviewer #2:

The manuscript reports a very short time course stimulation (total 15 minutes) of Jurkat T cells followed with the TT-seq method developed by the authors of this manuscript. TT-seq seems to be suitable to broadly detect the immediate changes in the genome activity like promoters of mRNA and ncRNA/lncRNAs and enhancers. One of the main results is that at least a part of the immediately early enhancers and promoters are co-activated.

I have the following comments for the authors.

- 1) Due to the growingly importance of the lncRNA/ncRNAs, it is important to clarify what they are. The Gencode annotation has been a bit confused as well, but this requires clarification here. The authors report them as ncRNA and lncRNAs, reflecting previous annotations. However, are many of the ncRNAs essentially lncRNAs? And are many of the lncRNA/ncRNAs also enhancer RNAs, which may be stabilized? Again, while I understand that the Gencode does not help at first, I think that this issue should be clarified, at least by inspecting a number of lncRNA and ncRNAs. I would at least expect an in depth discussion about this.
- 2) In my understanding, there are 4 libraries produced without replication at each of the 4 time

points. The author can leverage of the similarity of the response at 5, 10 and 15 minutes (at least in Fig 1 and 2 in the shown examples we can see similar responses with different amplitude). However, the authors also suggest that eRNA transcription is high variable (Page 8), however this needs either replication or validation of a number of them.

3) Importantly, the authors note that the activation of enhancers and promoters is simultaneous, and suggest that eRNAs are less stable. The authors suggest that conclusions are different from previous work, as cited in the manuscript, which suggested early activation of enhancers RNAs followed by promoter activation. We should be careful here for several reasons. (a) Actually, so short time intervals (5-15 minutes only activation) and Jurkat T cells activation are not used at least in Arner et al, one of the most comprehensive analysis so far (b) In the same Arner et al, there are several examples where there is "co-activation" of enhancers and promoters: the study contains many time points and biological, where many co-activation enhancer-promoters pairs can be seen and examples similar to this work may be seen. (c) Often, the mRNA expression peaks at much later stage (hours), in kinetics that are much longer.

1st Revision - authors' response

13 February 2017

Responses are in italics

Reviewer #1:

This manuscript reports the application of TT-seq, a powerful method recently developed by the authors, to study rapid transcriptional responses of enhancers and promoters. It is demonstrated here that 4sU labeling pulses as short as 5 minutes can be applied to detect rapid up- and down-regulation of transcriptional activity genome-wide after application of particular stimuli to cells. Because TT-seq appears to be a relatively simple method (compared to some other run-on transcription mapping methods) these results are important to a broad readership and worthy of publication in MSB. The manuscript is clear and the data look solid.

We would like to thank the reviewer for the support. We have gone through all of the points and updated the text and figures accordingly.

1. From the main text (incl Methods) it is unclear how many independent (?) biological replicates were done, and how these data were combined. How reproducible are the results between replicates? Scatterplots and correlation coefficients or similar metrics should be included (as Supp figures), so that readers get a clear sense of the reproducibility of the key results.

We are sorry that we had forgotten to add this information to the submitted manuscript. We have now included a paragraph in the Appendix methods and added scatterplots with correlations (Pearson correlation 0.97) as supplemental figures (Appendix Figures S1 and S2) for the time points where deep-sequenced replicates were collected and for total RNA at different time points (total RNA hardly changes within the first 15 minutes after stimulation). Generally, these analyses show that our data are highly reproducible and the quality of the data is very high.

2. At some points the text is quite dry, as it is stuffed with counts of transcripts and other numbers, e.g. "TT-seq uncovered 247 up- and 109 down-regulated transcripts after 5 min, 1,369 up- and 581 down-regulated transcripts after 10 min, and 2,447 up- and 1,297 down-regulated transcripts after 15 min following stimulation", followed by "Out of a total of 3,744 transcripts that showed significantly changed synthesis 15 min after stimulation, 638 were mRNAs, and 2,986 were ncRNAs, including 120 lncRNAs. Of the mRNAs, 311 were up-regulated and 327 were down-regulated". Perhaps some of this detailed information can be summarized more globally and/or moved to figures or tables.

We thank the reviewer for this comment and changed the text (p. 7) and figure (Fig 2) accordingly and added supplemental tables in the Appendix. We tried to move data from text to legends or

methods were possible, but some key numbers will have to remain in the main text.

3. Please provide a reviewer link to the GEO accession, which is currently private.

The GEO reviewer link is send by email to the editor, who we kindly ask to forward the link.

4. p6. Please clarify how half-life of mRNAs was calculated based on these TT-seq data. It seems to me that a pulse-chase design is needed, but here only a pulse is given without the chase. How reliable/reproducible are these estimates?

Indeed, it is not possible to estimate half-lives from TT-seq data alone. However, the combination of TT-seq data with total RNA-seq data enables us to estimate synthesis and degradation rates without a pulse-chase design. We have described this in detail elsewhere (Schwalb et al., Science 2016). Briefly, we use the information that total cellular RNA reflects the ratio of synthesis rate over degradation rate. As TT-seq gives an estimate for the synthesis rate, it is possible to determine the degradation rates from the RNA-seq data. We added an additional paragraph in the Appendix supplementary methods. This method was developed in the yeast system and there are multiple publications that use, test, and verify this approach. In particular, we had shown that the estimation of half-lives from 4sU-labeled RNA in combination with total RNA is less corrupted than rates derived by pulse-chase design due to shorter labeling times (Sun et al, Genome Research 2012).

In Fig 3C half life is shown on an "arbitrary scale"; why not % per time unit?

We corrected Figure 3C to show the half-life in absolute scale (minutes). We thank the reviewer for spotting this.

5. Fig 3D. Y-axis reads "percentage ..." The maximum value is about 0.5 -- which would be very low as a percentage (about 1 in 200). Or should this be "fraction"? Furthermore, the text states "the putative eRNAs were flanked by a region of high DNase hypersensitivity" but ignores that ~50% (or 99.5%?) are not. A more balanced discussion is warranted.

We see that this notion was misleading and thank the reviewer for pointing this out. Indeed we meant "fraction" and corrected accordingly. As the "presence of a region of high DNase hypersensitivity (=peak)" is based on a cutoff (for peak detection), we decided to change the figure in the revised manuscript and show raw coverage instead. We also carefully edited the text to reflect this.

6. I had hoped for a more critical attitude towards the simplistic assumption that "insulated neighborhoods" can be captured by taking CTCF-delineated ChIA-PET loops. In fact, it seems to me that the new TT-seq data offer an interesting opportunity to test how good this assumption is, compared to an alternative model in which simply distance between enhancer and promoter is the relevant parameter. Does the loop model indeed perform better in terms of correlations between enhancer and promoter activation? Such an analysis would be very informative for the "CTCF loop" community.

We thank the reviewer for this comment and agree that this is a partially open question. We think enhancer-promoter pairing within "CTCF loops" is less stringent than selecting the closest neighbor regarding genomic coordinates. Also, topological changes upon treatment could account for varying regulation of a promoter by different enhancers. To answer the question if pairing within CTCF loops improves correlation over pairing between close neighbors, we conducted additional analyses. We compared the correlation of enhancer-promoter pairs within "CTCF loops" and the correlation of synthesis changes for each mRNA and their closest eRNA. We see that the loop model leads to clear improvement of the correlation. Also closest enhancer-promoter pairs within a CTCF loop have a higher correlation than other closest pairs. We included supplemental figures in the Appendix and revised the manuscript (p. 9) accordingly.

Reviewer #2:

The manuscript reports a very short time course stimulation (total 15 minutes) of Jurkat T cells followed with the TT-seq method developed by the authors of this manuscript. TT-seq seems to be suitable to broadly detect the immediate changes in the genome activity like promoters of mRNA and ncRNA/lncRNAs and enhancers. One of the main results is that at least a part of the immediately early enhancers and promoters are co-activated.

We thank the reviewer for the careful evaluation and insightful comments.

1) Due to the growingly importance of the lncRNA/ncRNAs, it is important to clarify what they are. The Gencode annotation has been a bit confused as well, but this requires clarification here. The authors report them as ncRNA and lncRNAs, reflecting previous annotations. However, are many of the ncRNAs essentially lncRNAs? And are many of the lncRNA/ncRNAs also enhancer RNAs, which may be stabilized? Again, while I understand that the Gencode does not help at first, I think that this issue should be clarified, at least by inspecting a number of lncRNA and ncRNAs. I would at least expect an in depth discussion about this.

We are sorry if this was not clear. We called lincRNA all transcribed regions overlapping at least 20% with GENCODE-annotated lincRNAs. We called ncRNAs all the remaining transcribed regions non-overlapping coding genes. We next further classified the ncRNA as eRNAs when they overlapped an enhancer region. We understand the text needed clarification about this, which we did now (p. 6). We also checked that the labels in the supplemental figures are fine.

In addition, it is not easy to distinguish lincRNAs from eRNAs (Espinosa, Mol Cell 2016, and Paralkar, Mol Cell 2016). Hence, it is true that GENCODE-annotated lincRNAs could potentially also be eRNAs, and a large fraction (385/590) indeed overlaps with enhancer states in T cell lines. A conservative approach is to omit GENCODE-annotated lincRNAs from our eRNA analysis. We added a paragraph in the discussion (p. 12) and produced another Appendix Figure (S3), that shows difference between our lincRNA and eRNA classes, both in terms of length and half-life. We have expanded the discussion and this should address the concern.

2) In my understanding, there are 4 libraries produced without replication at each of the 4 time points. The author can leverage of the similarity of the response at 5, 10 and 15 minutes (at least in Fig 1 and 2 in the shown examples we can see similar responses with different amplitude). However, the authors also suggest that eRNA transcription is high variable (Page 8), however this needs either replication or validation of a number of them.

We are sorry that the information on replicates was missing. We included a paragraph and supplemental figures in the Appendix. Indeed, we show that our TT-seq data is highly reproducible. The statement on p. 8 that this reviewer refers to was "For a large fraction of eRNAs (29%) we observed significant changes in their synthesis between time points (Methods), showing that eRNA transcription is highly variable." This concerns changes with respect to time point 0 that are statistically significant and indicate that eRNAs are highly regulated. We rewrote the sentence to avoid such confusion in the future. Please also compare our answer to reviewer #1.

3) Importantly, the authors note that the activation of enhancers and promoters is simultaneous, and suggest that eRNAs are less stable. The authors suggest that conclusions are different from previous work, as cited in the manuscript, which suggested early activation of enhancers RNAs followed by promoter activation. We should be careful here for several reasons. (a) Actually, so short time intervals (5-15 minutes only activation) and Jurkat T cells activation are not used at least in Arner et al, one of the most comprehensive analysis so far (b) In the same Arner et al, there are several examples where there is "co-activation" of enhancers and promoters: the study contains many time points and biological, where many co-activation enhancer-promoters pairs can be seen and examples similar to this work may be seen. (c) Often, the mRNA expression peaks at much later stage (hours), in kinetics that are much longer.

We thank the reviewer for this comment. We have considerably revised and extended the text (p. 12) to account for the concerns of the reviewer. We agree that the main difference in these studies is the time frame; whereas we investigate immediate changes after activation within minutes, the published work generally investigates changes hours after stimulation, which allows for strong changes in the chromatin landscape. Note that when we cite the Arner paper we say that enhancer transcription CAN precede promoter transcription, not that it DOES, consistent with the co-activation pointed out by the reviewer. We made sure the text correctly reflects this. We do not wish to argue that there are any issues with the prior work and trust this message is correctly conveyed.

2nd Editorial Decision

15 February 2017

Thank you for sending us your revised manuscript. We think that the points raised by the referees have been satisfactorily addressed and I am pleased to inform you that your paper has been accepted for publication.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Patrick Cramer

Journal Submitted to: Molecular Systems Biology

Manuscript Number: MSB-16-7507

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures**1. Data****The data shown in figures should satisfy the following conditions:**

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if $n < 5$, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions**Each figure caption should contain the following information, for each panel where they are relevant:**

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/ varied/ perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
 - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
 - are tests one-sided or two-sided?
 - are there adjustments for multiple comparisons?
 - exact statistical test results, e.g., P values = x but not P values < x;
 - definition of 'center values' as median or average;
 - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	NA
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	We used the Mann-Whitney-U-Test which is used for non-normally distributed values.
Is there an estimate of variation within each group of data?	Yes.
Is the variance similar between the groups that are being statistically compared?	Yes.

C- Reagents**USEFUL LINKS FOR COMPLETING THIS FORM**<http://www.antibodypedia.com><http://1degreebio.org><http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo><http://grants.nih.gov/grants/olaw/olaw.htm><http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm><http://ClinicalTrials.gov><http://www.consort-statement.org><http://www.consort-statement.org/checklists/view/32-consort/66-title><http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun><http://datadryad.org><http://figshare.com><http://www.ncbi.nlm.nih.gov/gap><http://www.ebi.ac.uk/ega><http://biomodels.net/><http://biomodels.net/miriam/><http://jij.biochem.sun.ac.za>http://oba.od.nih.gov/biosecurity/biosecurity_documents.html<http://www.selectagents.gov/>

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	NA
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Jurkat cells were retrieved from DSMZ (Braunschweig, Germany), not recently authenticated but tested for mycoplasma contamination.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	TT-seq and RNA-seq at GEO under accession number GSE85201
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	Done.
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: Primary Data Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in <i>Shewanella oneidensis</i> MR-1. Gene Expression Omnibus GSE39462 Referenced Data Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	Done (differently at MSB)
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomedelis (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC (see link list at top right)). According to our biosecurity guidelines, provide a statement only if it could.	NA
---	----