

## Transcriptional regulatory networks underlying gene expression changes in Huntington's disease

Seth A. Ament, Jocelynn R. Pearl, Jeffrey P. Cantle, Robert M. Bragg, Peter J. Skene, Sydney R. Coffey, Dani E. Bergey, Vanessa C. Wheeler, Marcy E. MacDonald, Nitin S. Baliga, Jim Rosinski, Leroy E. Hood, Jeffrey B. Carroll and Nathan D. Price

---

### Review timeline:

Submission date:	10 November 2016
Editorial Decision:	23 January 2017
Revision received:	13 December 2017
Editorial Decision:	15 January 2018
Revision received:	23 January 2018
Accepted:	26 February 2018

---

Editor: Thomas Lemberger

### 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 three referees who agreed to evaluate your manuscript. First of all, I would like to apologize for the delay in getting back to you, which was due to the late arrival of the report from referees after the Christmas break. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, important concerns on your work, which should be addressed in a revision.

The main concerns are expressed by referee #3 and refer to the need more support for the biological relevance of some of the key results. We appreciate that you use separate functional genomics and proteomics datasets to replicate the prioritization of the core set of TFs and that a TF-TF regulatory network is inferred. The biological and functional significance of the inferred regulatory relationships displayed in this network remains however rather speculative at this stage and its relevance to design potential protective strategies is not clear. With regard to SMAD3, we appreciate that the chromatin location study is consistent with a down-regulation of SMAD3-regulated transcription in HD mouse models and with the previously protective role of TGFbeta in vitro (Ring et al 2015). Whether TGF/SMAD signaling is down-regulated in medium spiny neurons in HD mouse striata remains however unclear. In view of the comments of reviewer #3 asking for stronger validation, it seems that some advance on one or the other of these two major open issues could considerably strengthen the manuscript.

In terms of presentation, not sufficient details are provided to understand and reproduce the analysis. Thus it is unclear how the TF-TF network was actually inferred. Overlap between gene sets (differentially expressed, sets of target genes, bound genes) are reported in terms of statistical

significance but the size of the overlap ('effect size') is often not reported, making it very difficult to intuitively understand the potential biological relevance of some of the results.

When you resubmit your manuscript, please download our CHECKLIST (<[http://embopress.org/sites/default/files/Resources/EP\\_Author\\_Checklist\\_Master.xlsx](http://embopress.org/sites/default/files/Resources/EP_Author_Checklist_Master.xlsx)>) and include the completed form in your submission. \*Please note\* that the Author Checklist will be published alongside the paper as part of the transparent process <<http://msb.embopress.org/authorguide#transparentprocess>>.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favorable.

---

## REVIEWER REPORTS

Reviewer #1:

The manuscript "Genome-scale transcriptional regulatory network 1 models for the mouse and human striatum predict roles for SMAD3 and other transcription factors in Huntington's disease" is well written and very informative about the identification of transcriptional factors that play a role in HD. The study provides a mouse and human striatal-specific TRN and prioritizes a hierarchy of transcription factor drivers in HD. The work is outstanding and the following up on SMAD3 compelling. The authors are very clear about the precision of the methods. Some questions that remain on the analysis are the following:

- (1) Did the authors find transcription factors that regulate BDNF levels in their analysis? They mention other publications but did not discuss this. Also were the transcription factors identified overlapping with the BDNF knockout mice transcriptional changes in their study?
- (2) I think cortex TF analysis is highly relevant and should be included as a comparison.
- (3) Are the changes in SMAD3 relevant to a particular neuronal type in the brain? Did they compare their analysis to single cell transcriptional changes in HD?

Minor points

- (1) Maybe change wording on (However, to our knowledge a role for SMAD3 has not been described). "Consistent with our work Ring et al identified but did not characterize transcriptional changes in SMAD3. See "Targets of the transcription factor SMAD3 were overrepresented ( $p = 0.0018$ ), and the top WikiPathway was BDNF signaling ( $p = 0.0053$ )." Ring et al.
- (2) Reference line 30 Zuccato et al 2007 is in different font
- (3) Font on figures 3 and 4 should be larger

Reviewer #2:

This paper seems to be very mature, so perhaps I am seeing it after some rounds of revision. The method is a bit complex and not cutting-edge novel, but it is an interesting and novel use of TF-networks to investigate HD. Recent work has shown that CAG repeats yield tremendous phenotypic (and probably endotypic) complexity (Alexandrov et al. Nature Biotech 2016), so this paper is a very valuable resource for starting to tease apart the underlying molecular complexity of Htt repeat mutations. I applaud that the authors have released all of their analysis code in to the public domain via GitHub. If I were to be nit picky I think that Figure 1a gets a bit confusing where they go from linear models to networks as they don't really use the networks per se, though they can use the results of the linear models to paint networks. However, the paper is really about using data

integration to rank TFs and their targets. That being said it's a very minor comment. I think this is a well written paper with important results and clear and detailed analysis and is suitable for publication in MSB.

Reviewer #3:

#### Summary

Ament et al. reconstructed a genome scale transcriptional regulatory network (TRN) Huntington disease (HD) model. They did this by using digital genomic footprinting to predict the binding site of 871 TFs and tested the accuracy by comparing the TFBS with ChIP-seq data. They looked at co-expression patterns in RNA-seq transcriptome profiles to identify TF-target gene interactions in the mouse striatum and validated the predictive accuracy of the model by comparing the prediction with the observed expression level of each gene.

To identify TFs that are core regulators of transcriptional changes in HD they used published RNA-seq data from mice expressing different alleles of the HTT pathogenic variants. They identified a 48 TF core set whose target genes were differentially expressed (DEG) in 2, 6 and 10 months old mice. Published microarray and proteomics data sets were used to confirm the prediction for the selected 48 TFs core sets. In addition, the authors tested whether disease target genes identified in mouse, were differentially expressed in human late stage HD. They found a subset of 13 TFs of the initial 48 TFs differentially expressed in the mouse RNA-seq, mouse microarray, mouse proteomics and published human microarray data. Finally, the authors selected one TF, SMAD3, and performed ChIP-Seq experiments with 4 months old HttQ111 and wildtype striatum samples and found many SMAD3 targets downregulated.

Altogether, the authors aim in providing a global model between HD-related transcriptomic changes mediated by core TFs including a mouse and human striatal-specific TRN and a hierarchy of TFs.

#### General remarks

HD is a devastating disease and the authors merged and integrated published HD datasets of mice and men to extract TRNs that might trigger downstream pathological processes. Conceptually, this integration of published data into a TRN model, can be valuable to identify pathogenic molecular mechanisms that if known, might be drug targets resulting in neuroprotection. This HD TRN model can be a primer for biomedical research. The genome-scale TRN analysis sounds technically solid, but a more detailed explanation within the text and the method section is needed. However, it remains unclear whether the identified core TFs including SMAD3, have a biological impact in HD pathology or are "bystanders", since none of the identified core TFs have been tested for HTT interactions.

The in silico TRN analysis alone is interesting as a model, but without experimental validation, it remains elusive if the findings have a biological significance.

#### Major points

My major concerns are that it is unclear how biologically relevant is the TRN analysis, since none of the 13 core TFs have been experimentally validated in HD neither by the authors, nor by other researchers. The discussion is very short and superficial and a great opportunity to position the TRN analysis within the HD field is unused. Furthermore, taking advantage of published datasets has the downside that each set comprises different mouse models, ages, sample preparations and other methods, which might impede the analysis and must be better addressed in the manuscript.

1. The authors claim in the abstract to provide a "hierarchy of transcription factor drivers in HD". Where is this addressed in the manuscript? It is unclear.
2. Page 4 (lines 11-13): Since none of the TF drivers have been experimentally validated one should phrase "potential transcription factor drivers..." Otherwise, it is overstated.
3. Pages 6 and 7, the core TFs in independent datasets are not presented and discussed properly, for example only 22 of 48 TFs were enriched in protein samples and 13 of 48 core TFs were found in human samples. Please discuss this low fractions whether it stems from the analysis or from the samples (late human HD versus early onset mouse HD, ...).

4. Page 7 line 12 "This overlap was not statistically greater than expected by chance" and then the parameters were adjusted to obtain statistically significant differences? It sounds that the human data doesn't correlate with the mouse data and the conclusions are overstated (lines 19 -22).
5. Why was SMAD3 selected? Figure 3B does not support SMAD3 to be a top candidate in human HD.
6. Why were four-month-old HttQ111 mice used and not 2, 6 or 10 months old for SMAD3 ChIP-seq? This is inconsistent and a repetition with age-matched samples is needed.
7. Why were HttQ111 mice chosen and not Q140 or Q175 that have more DEGs at 6 and 10 months?
8. What is the biological and functional link between SMAD3 and the other core TFs with mutated HTT protein? An experimental validation for some of the core TFs is needed to support the in silico model.

#### Minor points

1. The TRN model (pages 4-5) needs a comprehensive description for a broader audience.
2. The mouse models used (page 6) need better explanation; for example when does the neurodegeneration starts. It is painful to look up multiple references to obtain an overview.
3. How old were the mice used for microarray data (page 6)?
4. The TF-to-TF regulatory interaction analysis is not sufficiently explained (page 7). Figure 4 is also unreadable; one should consider a comprehensive graphical presentation.
5. Why does the peak range to a TSS vary among analysis, initially set to 5kb and for SMAD3 it is 10kb. This is obviously inconsistent.
6. Overall, a more detailed methods section would strengthen the manuscript.
7. Figure 2 is better suited for a supplementary figure. One could also consider adapting the color code to the other figures.
8. Overall, the readability and clarity could be improved to invite readers from the biomedical field.

To: Thomas Lemberger, PhD, Chief Editor  
Molecular Systems Biology

Re: MSB-16-7435  
December 13, 2017

Dear Dr. Lemberger,

Thank you for providing us the opportunity to submit a revised version of our manuscript, "Transcriptional regulatory networks underlying gene expression changes in Huntington's disease." In response to your suggestions and those of the peer reviewers, we have performed new experiments assessing age- and *Htt* CAG length-dependent changes in the expression of SMAD3 transcripts and proteins. We also added new analyses and substantially revised the text for clarity.

The revision process took eight months to complete, and we apologize for the slow timeframe. Both lead authors on the manuscript, Drs. Ament and Pearl, transitioned to new positions over the last several months. Dr. Ament is now an Assistant Professor at the University of Maryland School of Medicine. Dr. Pearl is now a Fellow at the Altius Institute for Biomedical Sciences. These transitions led to inevitable delays in completing new experiments and analyses and in revising the manuscript. Under the circumstances, I hope you will still consider this as a revision rather than as a new submission.

Below is a summary of the changes we have made in response to specific referee comments.

#### Editorial Suggestions.

At your suggestion, we have expanded our methods section to clearly describe our analyses so that other groups are able to reproduce our work.

Another main concern was the need for additional biological relevance of some of the key results. In order to address this, we performed additional experiments to examine changes in SMAD3 and phosphorylated-SMAD3 in 4 and 11-month-old mice. On page 8, we added several sentences framing the literature and results around SMAD3 to better orient the reader to the importance of our results and others exploring the role of this transcription factor in HD.

#### Reviewer 1.

- (1) Did the authors find transcription factors that regulate BDNF levels in their analysis? They mention other publications but did not discuss this. Also were the transcription factors identified overlapping with the BDNF knockout mice transcriptional changes in their study?
- (2) I think cortex TF analysis is highly relevant and should be included as a comparison.
- (3) Are the changes in SMAD3 relevant to a particular neuronal type in the brain? Did they compare their analysis to single cell transcriptional changes in HD?

*We thank the reviewer for their helpful feedback and questions. In response to (1) we have expanded the results section "Biological associations of core TFs" on page 7. In response to point (2) we have added a paragraph on page 6 lines 27-41 which discusses gene expression changes in four additional tissues (cortex,*

*liver, hippocampus, and cerebellum). This analysis reveals a range of tissue specificity within the predicted TF modules. For example, predicted gene targets of RXRG were enriched in all five tissues, whereas predicted targets of IRF2 were enriched only in striatum. The results of this analysis are described in supplemental Figure 7. In response to (3), we were certainly interested in the relevance of particular TFs to neuronal or other cellular subtypes in the brain. We include in the lower panel of Figure 5 the cell-type specific enrichments of each TF's target genes across seven major cell types in the brain. We observe a predominant enrichment for these genes in Drd1 or Drd2+ neurons. We welcome the reviewer's question about single-cell transcriptional changes in HD. At this time, we are not aware of any single-cell transcriptomic data related to HD that is available to us. Data of this kind would be particularly helpful in understanding the contributions of particular cell types to gene expression changes observed in bulk tissues such as striatum.*

Reviewer 2.

This paper seems to be very mature, so perhaps I am seeing it after some rounds of revision. The method is a bit complex and not cutting-edge novel, but it is an interesting and novel use of TF-networks to investigate HD. Recent work has shown that CAG repeats yield tremendous phenotypic (and probably endotypic) complexity (Alexandrov et al. Nature Biotech 2016), so this paper is a very valuable resource for starting to tease apart the underlying molecular complexity of Htt repeat mutations. I applaud that the authors have released all of their analysis code in to the public domain via GitHub. If I were to be nit picky I think that Figure 1a gets a bit confusing where they go from linear models to networks as they don't really use the networks per se, though they can use the results of the linear models to paint networks. However, the paper is really about using data integration to rank TFs and their targets. That being said it's a very minor comment. I think this is a well written paper with important results and clear and detailed analysis and is suitable for publication in MSB.

*We thank the reviewer for their positive feedback.*

Reviewer 3.

Major points

My major concerns are that it is unclear how biologically relevant is the TRN analysis, since none of the 13 core TFs have been experimentally validated in HD neither by the authors, nor by other researchers. The discussion is very short and superficial and a great opportunity to position the TRN analysis within the HD field is unused. Furthermore, taking advantage of published datasets has the downside that each set comprises different mouse models, ages, sample preparations and other methods, which might impede the analysis and must be better addressed in the manuscript.

*We have added several new experiments and analyses to address these concerns. Focusing on SMAD3, we conducted new experiments characterizing the effects of age and HTT CAG length on SMAD3 transcript and protein levels. Together with our ChIP-seq experiment confirming an association between SMAD3 binding and HTT CAG-length dependent gene expression changes, as well as previous work linking the SMAD3 agonist TGF-beta to Huntington's disease, these data provide a compelling case for a biological connection of SMAD3 signaling in HD. In the discussion (page 10-11) we highlight literature that supports roles in HD for several TFs that were found to be core TFs in our analysis, including RARB, FOXO1, KLF16, and RXRG. In response to the reviewer's final comment about differences in mouse models, ages etc. we have added an additional sentence in the discussion (page 10, lines 40-43). In our view, observing enrichments for differential expression for the same TFs across multiple time points and mouse models suggests that the findings are robust. However, we agree with the reviewer that a more refined view of network dynamics will require careful attention to the mouse models being used.*

1. The authors claim in the abstract to provide a "hierarchy of transcription factor drivers in HD". Where is this addressed in the manuscript? It is unclear.

*We thank the reviewer for this point and have removed this phrase from the manuscript.*

2. Page 4 (lines 11-13): Since none of the TF drivers have been experimentally validated one should phrase "potential transcription factor drivers..." Otherwise, it is overstated.

*We have revised the language describing "potential transcription factor drivers" to "predictions" of transcription factor drivers throughout the manuscript.*

3. Pages 6 and 7, the core TFs in independent datasets are not presented and discussed properly, for example only 22 of 48 TFs were enriched in protein samples and 13 of 48 core TFs were found in human samples. Please discuss this low fractions whether it stems from the analysis or from the samples (late human HD versus early onset mouse HD, ...).

*We thank the reviewer for this comment. We have revised the section describing the datasets and hope that this has improved its clarity. For the comparison to four independent microarray datasets in mouse models of HD, we found that targets of 46 of the 48 core TFs were enriched for DEGs, an overlap which is significantly greater than expected by chance (Fisher's exact test:  $p = 5.7e-32$ ). We also found in a comparison with protein quantity data that targets of 22 of the 48 core TFs were enriched for differentially abundant proteins, an overlap which is significantly greater than expected by chance (Fisher's exact test:  $p = 5.7e-20$ ). We revised the paragraph describing the comparison to late-stage human HD and address that specific comment below.*

4. Page 7 line 12 "This overlap was not statistically greater than expected by chance" and then the parameters were adjusted to obtain statistically significant differences? It sounds that the human data doesn't correlate with the mouse data and the conclusions are overstated (lines 19 -22).

*We thank the reviewer for this point. We chose to perform two tests in which we considered either a restrictive set of TFs from the HD mouse models (the 48 core regulators), as well as a broader set of TFs (all 209 TFs whose predicted target genes were enriched in at least one condition from our primary mouse RNA-seq dataset). We state that for the overlap of 13 of 48 TFs, a p-value of 0.05 is not statistically significant (page 6, lines 9-10). We also reiterate that we are comparing advanced or late stage human disease to early pathogenic states in mouse models.*

5. Why was SMAD3 selected? Figure 3B does not support SMAD3 to be a top candidate in human HD.

*We address the selection of SMAD3 on page 8, lines 9-15. Briefly, SMAD3 was predicted to be a driver of gene expression changes across 4 independent datasets, and SMAD3 targets were predominantly down-regulated in an age- and CAG length-dependent manner beginning relatively early. In our view, samples from late-stage human HD are not useful for ranking top candidates as drivers of disease progression, since human samples include only the tail end of the disease progression. The extensive cell death and other gross pathology make it difficult to interpret the changes in the human samples. We, like many others in the HD research community, are therefore most focused on identifying early drivers, which can be detected most directly in the mouse models.*

6. Why were four-month-old HttQ111 mice used and not 2, 6 or 10 months old for SMAD3 CHIP-seq? This is inconsistent and a repetition with age-matched samples is needed.

*We utilized 4 month old mice because we wanted to capture early changes in SMAD3 binding. This decision was made based both on the saturation of differential gene expression changes observed at six months in the allelic series study (Langfelder et al), and also based on our knowledge of gene expression changes in a dense time series study of Htt<sup>Q111/+</sup> mice that was recently published in Human Molecular Genetics (Ament et al 2017). Specifically, we were interested in capturing a time point at which the down-regulation of neuronal genes has begun – these are the gene expression changes that SMAD3 is predicted to regulate -- but prior to the onset of neuroinflammatory gene expression changes that occur slightly later. By six months, neuroinflammatory gene expression changes are apparent in the Q111, Q140, and Q175 mouse models.*

7. Why were HttQ111 mice chosen and not Q140 or Q175 that have more DEGs at 6 and 10 months?

*A motivating factor of our study was to identify very early changes in the transcriptome that are also representative of human disease. For this reason, we prefer to use the shortest CAG-repeat length that allows us to detect early differences (Htt<sup>Q111/+</sup> mouse model). Additionally, many phenotypic studies are based in the Q111 mouse model, making it easier to understand transcriptional changes as they relate to other phenotypes.*

8. What is the biological and functional link between SMAD3 and the other core TFs with mutated HTT protein? An experimental validation for some of the core TFs is needed to support the in silico model.

*We thank the reviewer for this question. We agree that understanding the biological and functional link between the core TFs and mutated HTT protein is a promising future direction. On page 10, lines 32-43, we frame the limitations of our study, and suggest that the replication of 13 core TF predictions across four independent datasets sets up a clear set of hypotheses for follow-up studies. In an effort to further understand the protein-level changes of SMAD3, we looked at protein abundance differences in SMAD3 and phosphophorylated-SMAD3 in the striatum of Htt<sup>Q111/+</sup> mice compared to controls. This is included as an additional supplemental figure (page 33).*

Thank you very much for your consideration.

Sincerely,

Nathan Price



Nathan Price  
Professor & Associate Director  
Institute for Systems Biology  
Seattle, WA  
401 Terry Ave N  
Seattle, WA 98109



(206) 732-1204

[nprice@systemsbiology.org](mailto:nprice@systemsbiology.org)

Thank you again for submitting your work to Molecular Systems Biology. We are now globally satisfied with the modifications made and we will be able to accept your paper for publication in Molecular Systems Biology pending the following minor modifications:

#### # Figures

- The figure files are still too small. Please change the resolution to 300PPI/inch \*at the final size of the image\* (please see [http://www.embopress.org/sites/default/files/EMBOPress\\_Figure\\_Guidelines\\_061115.pdf](http://www.embopress.org/sites/default/files/EMBOPress_Figure_Guidelines_061115.pdf))
- The figures should only appear once, please remove the duplicates from the ms file.
- Figure 2 does not need to be named 2A since there is only one panel.
- For figure panels showing network visualization, we would encourage you to supply the corresponding zipped Cytoscape .cys files as "Figure Source Data Files" such that readers can download these files directly from the figures.

#### # Callouts

- All the main figure panels should be called out. Please add the missing ones: Fig 1E, Fig 3A-B, Fig 4A-D, Fig 5A-D, fig 6G-I.
- If possible, it would be better to call out Appendix figures from the main text. Please add callouts to Appendix figure S5 and S8.
- The callout to Appendix fig S2 appears before S1 (ideally they should be in numerical order)
- There is a callout to Appendix Dataset 1 on page 4, line 27, to Appendix Dataset 2 on page 5, line 22 and to Dataset 3 on page 9, line 8. Please update all Dataset callouts to "Dataset EV1-3" and please name the corresponding files Dataset EV1-3 (also see below).

#### # Datasets

- Mouse Striatum TRN Edges, TF to target gene module enrichments for reference datasets, TF module enrichment across 5 tissues: Please rename them Dataset EV1-3, add legends in separate tabs and update the callouts in the text.

#### # Supplementary information

- Please rename the figures and tables from 'SI Figure/Table [n]' to 'Appendix Fig. /Table S[n]'. Apologies for being picky, but The "S" for Fig. Sn is still missing, both in the Appendix and the callouts in the main ms.

Thank you for submitting this paper to Molecular Systems Biology.

Thank you for tentatively accepted our manuscript, "Transcriptional regulatory networks underlying gene expression changes in Huntington's disease" pending the following minor modifications. Listed below are the amendments we have made to the text and figures you requested.

#### # Figures

- The figure files are still too small. Please change the resolution to 300PPI/inch \*at the final size of the image\* (please see [http://www.embopress.org/sites/default/files/EMBOPress\\_Figure\\_Guidelines\\_061115.pdf](http://www.embopress.org/sites/default/files/EMBOPress_Figure_Guidelines_061115.pdf))

*We have increased the resolution of the figures and uploaded new versions.*

- The figures should only appear once, please remove the duplicates from the ms file.

*We have removed the figures from the manuscript text file.*

- Figure 2 does not need to be named 2A since there is only one panel.

*We removed the '2A' label and updated this figure to a higher resolution version.*

- For figure panels showing network visualization, we would encourage you to supply the corresponding zipped Cytoscape .cys files as "Figure Source Data Files" such that readers can download these files directly from the figures.

*We have included the .cys file.*

#### # Callouts

- All the main figure panels should be called out. Please add the missing ones: Fig 1E, Fig 3A-B, Fig 4A-D, Fig 5A-D, fig 6G-I.

*We have added callouts in the manuscript text for Fig 1E (page 4, line 27), Fig 3A-B (page 5 line 29 and 39), Fig4A-D (page 7 line 7), and Fig 6G-I (Page 9 line 40). We amended the description of Fig 5, and it is called out simply as Fig 5 in the text.*

- If possible, it would be better to call out Appendix figures from the main text. Please add callouts to Appendix figure S5 and S8.

*We have added callouts for S5 on page 5 line 20 and S8 on page 8 line 30 and 32.*

- The callout to Appendix fig S2 appears before S1 (ideally they should be in numerical order)

*We have reordered these two figures.*

- There is a callout to Appendix Dataset 1 on page 4, line 27, to Appendix Dataset 2 on page 5, line 22 and to Dataset 3 on page 9, line 8. Please update all Dataset callouts to "Dataset EV1-3" and please name the corresponding files Dataset EV1-3 (also see below).

*We have renamed the datasets and updated the callouts from (for example) "Appendix Dataset 2" to Dataset EV2.*

#### # Datasets

- Mouse Striatum TRN Edges, TF to target gene module enrichments for reference datasets, TF module enrichment across 5 tissues: Please rename them Dataset EV1-3, add legends in separate tabs and update the callouts in the text.

*We have renamed the three datasets and updated the callouts in the text.*

#### # Supplementary information

- Please rename the figures and tables from 'SI Figure/Table [n]' to 'Appendix Fig. /Table S[n]'. Apologies for being picky, but The "S" for Fig. Sn is still missing, both in the Appendix and the callouts in the main ms.

*We have updated the Appendix (for example) from Figure 1 to Figure S1.*

*We have updated the callouts in the manuscript to match the appendix names.*

We hope this completes the modifications necessary for the publication of our manuscript in Molecular Systems Biology. Thank you for working with us to improve the paper 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: Nathan D Price  
Journal Submitted to: Molecular Systems Biology  
Manuscript Number: MSB-16-7435

#### 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  $n \leq 5$ , 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  $\chi^2$  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?	No formal power analysis was conducted. However, we selected the largest available dataset (n=208 mice) to reconstruct the gene regulatory network described in our paper, and we used cross-validation to evaluate the robustness of our model. An advantage of network based approaches is that they often reduce the number of hypotheses that are tested in genomics experiments, in this case to 718 networks rather than ~20,000 genes. See p. 12, "Regression-based transcriptional regulatory network models." p. 13, "Enrichments of TF-target gene modules for differentially expressed genes."
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	The ChIP-seq experiment was performed with two independent replicates per group, with each replicate containing pooled ChIP DNA from three mice. Replicate samples were used to identify reproducible peaks and to perform exploratory analyses of differences between groups. See p. 14, "High resolution X-ChIP-seq."
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Sample hierarchical clustering and multidimensional scaling were used to detect outliers in the RNA-seq data. No outliers were detected or removed. See p. 12, "Regression-based transcriptional regulatory network models."
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.	Mice with HD mutations and wildtype controls were co-housed. Animals were selected for experiments at random except with respect to matching by age and home cage. See p. 14, "Mouse Breeding, Genotyping, and microdissection".
For animal studies, include a statement about randomization even if no randomization was used.	See above.
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.	This paper primarily reports the results of unbiased systems biology analyses, which were designed to evaluate all possible hypotheses about the system being studied. As there is inevitably some bias introduced during exploratory analyses of the primary dataset, we replicated all the key findings in independent samples. These independent samples were used solely for the purpose of the replication study, and no exploratory analyses were performed using the replication samples. See p. 13, "Enrichments of TF-target gene modules for differentially expressed genes."
4.b. For animal studies, include a statement about blinding even if no blinding was done	Researchers were blinded to the genotype of animals during rearing. Researchers were not blinded to the genotypes of animals during the preparation of ChIP DNA.
5. For every figure, are statistical tests justified as appropriate?	Yes. See page 16, "Figure Legends".
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	For network reconstruction, we performed linear modeling on RNA-seq data that were normalized using the RPKM method. During the initial formulation of our model, we evaluated a number of alternative normalization procedures and evaluated the distributions of expression values for some genes. The methods described in the paper produced the most robust results. For statistical analyses of differential expression, we used statistical approaches specifically designed for the analysis of RNA-seq, microarray, and ChIP-seq data, respectively. Specifically, since RNA-seq and ChIP-seq counts data have been shown to fit a negative binomial distribution, we used the edgeR software, which uses a negative binomial generalized linear model. Microarray gene expression data are often log-normal. We performed data normalization and linear models to evaluate expression changes in log-transformed data, using the affy, RMA, and limma R packages. For key results, we supplemented these parametric approaches with permutation-based methods, which make fewer assumptions about the shape of the data. See p. 13, "Enrichments of TF-target gene modules for differentially expressed genes."

#### 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-tur>

<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://ijb.biochem.sun.ac.za>  
[http://oba.od.nih.gov/biosecurity/biosecurity\\_documents.html](http://oba.od.nih.gov/biosecurity/biosecurity_documents.html)  
<http://www.selectagents.gov/>

Is there an estimate of variation within each group of data?	Yes. For variance estimates related to network reconstruction, see Fig. 1b. We report z-scores to summarize the variance in the expression of individual genes (e.g., Fig. 6a,f).
Is the variance similar between the groups that are being statistically compared?	Yes. Multidimensional scaling suggests that there are no major differences in variance between the groups, at least with respect to the major variance components. See p. 12, "Regression-based transcriptional regulatory network models."

### C- Reagents

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).	Page 14, Lines 24-36. Abcam Anti-SMAD3 antibody ab28379 [ChIP grade]; Anti-RNA polymerase II CTD repeat YSPSPS antibody [8WG16] [ChIP Grade] ab817
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

\* 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.	Page 14, Lines 10-22. Mouse Breeding, Genotyping, and microdissection. The B6.HttQ111/+ mice (Strain 003456; JAX) used for the ChIP-seq study have a targeted mutation replacing a portion of mouse Htt (formerly Hdh) exon 1 with the corresponding portion of human HTT (formerly IT15) exon 1, including an expanded CAG tract (originally 109 repeats). Mice used in the present study were on the C57BL/6J inbred strain background. The targeted Htt allele was placed from the CD-1 background onto the C57BL/6J genetic background by selective backcrossing for more than 10 generations to the C57BL/6J strain at Jackson laboratories. Cohorts of heterozygote and wild-type littermate mice were generated by crossing B6.HttQ111/+ and B6.Htt+/+ mice. Male mice were sacrificed at 122 ± 2 days of age (or 16 weeks) via a sodium phenobarbital based euthanasia solution (Fatal Plus, Henry Schein). Both hemispheres of each animal's brain was microdissected on ice into striatum, cortex, and remaining brain regions. These tissues were snap frozen and stored in -80°C.
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Page 14, line 21. Experiments were approved by an institutional review board in accordance with NIH animal care guidelines.
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.	We confirm compliance.

### 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	Page 15, lines 7-12. SMAD3 and RNA Pol2 ChIPseq data deposited in GEO; accession #GSE88775 The following link has been created to allow review of record GSE88775 while it remains in private status (prior to publication): <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=oryzqeerzmdaf&amp;acc=GSE88775">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=oryzqeerzmdaf&amp;acc=GSE88775</a>
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).	NA
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: <b>Primary Data</b> 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 <b>Referenced Data</b> 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	Yes - included section "Referenced Datasets" Page 11 and section included "Software and Primary Data Resources" Page 15
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 Biocompare (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.	Page 15, lines 7-12. Code for analysis of gene expression data, modeling transcriptional regulatory networks, and analyzing ChIPseq data have been deposited in the public github repository located at <a href="http://github.com/seth-ament/hd-trn">http://github.com/seth-ament/hd-trn</a>

### 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
---	----