THE EMBO JOURNAL

Manuscript EMBO-2015-91206

NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program

Abhijeet Pataskar, Johannes Jung, Pawel Smialowski, Florian Noack, Federico Calegari, Tobias Straub and Vijay K. Tiwari

Corresponding author: Vijay K. Tiwari, Institute of Molecular Biology

Review timeline:

Submission date:	14 February 2015
Editorial Decision:	20 April 2015
Revision received:	28 July 2015
Editorial Decision:	11 August 2015
Revision received:	31 August 2015
Accepted:	14 September 2015

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.)

Editor: Anne Nielsen

1st Editorial Decision 20 April 2015

First of all - thank you very much for your patience here and my apologies for the very unusual delay in the review process for your manuscript!

Thank you again for submitting your study for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below. As you will see from the reports, all referees express interest in the findings reported, however they also a raise a number of concerns - largely related to the data description and analysis - that will have to be addressed in full before they can support publication of the manuscript.

For the revised manuscript I would particularly ask you to focus your efforts on the following points:

-> Please Include data to provide in vivo validation for the NeuroD1 targets identified through the analysis presented here (ref #3, major point)

-> Please revise the data presentation and description of experimental conditions used (as both refs #2 and #3 point out the need for significant revision on this point).

-> Please also strengthen the quantitative aspects of the study to support the significance of your findings (as extensively outlined by ref #2).

In addition to the three referees, we also had input from an additional expert advisor who raised the following points and suggested these to be addressed in a revised manuscript:

'First, ES cells like to make neurons. Sox2, a pluripotency factor, is involved in the neurogenic lineage. So I feel it is a stretch to say that NeuroD1 is sufficient to reprogram cells to neurons. It can direct pluripotent cells to do so, but seems likely to be working with several of the pluripotency factors that make ES cells so happy to generate neurons. The authors are not converting fibroblasts to neurons here. This is a major point where the ms needs reworking.

Second, the method of comparing bound vs not bound sites (by NeuroD) to other chromatin features, in Fig. 4A, is not a standard method that I have seen. Furthermore, it seems that conclusions are drawn from this in a qualitative way, not quantitatively (e.g. with P value comparisons of the metadata). I realize that the authors do qPCR ChIP assays to further assess these relationships, but this part of the paper seems qualitative to me. Perhaps they need to better explain the plots in 4A or repeat the analysis more quantitatively. '

Given the overall positive recommendations from both referees and advisor, I would like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE COMMENTS

Referee #1:

first I apologize for being late with my review. In this manuscript Tiwari and colleagues test the possibility that neuroD1 act as a pioneering factor able to bind to heterochromatic regions and activate a specific set of genes leading to neuronal differentiation. They test this by inducing the expression of neuroD1 in ES cells and check all of their results in the more natural situation of neuronal differentiation. The work is of high technical quality with appropriate controls. It shows that neuroD1 indeed acts as a pioneering factors directly inducing a number of neuronal genes (and transcription factors that in turn activate neuronal genes). It addition it supresses non neuronal genes. The activation is convincingly demonstarted by following chromatin changes in time in particular H3K27 Ac versus Me3. After establishing this they do the final check by looking at the inactivation of pluripotency genes and the induction of neuroD1, which fits their earlier observations. I am very positive and think it could be published directly without quibbling about some unimportant aspects.

Referee #2:

The manuscript by Pataskar et al examines the downstream transcriptional response of NeuroD1 in the neurogenic program. Overall the data itself looks of high quality, the two key issues are really in the data analyses and how the results are presented. There is a general lack of description about the data itself and how it was analyzed and second the strong wording (over interpretation) in the text

Data: The description of the data itself is vague at best.
Pg 7 - How many ChIP replicates of NeuroD1 were performed? How many peaks were called?

Give percentages in the text for how many peaks are at the promoters - It currently states "NeuroD1 was preferentially bound to promoters". Figure 2C shows 341 peaks - The reader is left to figure out

for themselves that this is actually only about 14% of all peaks. There are 3 times more in introns and almost 4 times more in intergenic regions - Promoter binding is hardly preferential

Page 7 - in referencing papers for H3k27ac the authors omitted two important ones (both published before Zhu et al) - PMIDs: 22231485, 21160473

The entire global analysis of NeuroD1 binding at enhancers is very vague.
- How were enhancers defined? What distance from the promoters? Is this based on chromatin marks? If so, these should be called "potential or predicted enhancers" or something to that sort e.g. pg 11 "Having observed NeuroD1 also targeted a large number of distal enhancers" No information is given about the definition of distal and these are putative enhancers. They are not enhancers until they have been shown to function as an enhancer in an embryo
- Were all annotated and non-annotated TSSs excluded? H3K27ac is also at active promoters.
- How were the enhancers linked to their putative target gene? (What distance, etc),
This is very important for the point below

"Comparing NeuroD1 binding with transcritomic changes induced its expression, we found that a substantial fraction of genes that were directly bound by NeuroD1 at their promoter, enhancers or both was significantly unregulated (n=252)"
Is this 252 from the 2300 genes with differential expression? Or 252 from the genes associated with NeuroD1 binding

Pg 9 "...NeuroD1 induced genes encompass a large number of genes that are unregulated during EMT..." How many? Give a percentage and the number x/x

When looking for other proteins and chromatin marks that are colocalized with NeuroD1, the authors used published data from ES cells. The subsequent analysis of this is not at all clear. Why do you need a Bayesian model to simply look for enrichment?
What advantage does this model give?
In the description of the results - it predicted 9 factors that "distinguish between NeuroD1-bound and non-bound promoters" Give the numbers. How many bound (341?)
How many unbound (the rest of the genome?)
The model could over fit with such an unbalanced scheme

The way these results are presented in Fig 4A and 5A is really not intuitive.
Why are some panels (Tbx3 in Fig 4A) shown with lines and others like UTF1 as scatter plots?
What do the contour plots bring over averograms?

Pg 10 "The chromatin at NeuroD1 target promoters was substantially less accessible compared with that of non-target promoters. These findings suggested that in the absence of NeuroD1, its target promoters are repressed by distinct factors" ?? But in the presence of NeuroD1 the promoters are less accessible (presumably repressed). This point seems to contradict itself.

"....after NeuroD1 was expressed, we performed a ChIP assay for Tbx3 , H3k27me3 and H3k27ac...." How long was NeuroD1 expressed? This is important given the later time-course experiments.

Text: There is an overall tendency in the text, especially in the abstract, to use to over claim what the results show. The wording is too strong. A couple of examples;
'NeuroD1 is able to bind its target sites at heterochromatic promoters and causes the loss of PcG associated recessive mark.
It's associated with the loss of H3K27me3. What actually causes that loss is another question. Clearly it cannot be NeuroD1 directly. If NeuroD1 triggers a cascade of events, remains to be determined

"..gain of the active mark H3K27ac, as well as increased chromatin accessibility, resulting in gene expression". Again, these are associated with gene expression - but there is no causal evidence that H3K27ac is required for, or results in, gene expression.
The two are certainly linked.

In both the abstract and text (pg 12) it states that NeuroD1 targeting to intergenic sites "is both necessary and sufficient to activate enhancers". There is not enough supporting evidence for this statement at several levels.
First, necessary - not strictly, as without NeuroD1 (Dox-) there is already high relative enrichment of H3K27ac - 0.5 before Dox, and only reaching 0.7ish after Dox (Fig 5I-K).
Second, Sufficiency - Certainly there is a modest (2-fold on average) increase of H27ac after Doc (Fig 5B). But the levels don't really increase in the time course and don't correlate particularly well with the maximum expression of NeuroD1 at 12h
Third, '..to active enhancers' - the results show that there is a link to NeuroD1 and a change in chromatin state at enhancers, which are correlated with active enhancers.
You haven't show that in NeuroD1 mutants these enhancers are inactive.

In the experiments using 'in vivo neuronal differentiation" the system is not described. Is this primary cells? If so this is ex-vivo or simply just call it primary neurons

"our comprehensive findings uncovered the entire gene regulatory program through which NeuroD1 specifies the neuronal fate..." Do you really believe that? All genomic experiments are based on cut-offs. We can call them global or comprehensive, but no experiment will ever get 'the entire'. If slightly different culture conditions were used, or if the experiment was truly done in vivo in developing mouse embryos, the genes and peak lists would change, even if by only a small percentage

Referee #3:

This work characterizes the transcriptional response of neural stem cells to NeuroD1 overexpression and its binding sites. The authors found genes enriched in 'neurogenic factors' supporting the notion that NeuroD1 alone is sufficient to initiate a neurogenic program. More interestingly, NeuroD1 binds regulatory regions (enhancers) that in its absence are in a 'closed' chromatin state making them 'open' and allowing the epigenetic, long-term expression of the factors themself.

This is an interesting, novel and well-conducted study. The manuscript itself is well written and presented (except minor points, below). The data seem robust and supportive of the authors' conclusions but there is still one major issue that makes me uneasy, which is that it is entirely performed in a rather 'artificial' system combining cultured cells and ectopic over-expression. Admittedly, the cell line used is probably as good as it gets in culture and many experiments could not have been done otherwise. Yet, particularly for the high ranking of the journal chosen, I would expect some corroboration in vivo of at least some of the most critical findings.

For example:

With the use of a reporter line, the authors could have made ChIP-Seq (or ChIP-qPCR) on FAC-sorted NeuroD1+/- progenitors. Even without a reporter mouse, ChIP from total brains should allow the validation of its binding domains, and chromatin state, in physiological conditions. For the epigenetic effects of NeuroD1 manipulations, in utero electroporation and FAC-sorting could have been used, in principle both after overexpression or RNAi should a conditional NeuroD1-KO mouse not be available. The authors
could come up with alternative ideas about this, but the tools mentioned are available and in vivo confirmation of some of the key observations is important.

Minor points

1) The authors indicate the number of genes overexpressed after NeuroD1 induction that are also direct targets (e.g. Fig 2D). Yet, they make no comment on the opposite relationship, i.e. the number of down-regulated genes that are also direct targets. If their model on NeuroD1-mediated chromatin opening should be correct, I would expect that virtually no direct target should ever be down regulated since the primary effect of NeuroD1 may be that of opening the chromatin to make it more accessible for upregulation. Is this the case or can NeuroD1 also act as a repressor?

Moreover, is there a correlation in the fold-change (FC) among groups of direct vs. secondary targets? If the chromatin-effect of NeuroD1 is overall comparable for all targets, this may lead to a similar FC among direct targets while secondary targets should be more diverse. Is this the case?

2) I do not understand why binding domains are so different in Fig. 2N-O and what is the reason for comparing them. If the authors want to propose a binding domain for NeuroD1 they should validate it by luciferase assays. In the absence of this, that table seems rather useless.

3) The authors often use 'qualitative' terms in the text (e.g. the majority of; a substantial fraction of; a large number of; etc.). While in most - but not all - cases quantifications are provided in the figures it would be
more appropriate and simpler for the reader to use quantitative terms already in the text.

4) The authors validate the expression of only 1 target: Lzts1. Many more should be provided in a supplemental figure using ISHs from the Allen Brain Atlas, Genepaint, Eurexpress or similar resources. Also, these targets should be absent prior to NeuroD1 expression, is this the case? Genome-wide
automated ISH at different developmental stages is available with a click of a mouse and no pipetting involved.

5) In a few cases the authors mentioned data from in vivo studies where are these data coming from? A short description and citation should be provided. Also, if these come from different publications, have the authors normalized for total number of reads in each case?

6) The two statements in the introduction: 'was successfully used to reprogram' and 'cell fate specification involves' need proper reference(s).

7) Figures: i) It seems appropriate to add a model for the mechanism proposed. ii) Most labels are extremely small and impossible to read if not magnified many times. iii) No significance is given for the bars and box plots. iv) I believe that relative expression levels (e.g. 1D) should have
all controls set to 1 and the scale values be also 1 in all cases because these are a.u. and there is no reason for such extreme numbers ranging from 4,000 (1A) to 0.005 (1D). v) Enhancers can be added to 2C.

1st Revision - authors' response 28 July 2015

RESPONSES TO REVIEWERS

Input from the expert advisor:

We are highly thankful to the adviser for providing positive recommendations for our study. In addition, he/she also raised few points that we have attempted to address as follows:

Comment 1) 'First, ES cells like to make neurons. Sox2, a pluripotency factor, is involved in the neurogenic lineage. So I feel it is a stretch to say that NeuroD1 is sufficient to reprogram cells to neurons. It can direct pluripotent cells to do so, but seems likely to be working with several of the pluripotency factors that make ES cells so happy to generate neurons. The authors are not converting fibroblasts to neurons here. This is a major point where the ms needs reworking.

Authors' response: We fully understand the expert advisor's concern. However, a major foundation of this study was that NeuroD1 has been known to exhibit remarkable capacity to reprogram other cell types towards neurons in various contexts. For example, recently NeuroD1 alone was shown to be able to reprogram reactive glial cells into neurons *in vivo* (Guo et al., 2014) as well as human astrocytes

into glutamatergic neurons (Amamoto and Arlotta, 2014; Guo et al., 2014).

Having known that NeuroD1 has a strong neurogenic potential, we were tempted to uncover the gene regulatory circuitry through which NeuroD1 elicits neuronal fate and chose uncommitted embryonic stem (ES) cells as a model system to unravel epigenetic and transcriptional dynamics induced by NeuroD1 during specification of neuronal fate. In this study, we show that the ectopic induction of this bHLH factor in ES cells is able to induce neuronal differentiation within 48 hours even in the presence of LIF, a strong inducer of pluripotency (Fig 1). Interestingly, under these conditions the expression of three major pluripotency factors (Oct4, Klf4 and Nanog) was significantly reduced, with the exception of Sox2 (Fig 1E). Such maintained expression of Sox2 could be of physiological relevance since it is also known to be required for neurogenesis, as also mentioned by the expert advisor. We furthermore show that even a transient expression of NeuroD1, as also occurs during neurogenesis *in vivo,* is sufficient to successfully induce neurogenesis (Fig 7I-O). In addition, our timecourse analysis suggests that such program is initiated by NeuroD1 binding to regulatory elements of its target genes, followed by displacement of repressive factors and heterochromatin machinery and subsequently induction of gene expression (Fig 4 J-L and Fig 5J-L), a finding that was further validated *in vivo* (new figure: Fig 6A-D).

Motivated by the expert advisor's comment, we have now also tested the transcriptional consequences of overexpressing NeuroD1 in murine fibroblasts. We find that NeuroD1 is able to locate its genomic sites in fibroblasts (new figure: Supplementary Fig S4A) and induce target gene expression (new figure: Supplementary Fig S4B-G). Furthermore, such ectopic NeuroD1 expression was also able to upregulate expression of the established neuronal marker Tubb3 (new figure: Supplementary Fig S4H). Altogether, these observations suggest that NeuroD1 is able to bind its target sequences and elicit neuronal gene expression program irrespective of the cell type.

While having described these observations that suggest that NeuroD1 is able to bind and induce its targets irrespective of cell type and reprogram chromatin and transcription factor landscape to initiate neurogenic transcription program, we agree with the expert advisor that given the nature of bHLH factors to work in complexes (Jones, 2004) and observed epigenetic changes that follow upon NeuroD1 binding, our future study should aim to uncover other factors that participate in NeuroD1 function during physiological and forced neurogenesis.

Comment 2) Second, the method of comparing bound vs not bound sites (by NeuroD) to other chromatin features, in Fig. 4A, is not a standard method that I have seen. Furthermore, it seems that conclusions are drawn from this in a qualitative way, not quantitatively (e.g. with P value comparisons of the metadata). I realize that the authors do qPCR ChIP assays to further assess these relationships, but this part of the paper seems qualitative to me. Perhaps they need to better explain the plots in 4A or repeat the analysis more quantitatively.'

Authors' response: Bayesian model is a quantitative mathematical model and here we applied it to predict transcription factors and chromatin marks (from all published ES ChIP-seq datasets) that statistically significantly discriminate NeuroD1 targets from those of the non-targets from among the genes induced following ectopic expression of NeuroD1 (promoters in Fig 4A and enhancers in Fig 5A; also see Supplementary Fig 5A and 6A). Towards this, we screened public databases for all existing histone modification and transcription factor ChIP-seq datasets in ES cells and took only those for further analysis that passed our stringent quality criteria. We then used this dataset of 50 transcription factors (TFs) and 8 chromatin marks and generated a matrix of normalized (Trimmed Mean of M-values) tag counts. We then implemented Bayesian statistics, specifically Naïve Bayesian classifier, because of its proven robustness against outliers (Smialowski et al., 2010). The set of the most relevant (based on information gain coefficients) features were selected using best

first mathematical approach (Frank et al., 2004). Thus, the major advantage the Bayesian model provides in this case is to arrive at a list of factors (TF and chromatin features) from a high scale dataset, that robustly and with high significance classifies targets vs control regions. This automatable methodology is quicker and efficient in characterizing the chromatin and TF landscape than individual ChIP-seq enrichment comparison analysis.

To identify factors classifying NeuroD1 promoter induced targets vs control, we took the list of 88 promoter targets (Fig 2D) as the experimental set together with equal number of randomly selected non-target promoters of upregulated genes (URG) (n=88) as control. Similarly for analysis at enhancers, we took list of 169 target enhancers associated to upregulated genes (URG) as the experimental set and equal number of non-target enhancers associated with upregulated genes (URG) (n=169) as control. Using this analysis we uncovered transcription factors and epigenetic features that are enriched (e.g. TBX3, UTF1 and H3K27me3 at promoters; MBD3, TBX3 and H3K4me1 at enhancers) and depleted (e.g. H3K27ac and chromatin accessibility at promoters) at NeuroD1 target sites in the absence of NeuroD1 when these genes are inactive (Fig 4A and Fig 5A).

Indeed, we had also validated these observations by ChIP qPCRs which showed that these predicted transcription factors and epigenetic features are present at the silent NeuroD1 target genes and get remodeled following NeuroD1 binding, which causes transcriptional activation (Fig 4F-L, Fig 5F-L, Fig 7E-H and Fig 7M-N).

However, we highly agree with the expert advisor of the necessity of a more intuitive representation of this finding. Hence, we have now complemented the distribution plots (Fig 4A and Fig 5A) with boxplots for ChIP-seq enrichment of predicted factors and chromatin features at NeuroD1 target and non-targets sites (new figures: Fig 4B and Fig 5B). This analysis also clearly shows statistically significant differences, as marked by P-values, between NeuroD1 targets and non-target controls for the transcription factors and chromatin marks predicted previously.

We have now explained Bayesian model in detail in materials and methods section and included this new box plot in the figures (Fig 4B and 5B). We thank the expert advisor for motivating us to provide these additional analysis and clarifications which would definitely facilitate reader's appreciation of our findings.

Reviewer 1 (Remarks to the Author):
We thank the reviewer for many positive remarks and for the opinion that *"In this manuscript Tiwari and colleagues test the possibility that neuroD1 act as a pioneering factor able to bind to heterochromatic regions and activate a specific set of genes leading to neuronal differentiation. They test this by inducing the expression of neuroD1 in ES cells and check all of their results in the more natural situation of neuronal differentiation. The work is of high technical quality with appropriate controls. It shows that neuroD1 indeed acts as a pioneering factors directly inducing a number of neuronal genes (and transcription factors that in turn activate neuronal genes). It addition it supresses non neuronal genes. The activation is convincingly demonstarted by following chromatin changes in time in particular H3K27 Ac versus Me3. After establishing this they do the final check by looking at the inactivation of pluripotency genes and the induction of neuroD1, which fits their earlier observations"*. He/she continues with excitement for our study *"I am very positive and think it could be published directly without quibbling about some unimportant aspects"*. We are glad with such a high recommendation in support of our study and thank the reviewer for his/her support.

This reviewer did not raise any concerns that needed to be addressed in the revised version.

Reviewer 2 (Remarks to the Author):

The reviewer comments *"The manuscript by Pataskar et al examines the downstream transcriptional response of NeuroD1 in the neurogenic program.*" We thank the reviewer for the positive remark that, *"Overall the data itself looks of high quality."*

In addition, the reviewer had concerns on the data analyses and how the results are presented. He finds that there is a general lack of description about the data itself and how it was analyzed and second the strong wording (over interpretation) in the text. We have attempted to address each of these concerns as follows:

Comment 1) Data: The description of the data itself is vague at best. Pg 7 - How many ChIP replicates of NeuroD1 were performed? How many peaks were called?

Authors' response: We apologize to the reviewer for not making the data description properly visible. We derived ChIP-seq data for NeuroD1 from two biological replicates to identify its genomic targets. Using these data, we reproducibly called 2409 enriched peaks in both the biological replicates and their distribution was shown in Fig 2C (after genome size normalization). We have now updated the materials and methods and figure legends as applicable to further convey this information more comprehensibly.

Comment 2) Give percentages in the text for how many peaks are at the promoters - It currently states "NeuroD1 was preferentially bound to promoters". Figure 2C shows 341 peaks - The reader is left to figure out for themselves that this is actually only about 14% of all peaks. There are 3 times more in introns and almost 4 times more in intergenic regions - Promoter binding is hardly preferential.

Authors' response: We agree with the reviewer that there are 14% of the absolute number of NeuroD1 peaks that fall at promoters. However, this is significant, given the much smaller total genomic size promoter regions encompass in the genome as compared to exons, introns and intergenic regions. Given this, to appreciate patterns better, we and many other authors chose to show distribution of ChIP-seq peaks at various regions in the genome after normalization to the genomic size they constitute in the genome (Kraushaar et al., 2013; Thakurela et al., 2013; Tiwari et al., 2012). Similar genome size normalization analysis for NeuroD1 peaks showed the NeuroD1 peaks are most strongly enriched at the promoters compared to other regions in the genome (~42% after normalization as shown in the pie chart in Fig 2C) and thus we had used the word "preferential" to explain the pattern. Importantly, following reviewer's advice, we have now also included the percentage of the peaks in each of the genomic regions after this normalization step in figure legends. We thank the reviewer for these suggestions that would help readers better appreciate our findings.

Comment 3) Page 7 - in referencing papers for H3k27ac the authors omitted two important ones (both published before Zhu et al) - PMIDs: 22231485, 21160473.

Authors' response: We agree with the reviewer that these two are very important references and have now included them in the manuscript. We apologize for this mistake and thank the reviewer for pointing this to us.

Comment 4) The entire global analysis of NeuroD1 binding at enhancers is very vague.

- How were enhancers defined? What distance from the promoters? Is this based on chromatin marks? If so, these should be called "potential or predicted enhancers" or something to that sort e.g. pg 11 "Having observed NeuroD1 also targeted a large number of distal enhancers" No information is given about the definition of distal and these are putative enhancers. They are not enhancers until they have been shown to function as an enhancer in an embryo.

- Were all annotated and non-annotated TSSs excluded? H3K27ac is also at active

promoters.

- How were the enhancers linked to their putative target gene? (What distance, etc), This is very important for the point below.

Authors' response: Previous studies have already established that non-promoter H3K27ac mark closely predicts enhancers and thus has been frequently used to define enhancers (Bonn et al., 2012; Cheng et al., 2014; Creyghton et al., 2010; Kwiatkowski et al., 2014; Rada-Iglesias et al., 2011; Shlyueva et al., 2014; Zhu et al., 2013). Very recently, it was shown that the targeting of acetyltransferase p300 (that acetylates H3 at K27) to enhancers is sufficient to activate enhancers and the associated gene expression (Hilton et al., 2015). In line with these approaches, we took all intergenic regions at TN d1 (the phase of transition from neuronal progenitors to neurons in our *in vitro* neuronal differentiation system) that exhibit high H3K27ac enrichment at enhancer regions as this is the phase of neuronal differentiation where NeuroD1 is most highly induced as well as bound at the target regions, including intergenic sites (Fig 1B, Fig 2B, Fig 2D, Fig 2L-M). As mentioned by the reviewer, since H3K27ac also occurs at active promoters, we had excluded all regions from -800 bp to +200 from the UCSC mm9 annotated TSS for enhancer analysis as they fall under our definition of promoters. Next, among all TN d1 enhancers, to further shortlist potential enhancers regulated by NeuroD1, we took +/- 1000 bp from H3K27ac peak summit, calculated the NeuroD1 enrichment at these sites and took intergenic regions that exhibit high enrichment for both H3K27ac and NeuroD1 for further analysis of NeuroD1 function at enhancers (new table: Supplementary Table S6). Given the stringency of our approach, we have been able to discover a high confidence list of enhancers potentially regulated by NeuroD1. For example, most NeuroD1 target intergenic regions gain H3K27ac concomitant with NeuroD1 expression during neuronal differentiation (Fig 5D-E). Strikingly, we find that following ectopic expression of NeuroD1 in ES cells, it binds to these distal sites, which then gain H3K27ac followed by transcriptional induction of the associated genes (nearest gene approach) (Fig 2L and 5J-L). We also provide new data in the revised version of the manuscript where we show that these enhancer regions are also occupied by NeuroD1 in the embryonic cortex (new figure: Fig 6A). In addition, ectopic overexpression of NeuroD1 during cortical development induces gain of H3K27ac at these distal NeuroD1 target sites, which further accompanies upregulation in the transcription of the associated genes (new figure: Fig 6B-D). Last but not least, we also show that a significant number of identified NeuroD1 target enhancer sites overlap with previously published validated tissue-specific enhancers from E14.5 brain and cortex (Shen et al., 2012) (new figure: Supplementary Fig S3B). Altogether, these range of findings strongly argue that identified NeuroD1-target distal regions are enhancers that likely function in mediating NeuroD1-dependent gene expression program during neurogenesis.

We have linked the enhancers to their putative target genes by the nearest gene approach (without any distance cutoff) as used in many others studies (Creyghton et al., 2010; Lodato et al., 2013; Rada-Iglesias et al., 2011). This is based on the assumption that the enhancers most likely regulate the genes in closest proximity. In the absence of any genomewide long-range interaction dataset (e.g. Hi-C), which would be the most ideal way of identifying functional promoter-enhancer pairs in the genome, we considered the nearest gene approach as an alternate approach for addressing our questions. In support of the validity of our strategy, we find that a significant number of enhancers identified to be activated by NeuroD1 following its induction in ES cells also gained H3K27ac (Fig 5D-E) as well as associated with transcriptional induction of the nearest genes concomitant with NeuroD1 expression during neurogenesis *in vitro* and *in vivo* (Supplementary Fig S6D-l; new figure: Fig 6B-D, Fig 7E, Fig 7H).

We have updated the manuscript text and materials and methods to document these details more comprehensively. We thank the reviewer for these comments that have helped us better describe our approach and findings.

Comment 5) "Comparing NeuroD1 binding with transcritomic changes induced its expression, we found that a substantial fraction of genes that were directly bound by NeuroD1 at their promoter, enhancers or both was significantly unregulated (n=252)" Is this 252 from the 2300 genes with differential expression? Or 252 from the genes associated with NeuroD1 binding?

Authors' response: The ectopic expression of NeuroD1 in ES cells led to upregulation of 2209 genes, out of which 195 (previously 252, the change explained at the end of this paragraph) genes were directly bound by NeuroD1 at their regulatory elements (referred to as "URT" (Upregulated Target Genes) throughout the manuscript) (Fig 2D). This further suggests that the remaining upregulated genes might represent secondary consequences acquired following induction of the direct NeuroD1 targets and is supported by our observation that the induction of the secondary targets was lower compared to the direct NeuroD1 targets (new figure: Supplementary Fig S3A). These secondary changes could additionally be explained by the fact that many of the direct NeuroD1 targets are transcription factors (Fig 3AD), which, following induction may function in activation of the secondary target genes. Importantly, to retrieve a high confidence target genes that are under direct transcriptional control of NeuroD1, we had applied stringent cut-offs for expression change (at least 1.5 fold and statistically significant while correcting for multiple testing) and NeuroD1 binding (enrichment at least 0.75 and statistically significant) (please refer to materials and methods for details). In the current version these numbers are slightly altered for the enhancer targets as we decided to put number of unique gene identifiers associated to enhancer peaks rather than number of peaks for enhancer targets. We have now further improved the description in the manuscript to properly reflect these findings.

Comment 6) Pg 9 "...NeuroD1 induced genes encompass a large number of genes that are unregulated during EMT..." How many? Give a percentage and the number x/x.

Authors' response: We thank the reviewer for pointing this out. We have now provided the percentage and the number of genes overlapping between NeuroD1-induced genes and those upregulated during EMT in the text and figure legends. We have additionally provided significance calculation which shows that this overlap is statistically significant.

Comment 7) When looking for other proteins and chromatin marks that are colocalized with NeuroD1, the authors used published data from ES cells. The subsequent analysis of this is not at all clear. Why do you need a Bayesian model to simply look for enrichment?
What advantage does this model give?
In the description of the results - it predicted 9 factors that "distinguish between NeuroD1-bound and non-bound promoters" Give the numbers. How many bound (341?) How many unbound (the rest of the genome?) The model could over fit with such an unbalanced scheme.

Authors' response: We screened public databases for all existing histone modification and transcription factor ChIP-seq datasets in ES cells and took only those for further analysis that passed our stringent quality criteria. We then used this dataset of 50 transcription factors (TFs) and 8 chromatin marks and generated a matrix of normalized (Trimmed Mean of M-values) tag counts.

We implemented Bayesian statistics, specifically Naïve Bayesian classifier, because of its proven robustness against outliers (Smialowski et al., 2010). The set of the most relevant (based on information gain coefficients) features were selected using best first mathematical approach (Frank et al., 2004). Thus, the major advantage the Bayesian model provides in this case is to arrive at a list of factors (TF and chromatin features) from a high scale dataset, that robustly and with high significance classifies

targets vs control regions. This automatable methodology is quicker and efficient in characterizing the chromatin and TF landscape than individual ChIP-seq enrichment comparison analysis.

To identify factors classifying NeuroD1 promoter induced targets vs control, we took the list of 88 promoter targets (Fig 2D) as the experimental set together with equal number of randomly selected non-target promoters of upregulated genes (URG) (n=88) as control. Similarly for analysis at enhancers, we took list of 169 target enhancers of upregulated genes (URG) as the experimental set and equal number of non-target enhancers associated with upregulated genes (URG) (n=169) as control. Thus, this setting helped us to avoid overfitting, which, as pointed by the reviewer can occur if we take rest of the genome as control. Using this analysis we uncovered transcription factors and histone modifications that are enriched (e.g. TBX3, UTF1 and H3K27me3 at promoters; MBD3, TBX3 and H3K4me1 at enhancers) and depleted (e.g. H3K27ac and chromatin accessibility at promoters) at NeuroD1 target sites in the absence of NeuroD1 induction. We had also validated these observations by ChIP qPCRs which showed that indeed these predicted transcription factors and epigenetic features are present at the silent NeuroD1 target genes and get remodelled following NeuroD1 binding, which causes transcriptional activation (Fig 4F-L, Fig 5F-L, Fig 7E-H and Fig 7M-N).

Comment 8) The way these results are presented in Fig 4A and 5A is really not intuitive.
Why are some panels (Tbx3 in Fig 4A) shown with lines and others like UTF1 as scatter plots? What do the contour plots bring over averograms?

Authors' response: The plots in Fig 4A and Fig 5A are aimed to visualize correlation of TF and chromatin feature at NeuroD1 target sites prior to NeuroD1 binding. These plots help to appreciate that the model predicted factors are indeed classifying features given different densities and distributions in one and two dimensional space, which are further affirmed by a quantitative measurement of information gain, i.e. measure of importance of a feature towards classifying the datasets. There are three subsets of charts within the figures: 1) the diagonal plots display density of ChIP enrichment at NeuroD1 target promoters (shown in red) and control promoters (shown in cyan); 2) Upper right diagonal contour plots convey the same information in two dimensions (two features at a time); 3) Scatter plots in lower left diagonal show cross-correlation of identified factors in experimental and control dataset.

However, we agree with the reviewer that we need to represent our findings more intuitively. We have now also complemented the previous distribution plots (Fig 4A and Fig 5A) with boxplots for ChIP-seq enrichment of predicted factors and chromatin features at NeuroD1 target and non-targets sites (new figures: Fig 4B and Fig 5B). This analysis also showed statistically significant differences between NeuroD1 targets and non-target controls for the transcription factors and chromatin marks predicted by the Bayesian model. We thank the reviewer for motivating us to provide these additional analysis and clarifications which would definitely facilitate reader's appreciation of our findings.

Comment 9) Pg 10 "The chromatin at NeuroD1 target promoters was substantially less accessible compared with that of non-target promoters. These findings suggested that in the absence of NeuroD1, its target promoters are repressed by distinct factors" ?? But in the presence of NeuroD1 the promoters are less accessible (presumably repressed). This point seems to contradict itself.

Authors' response: We thank the reviewer for pointing this out and regret to not have conveyed the message properly. As shown in the Fig 4A, the Bayesian modeling revealed that in the absence of NeuroD1, its target sites are kept in a heterochromatic state and they show less accessibility. The new boxplots for ChIPseq enrichment of predicted transcription factors and chromatin features at NeuroD1

target and non-targets sites in the absence of NeuroD1 provide a much better conveying of the message and are in line with the findings from Bayesian modeling (new figure: Fig 4B). Following NeuroD1 induction, it binds its target sites, overrides repressive TF and chromatin landscape to induce a euchromatic state which accompanies gain of chromatin accessibility (Fig 4F-L, Fig 7D-G).

Comment 10) "....after NeuroD1 was expressed, we performed a ChIP assay for Tbx3 , H3k27me3 and H3k27ac...." How long was NeuroD1 expressed? This is important given the later time-course experiments.

Authors' response: We induced NeuroD1 in ES cells for 48 hours prior to ChIP assay for the described transcription factors and histone modifications, the same time point where we had performed RNA-seq assay following NeuroD1 induction. We have now described this in the main text in addition to the materials and methods section.

Comment 11) Text: There is an overall tendency in the text, especially in the abstract, to use to over claim what the results show. The wording is too strong. A couple of examples;
'NeuroD1 is able to bind its target sites at heterochromatic promoters and causes the loss of PcG associated recessive mark.
It's associated with the loss of H3K27me3. What actually causes that loss is another question. Clearly it cannot be NeuroD1 directly. If NeuroD1 triggers a cascade of events, remains to be determined.

Authors' response: We apologize to the reviewer for giving an impression of overstating the results. Since NeuroD1 has only a DNA binding domain and is known to bind DNA and interact with other proteins, it is almost certain that it cannot change the chromatin landscape itself on its own and likely recruits other factors and/or epigenetic regulatory proteins to its target sites for this purpose. To dissect the order of events during NeuroD1-dependent gene activation, we had performed timecourse experiments following ectopic expression of NeuroD1 which show that it binds to its targets within 6 hours of expression, followed by a loss of repressive transcription factor and heterochromatin landscape and a gain of H3K27ac and chromatin accessibility (Fig 4J-L; Fig 5J-L). In this revised version we now complemented this figure with the timecourse expression analysis of associated target genes which shows that such acquisition of a euchromatic state accompanies their transcriptional induction. These patterns are highly suggestive that NeuroD1 binding triggers changes in transcription factor and chromatin landscape at its target sites towards a transcriptionally permissive state which ultimately results in the expression of associated genes. Similarly, we could now also show that NeuroD1 binds its induction of H3K27ac as well as transcription following ectopic expression of NeuroD1 during cortical development, further validating our findings *in vivo* (new figure: Fig 6B-D). We have now also moderated our sentences and further clarified this in the manuscript.

Comment 12) "..gain of the active mark H3K27ac, as well as increased chromatin accessibility, resulting in gene expression". Again, these are associated with gene expression - but there is no causal evidence that H3K27ac is required for, or results in, gene expression.
The two are certainly linked.

Authors' response: We agree with the reviewer that one cannot be absolutely sure that H3K27ac occurrence at genomic regions causes transcriptional induction. We had proposed such claims based on existing literature that H3K27ac mark closely predicts active regulatory elements (Bonn et al., 2012; Cheng et al., 2014; Creyghton et al., 2010; Kwiatkowski et al., 2014; Rada-Iglesias et al., 2011; Shlyueva et al., 2014; Zhu et al., 2013). Very recently, it was also shown that the targeting of acetyltransferase p300 (that acetylates H3 at K27) to promoters and enhancers is sufficient for their activation and induction of gene expression (Hilton et al., 2015).

Our results very well comply with these existing evidences. For example, most NeuroD1 target promoter and enhancer regions gain H3K27ac concomitant with high NeuroD1 expression during neuronal differentiation (Fig 4D-E, Fig 5D-E) and this accompanies transcriptional activation of associated genes (Supplementary Fig S5DE, Supplementary Fig S6D-E). Importantly further, these NeuroD1 target regions lack H3K27ac in tissues from non-neuronal lineages where NeuroD1 is not expressed (Supplementary Fig S5H-I, Supplementary Fig S6H-I) and this associates with a transcriptionally silent state of these genes (Supplementary Fig S5F-G, Supplementary Fig S6F-G). Furthermore, following ectopic expression of NeuroD1, it binds to its target regions (Fig 2D, Fig 2K-L), which follows a gain in H3K27ac (Fig 4F, Fig 4J-L, Fig 5G, Fig 5J-L) and transcriptional induction of associated genes (Fig 2D, Fig 4J-L, Fig 5J-L). We also provide new data in the revised version of the manuscript where we show that NeuroD1 overexpression during cortical development induces gain of H3K27ac at its target sites which further accompanies upregulation in the transcription of the associated genes *in vivo* (new figure: Fig 6AD). Moreover, we now also show that a significant number of NeuroD1 target enhancer sites overlap with the previously published validated tissue-specific enhancers from E14.5 brain and cortex (Shen et al., 2012)(new figure: Supplementary Fig S3B). These comprehensive findings in our study relating H3K27ac with the induction of gene expression in combination with the existing literature strongly suggests that NeuroD1 targeting to distinct loci results in a gain of H3K27ac which then leads to the transcriptional induction of associated genes. The actual events that allow H3K27ac modified regulatory elements to be permissive for transcription has been studied extensively in last decades (Grunstein, 1997; Shlyueva et al., 2014; Struhl, 1998; Tessarz and Kouzarides, 2014; Verdone et al., 2005) . We have now further edited the text to explain these points.

Comment 13) In both the abstract and text (pg 12) it states that NeuroD1 targeting to intergenic sites "is both necessary and sufficient to activate enhancers". There is not enough supporting evidence for this statement at several levels.
First, necessary - not strictly, as without NeuroD1 (Dox-) there is already high relative enrichment of H3K27ac - 0.5 before Dox, and only reaching 0.7ish after Dox (Fig 5IK).
Second, Sufficiency - Certainly there is a modest (2-fold on average) increase of H27ac after Doc (Fig 5B). But the levels don't really increase in the time course and don't correlate particularly well with the maximum expression of NeuroD1 at 12h
Third, '..to active enhancers' - the results show that there is a link to NeuroD1 and a change in chromatin state at enhancers, which are correlated with active enhancers. You haven't show that in NeuroD1 mutants these enhancers are inactive.

Authors' response: Our statement that NeuroD1 targeting to intergenic sites is both necessary and sufficient to activate enhancers was based on several observations that we made during our study. Importantly, our time course analysis showed that within hours of its expression, NeuroD1 binds to the intergenic sites and these sites then gain H3K27ac (Fig 5J-L). Such gain of H3K27ac then accompanies transcriptional induction of associated genes (Fig 5J-L). In addition, during neuronal differentiation of ES cells, the identified NeuroD1 target intergenic regions gain H3K27ac only after NeuroD1 is expressed (Fig 5D-E). Importantly further, in the revised manuscript we provide evidence that NeuroD1 overexpression during cortical development induces gain of H3K27ac at the NeuroD1 target intergenic regions which accompanies upregulation in the transcription of the associated genes (new figure: Fig 6B-D). Lastly, we now also show that a significant number of our NeuroD1 target enhancer sites overlap with the previously published validated tissue-specific enhancers from E14.5 brain and cortex (Shen et al., 2012)(new figure: Supplementary Fig S3B). These findings together with the existing literature suggesting that H3K27ac deposition serves as a guiding cause for a transcription activating state and predicts active enhancers with high accuracy (Bonn et al., 2012; Cheng et al., 2014; Creyghton et al., 2010; Kwiatkowski et al., 2014; Rada-Iglesias et al., 2011; Shlyueva et al., 2014; Zhu et al., 2013) and a recent finding that the targeting of acetyltransferase p300 (enzyme that places an acetyl group on H3K27) to distal regions is sufficient to activate enhancers and induce gene expression

(Hilton et al., 2015) would imply that indeed targeting of NeuroD1 to intergenic regions is the trigger for H3K27ac-mediated activation of these enhancer sites.

With reference to the reviewers question on an already high enrichment of H3K27ac without NeuroD1, we apologize to not have explained enough about how these data are plotted, which led to this confusion. For better visualization, the relative enrichment in the mentioned timecourse ChIP plots (Fig 5J-L) was calculated based on the highest enrichment set to 1 in each timecourse replicate. We further normalized the remaining enrichment values to this highest enrichment and plotted these values as relative enrichment. We agree that these overall increases are not massive, but are statistically significant, as calculated in another plot for similar data (Fig 5G). The lesser induction could be explained by various possible reasons. For example, the induction of NeuroD1 is not homogenous in ES cells, where certain cells express high levels of NeuroD1 and other low. ChIP assay, being done on the total population, would provide average data for such heterogeneous populations. Furthermore, ChIP analysis was done at the maximum time point of only 48 hours post-induction of NeuroD1 and it is likely that analysis after a longer period gives a higher enrichment for H3K27ac.

We agree that for one of the examples shown, i.e. the Ncam1 enhancer, the changes in H3K27ac levels are not as prominent, but the additional examples shown in Fig 5JL show the stated behavior. In this revised version of the manuscript, we have now included the expression dynamics of the nearest gene of the exemplified enhancers during the time course analysis of NeuroD1 binding and H3K27ac enrichment following NeuroD1 expression (Fig 5J-L) which shows that following NeuroD1 binding, H3K27ac levels rise at these intergenic sites and then the transcription of the associated genes is induced (maximally in between 18-48 hours) (Fig. 5J-L). It is easily conceivable that the fine kinetics of chromatin dynamics at individual loci following NeuroD1 targeting would slightly differ based on multiple local features such as sequence composition, epigenetic landscape, regulatory factors occupancy and higher order chromatin conformation which are known to differ between genes.

With regard to the last comment from the reviewer that we do not have any analysis on NeuroD1 mutants to show that these enhancers are inactive, we have now attempted to address this question in a different way as we did not have an immediate access to NeuroD1 mutant mice. A ChIP assay for NeuroD1 in the embryonic cortex showed that that targets we identified, including enhancers, were also bound by NeuroD1 *in vivo* (new figure: Fig 6A). To functionally test the impact of NeuroD1 on its targets during neuronal development, we decided to overexpress NeuroD1 during cortical development and assessed its effects on H3K27ac at identified NeuroD1 target regulatory elements as well as expression of the associated genes. Towards this, we performed *in utero* electroporation assay in mouse cortex at E13.5 with either an empty or NeuroD1 expression vector (vector also contains an IRES-RFP element) as described previously (Aprea et al., 2013; Saito, 2006). The animals were sacrificed at E15.5 for FAC-sorting of RFP positive cells for further analysis (new figure: Fig 6B). Interestingly, such overexpression of NeuroD1 led to a significant increase in H3K27ac levels at NeuroD1 target sites (both promoters and enhancers) as compared to the control (new figure: Fig 6C). Furthermore, this also accompanied a significant transcriptional induction of associated genes (new figure: Fig 6D). These *in vivo* observations together with our similar previous findings indeed establish that NeuroD1 targeting to distinct regulatory elements induces a euchromatic state which results in transcriptional activation of associated genes.

We have also discussed these points in the revised version of the manuscript. We thank the reviewer for motivating us to do these edits and experiments that have clearly made highly relevant additions to the manuscript and increased the impact of our findings.

Comment 14) In the experiments using 'in vivo neuronal differentiation" the system is

not described. Is this primary cells? If so this is ex-vivo or simply just call it primary neurons.

Authors' response: We apologize for this mistake in the text where we wrote "*in vivo* neuronal differentiation" in the section regarding our newly developed iTN system. Here we had referred to the *in vitro* neuronal differentiation protocol where we derive neurons from ES cells. We have now changed the wordings in the revised version of the manuscript and thank the reviewer for pointing this to us.

Comment 15) "our comprehensive findings uncovered the entire gene regulatory program through which NeuroD1 specifies the neuronal fate..." Do you really believe that? All genomic experiments are based on cut-offs. We can call them global or comprehensive, but no experiment will ever get 'the entire'. If slightly different culture conditions were used, or if the experiment was truly done in vivo in developing mouse embryos, the genes and peak lists would change, even if by only a small percentage.

Authors' response: We fully agree with the reviewer and we have now modified the text accordingly as advised. We thank the reviewer for this suggestion.

Reviewer 3 (Remarks to the Author):
We thank the reviewer for many positive remarks and for the opinion that, *"This work characterizes the transcriptional response of neural stem cells to NeuroD1 overexpression and its binding sites. The authors found genes enriched in 'neurogenic factors' supporting the notion that NeuroD1 alone is sufficient to initiate a neurogenic program. More interestingly, NeuroD1 binds regulatory regions (enhancers) that in its absence are in a 'closed' chromatin state making them 'open' and allowing the epigenetic, long-term expression of the factors themself."* He/she continues with excitement for our study *"This is an interesting, novel and well-conducted study. The manuscript itself is well written and presented. The data seem robust and supportive of the authors' conclusions".*

In addition, the reviewer had one major comment and several minor comments that we have addressed as follows:

Comment 1) there is still one major issue that makes me uneasy, which is that it is entirely performed in a rather 'artificial' system combining cultured cells and ectopic over-expression. Admittedly, the cell line used is probably as good as it gets in culture and many experiments could not have been done otherwise. Yet, particularly for the high ranking of the journal chosen, I would expect some corroboration in vivo of at least some of the most critical findings.
For example: With the use of a reporter line, the authors could have made ChIP-Seq (or ChIP-qPCR) on FAC-sorted NeuroD1+/- progenitors. Even without a reporter mouse, ChIP from total brains should allow the validation of its binding domains, and chromatin state, in physiological conditions. For the epigenetic effects of NeuroD1 manipulations, in utero electroporation and FAC-sorting could have been used, in principle both after overexpression or RNAi should a conditional NeuroD1-KO mouse not be available. The authors could come up with alternative ideas about this, but the tools mentioned are available and in vivo confirmation of some of the key observations is important.

Authors' response: We fully agree with the reviewer that we must validate some of the critical findings *in vivo* and we have followed along the suggested lines. Towards this, we first performed ChIP assay against endogenous NeuroD1 in murine E14.5 cortex and tested its binding at the target sites we identified *in vitro*. This analysis showed that all tested targets (both promoters and enhancers) were occupied by NeuroD1 in the embryonic cortex arguing that our approach has identified authentic genomic sites bound by NeuroD1 during cortical development (new figure: Fig 6A).

We next attempted to functionally test the impact of NeuroD1 on the epigenetic landscape of its targets during neuronal development. While discussing with our collaborators, we felt that the shRNA-mediated knockdown during corticogenesis could accompany potential side effects including apoptosis, off-target effects etc. and would require much more analysis to raise confidence on any findings. Therefore, we decided to overexpress NeuroD1 during cortical development and assess its effects on H3K27ac levels at identified NeuroD1 target sites as well as expression of associated genes. Towards this, we performed *in utero* electroporation assay in mouse cortex at E13.5 with a plasmid expressing NeuroD1 (NeuroD1-IRES-RFP) or the same backbone without an inserted CDS as control (Control-IRES-RFP) as described previously (Aprea et al., 2013; Saito, 2006). We sacrificed the animals at E15.5 and FAC-sorted for RFP positive cortical cells for further analysis (new figure: Fig 6B). Interestingly, such overexpression of NeuroD1 led to a significant increase in H3K27ac levels at NeuroD1 target sites (both promoter and intergenic sites) as compared to the control (new figure: Fig 6C). Furthermore, this also accompanied a significant transcriptional induction of the associated genes (new figure: Fig 6D). These *in vivo* observations are fully supportive of our previous observations that NeuroD1 targeting to distinct regulatory elements results in a euchromatic state which causes transcriptional activation. We are highly thankful to the reviewer for suggesting these experiments that have clearly made highly relevant additions to the manuscript and increased the impact of our findings.

Minor points:
Comment 1) The authors indicate the number of genes overexpressed after NeuroD1 induction that are also direct targets (e.g. Fig 2D). Yet, they make no comment on the opposite relationship, i.e. the number of down-regulated genes that are also direct targets. If their model on NeuroD1-mediated chromatin opening should be correct, I would expect that virtually no direct target should ever be down regulated since the primary effect of NeuroD1 may be that of opening the chromatin to make it more accessible for upregulation. Is this the case or can NeuroD1 also act as a repressor?
Moreover, is there a correlation in the fold-change (FC) among groups of direct vs. secondary targets? If the chromatin-effect of NeuroD1 is overall comparable for all targets, this may lead to a similar FC among direct targets while secondary targets should be more diverse. Is this the case?

Authors' response: We thank the reviewer for advising us to look into this aspect. An overlap of down-regulated genes with genes bound by NeuroD1 at their regulatory elements showed that very few downregulated genes are direct targets of NeuroD1 (statically not significant) (new figure: Supplementary Fig 1J) as compared to the upregulated genes (statistically significant). These observations support a dominantly activating function of NeuroD1.

Prompted by the reviewers comment, we investigated whether there a correlation in the fold-change among groups of direct vs. secondary targets. Our analysis showed that while the range of induction within NeuroD1 direct target and non-target genes varies, the direct target genes are significantly higher upregulated than non-targets (secondary targets) following NeuroD1 expression (new figure: Supplementary Fig S3A). This has indeed provided additional validation that NeuroD1 function at its target sites is specific and is primarily activating.

Comment 2) I do not understand why binding domains are so different in Fig. 2N-O and what is the reason for comparing them. If the authors want to propose a binding domain for NeuroD1 they should validate it by luciferase assays. In the absence of this, that table seems rather useless.

Authors' response: We apologize for having confused the reviewer with this figure. In fact, the motifs shown in Fig. 2N-O were identified by analyzing known DNA binding motifs at NeuroD1 bound promoters and enhancers. This analysis led to at least two key conclusions. First, the quality of our ChIP-seq data is high as the known

NeuroD1 binding motif was among the top three most highly enriched motifs in the peaks from both promoters and enhancers. Second, NeuroD1 binding is highly sequence-specific, suggesting that in these cases genetic mechanisms might guide epigenetic state and transcriptional program via recruitment of critical sequence16 dependent transcription factors. We have now also performed *de novo* motif analysis at NeuroD1 peaks which shows that approximately 95% of high confidence peaks exhibit the classical E-box motif, which occurs at the peak center (new figure: Supplementary Fig S3G). The E-box motif is known to be among preferred target sequences of bHLH transcription factors (Jones, 2004), of which NeuroD1 is a member. Overall, these findings conclude that NeuroD1 is a highly sequence-specific transcription factor.

Comment 3) The authors often use 'qualitative' terms in the text (e.g. the majority of; a substantial fraction of; a large number of; etc.). While in most - but not all - cases quantifications are provided in the figures it would be more appropriate and simpler for the reader to use quantitative terms already in the text.

Authors' response: We thank the reviewer for the suggestion and we have now revised all figures and the text to reflect quantitative nature of our datasets by adding statistical tests and wordings wherever applicable.

Comment 4) The authors validate the expression of only 1 target: Lzts1. Many more should be provided in a supplemental figure using ISHs from the Allen Brain Atlas, Genepaint, Eurexpress or similar resources. Also, these targets should be absent prior to NeuroD1 expression, is this the case? Genome-wide automated ISH at different developmental stages is available with a click of a mouse and no pipetting involved.

Authors' response: We are glad that the reviewer made this suggestion and indicated us the appropriate publically available tools. Looking into these databases we found that ISH images were available for few of our NeuroD1 targets in the Allen Brain Atlas for E15.5 cortex and we have now included them in the current version of the manuscript (new figures: Fig 3E-J, upper panels). These images show that indeed the NeuroD1 targets are relatively less transcribed in the ventricular zone (VZ), but are very highly expressed in subventricular zone (SVZ) (where NeuroD1 levels are highest) and/or cortical plate (CP), suggesting that in cortical layers which do not highly express NeuroD1 these target genes are not strongly expressed. These observations were further validated by the analysis of the target gene expression using the RNA-seq data derived from the three layers of the embryonic cortex (VZ, SVZ and CP) and data are in full agreement with the ISH images (new figures: Fig 3E-J, lower panels). Moreover, we have also included ISH images for different development stages of a target gene, Lzts1, as well as NeuroD1 itself, which shows that Lzts1 expression pattern closely parallels NeuroD1 expression during cortical development. These data nicely complement our findings and provide additional *in vivo* validations of target genes. We are highly thankful to the reviewer for these suggestions.

Comment 5) In a few cases the authors mentioned data from in vivo studies where are these data coming from? A short description and citation should be provided. Also, if these come from different publications, have the authors normalized for total number of reads in each case?

Authors' response: We thank the reviewer for pointing this out. We have now cited references for these datasets wherever applicable and provided GEO IDs for all datasets used in this study as Supplementary Table S1 as well as referred to this in every figure that employed them. Following processing of these datasets, indeed we had normalized all data together that were being compared for both RNA-seq and ChIP-seq datasets. In detail, RNA-seq datasets were normalized by implementing library size normalization from DESeq package. *In vivo* RNA-seq datasets were normalized by implementing RPKM normalization. For ChIP-seq of transcription

factors and chromatin marks from ES cells, trimmed mean of M-values were used for normalization.

Comment 6) The two statements in the introduction: 'was successfully used to reprogram' and 'cell fate specification involves' need proper reference(s).

Authors' response: We are sorry for missing to cite important references and thank the reviewer for indicating this to us. The revised version of the manuscript contains proper citations.

Comment 7) Figures: i) It seems appropriate to add a model for the mechanism proposed. ii) Most labels are extremely small and impossible to read if not magnified many times. iii) No significance is given for the bars and box plots. iv) I believe that relative expression levels (e.g. 1D) should have all controls set to 1 and the scale values be also 1 in all cases because these are a.u. and there is no reason for such extreme numbers ranging from 4,000 (1A) to 0.005 (1D). v) Enhancers can be added to 2C.

Authors' response: We have addressed each of these comments as follows:
(i) We agree that a model would be a nice addition to the manuscript and we have now done so to reflect the most critical findings in a simplistic fashion (new figure: Fig 8)
(ii) We apologize for the small font size used previously for labeling in the figures. We have now increased the font size in all figures for a better visualization.
(iii) We have added significance wherever applicable throughout the manuscript.
(iv) We agree with the reviewer and we have now plotted fold changes with respect to the control as advised.
(v) We apologize for not making it clear that out of all NeuroD1 peaks falling in the intergenic regions, those ones showing H3K27ac at TN d1 (a timepoint immediate to the highest NeuroD1 expression during neuronal differentiation) are classified as enhancers and followed-up in this study. We have now added additional text in the manuscript to further clarify this point.
We are highly thankful to the reviewer for each of these comments that have improved the presentation of the manuscript and will have a huge impact on the readers' appreciation of our findings.

References:
Amamoto, R., and Arlotta, P. (2014). Development-inspired reprogramming of the mammalian central nervous system. Science *343*, 1239882.
Aprea, J., Prenninger, S., Dori, M., Ghosh, T., Monasor, L.S., Wessendorf, E., Zocher, S., Massalini, S., Alexopoulou, D., Lesche, M.*, et al.* (2013). Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. The EMBO journal *32*, 3145-3160.
Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nature genetics *44*, 148-156.
Cheng, Y., Ma, Z., Kim, B.H., Wu, W., Cayting, P., Boyle, A.P., Sundaram, V., Xing, X., Dogan, N., Li, J.*, et al.* (2014). Principles of regulatory information conservation between mouse and human. Nature *515*, 371-375.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.
Frank, E., Hall, M., Trigg, L., Holmes, G., and Witten, I.H. (2004). Data mining in bioinformatics using Weka. Bioinformatics *20*, 2479-2481.
Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. Nature *389*, 349-352.
Guo, Z., Zhang, L., Wu, Z., Chen, Y., Wang, F., and Chen, G. (2014). In vivo direct

reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell stem cell *14*, 188-202.
Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nature biotechnology *33*, 510-517.
Jones, S. (2004). An overview of the basic helix-loop-helix proteins. Genome biology *5*, 226.
Kraushaar, D.C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. Genome biology *14*, R121.
Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B.*, et al.* (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature *511*, 616-620.
Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS genetics *9*, e1003288.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.
Saito, T. (2006). In vivo electroporation in the embryonic mouse central nervous system. Nature protocols *1*, 1552-1558.
Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.
Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nature reviews Genetics *15*, 272-286.
Smialowski, P., Frishman, D., and Kramer, S. (2010). Pitfalls of supervised feature selection. Bioinformatics *26*, 440-443.
Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes & development *12*, 599-606.
Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nature reviews Molecular cell biology *15*, 703-708.
Thakurela, S., Garding, A., Jung, J., Schubeler, D., Burger, L., and Tiwari, V.K. (2013). Gene regulation and priming by topoisomerase IIalpha in embryonic stem cells. Nature communications *4*, 2478.
Tiwari, V.K., Stadler, M.B., Wirbelauer, C., Paro, R., Schubeler, D., and Beisel, C. (2012). A chromatin-modifying function of JNK during stem cell differentiation. Nature genetics *44*, 94-100.
Verdone, L., Caserta, M., and Di Mauro, E. (2005). Role of histone acetylation in the control of gene expression. Biochemistry and cell biology = Biochimie et biologie cellulaire *83*, 344-353.
Zhu, Y., Sun, L., Chen, Z., Whitaker, J.W., Wang, T., and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. Nucleic Acids Res *41*, 10032-10043.

2nd Editorial Decision 11 August 2015

Thank you for submitting a revised version of your manuscript and my apologies for the slight delay in the re-review process here. We have had a very high load of manuscript submissions in the last weeks and this has unfortunately led to delays in handling times.

Your manuscript has now been seen by one of the original referees (comments included below) and I am glad to let you know that this person finds that all major criticisms have been sufficiently addressed. However, before we can go on to officially accept the manuscript for publication there

are a few editorial issues concerning text and figures that I have to ask you to address. I would therefore encourage you to submit a final revision of your manuscript addressing the points outlined below:

-> You will see that our referee finds the box plots in fig 4B to describe your data more clearly than the Bayesian model in Fig 4A. However, I will leave it up to you to decide if fig 4A should stay in the main figure or rather be moved to the expanded view section.

-> Please fill out and include an author checklist as listed in our online guidelines (http://emboj.embopress.org/authorguide)

-> Please reduce the size of fig 3 if possible since the current 50MB size may cause difficulties in the transfer to our production team

-> As you will see in our online guide to authors we have changed our format to include either up to 5 Expanded view figures or a joined appendix pdf (as currently done for your supplemental figures). To avoid too much further restructuring at this stage, I would encourage you to maintain the current appendix format but I do need you to correct the labeling of these figures when referring to them in the main text (see http://emboj.embopress.org/authorguide for further details).

-> Regarding the separate data sets I would suggest that you relabel them as EV tables and also refer to them as such in the main text. If possible, we would also ask you to include the legends to the tables as a separate tab in the tables themselves rather than as part of the appendix file. Please feel free to contact me with any questions about this.

-> I have made a few changes in the abstract (below) that I would suggest you to consider for the final revision

'Cell fate specification relies on the action of critical transcription factors that become available at distinct stages of embryonic development. One such factor is NeuroD1, which is essential for eliciting the neuronal development program and possesses the ability to reprogram other cell types into neurons. Given this capacity, it is important to understand its targets and the mechanism underlying neuronal specification. Here, we show that NeuroD1 directly binds regulatory elements of neuronal genes that are developmentally silenced by epigenetic mechanisms. This targeting is sufficient to initiate events that confer transcriptional competence, including reprogramming of transcription factor landscape, conversion of heterochromatin to euchromatin and increased chromatin accessibility, indicating potential pioneer factor ability of NeuroD1. The transcriptional induction of neuronal fate genes is maintained via epigenetic memory despite a transient NeuroD1 induction during neurogenesis. Our study not only reveals the NeuroD1-dependent gene regulatory program driving neurogenesis but also increases our understanding of how cell-fate specification during development involves a concerted action of transcription factors and epigenetic mechanisms.'

-> Papers published in The EMBO Journal include a 'Synopsis' to further enhance discoverability. Synopses are displayed on the html version of the paper and are freely accessible to all readers. The synopsis includes a short standfirst - written by the handling editor - as well as 2-5 one sentence bullet points that summarise the paper and are provided by the authors. I would therefore ask you to include your suggestions for bullet points

-> In addition, I would encourage you to provide an image for the synopsis. This image should provide a rapid overview of the question addressed in the study but still needs to be kept fairly modest since the image size cannot exceed 550x400 pix

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to your revision.

REFEREE COMMENTS

Referee #2:

The authors have done a thorough job of addressing most of my concerns. They key issues were the lack of transparency and description of the data and bioinformatics analysis as well as the general language in the paper. The authors have obviously worked hard to address these. The new data is a great addition.

I still think that the Bayesian analysis does not add to the paper (Fig 4a) - the figure panel is very confusing (and non-standard) and the new box plots (Fig 4b) show exactly what the authors want to convey in a very simply intuitive manner - The data speaks for itself - there is higher quantitative levels of H3K27me3 on Targets and lower H3K27ac compared to control regions.

2nd Revision - authors' response 31 August 2015

Reviewer 2 (Remarks to the Author):

We thank the reviewer for many positive remarks and for the opinion that *"The authors have done a thorough job of addressing most of my concerns. They key issues were the lack of transparency and description of the data and bioinformatics analysis as well as the general language in the paper. The authors have obviously worked hard to address these. The new data is a great addition."*

The reviewer had one further comment that we have answered as follows:

Comment 1) I still think that the Bayesian analysis does not add to the paper (Fig 4a) - the figure panel is very confusing (and non-standard) and the new box plots (Fig 4b) show exactly what the authors want to convey in a very simply intuitive manner - The data speaks for itself - there is higher quantitative levels of H3K27me3 on Targets and lower H3K27ac compared to control regions.

Authors' response: We thank the reviewer for appreciating the introduction of new Fig 4B and commenting that this plot is able to convey the required information in a simpler and more intuitive manner. However, with due respect to the reviewer, we would like to keep Fig 4A as a main figure in addition to the newly provided Fig 4B. In our opinion it provides a new way of analyzing large scale epigenomics datasets to infer gene regulatory networks. This point has now been further elaborated in the main text.

