THE EMBO JOURNAL

Manuscript EMBO-2010-75716

Differential genomic targeting of the transcription factor TAL1 in alternate hematopoietic lineages

Carmen Palii, Carolina Perez-Iratxeta, Zizhen Yao, Yi Cao, Fengtao Dai, Harold Atkins, David Allan, F. Jeffrey Dilworth, Robert Gentleman, Stephen J. Tapscott and Majorie Brand

Corresponding author: Marjorie Brand, Sprott Center for Stem Cell Research-Ottawa Hospital Research Institute

Review timeline:

Submission date:	19 August 2010
Editorial Decision:	30 September 2010
Revision received:	12 November 2010
Editorial Decision:	30 November 2010
Accepted:	30 November 2010

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 30 September 2010

Thank you for submitting your manuscript for consideration at The EMBO Journal. I have now received the final report from the three referees that have evaluated your study, which I enclose below. As you will see all three referees find the study describing the differential binding of TAL1 in erythroid and leukemic T cells to be in general interesting and potentially important. Nevertheless they also raise a number of important issues regarding the current data that need to be addressed before the manuscript can be further considered for publication. Upon reflection these concerns are central to the main conclusions of the paper and include more indepth representation of the data, the bioinformatic analysis and overlap between TAL1 and ETS/RUNX datasets. Should be able to address these issues we would be happy to consider a revised version of the manuscript.

I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript. When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website: http://www.nature.com/emboj/about/process.html

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

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #1 (Remarks to the Author):

The manuscript by Palii et al. describes a comparative analysis of chromatin occupancy by the critical regulator of hematopoiesis TAL1 in erythroid versus malignant T-cells. As prior studies have described TAL1 chromatin occupancy genome-wide (and relevant non-genome-wide studies) in erythroid and T-cells, and also mechanistic studies on how TAL1 occupies chromatin, this reduces the potential novelty of the current body of work. Importantly, the authors have not appropriately presented their work in the context of what was known previously from studies of TAL1 function in erythroid cells, including how TAL1 interfaces with GATA-1 on chromatin. Nevertheless, no prior studies compared TAL1 occupancy and targets in erythroid versus malignant T-cell, and assessing how trans-acting factors select chromatin targets and function through cis-elements at the targets in distinct cellular contexts represents an important area of investigation in which many unknowns exist. The authors noted that numerous TAL1-occupied sites differed between the two systems and used bioinformatics and molecular approaches to evaluate the potential contribution of other factors to the distinct chromatin binding patterns. This analysis implicated ETS and RUNX factors as being determinants for TAL1 function in malignant T-cells, which represents a significant advance. Overall, the work appears to have been carefully conducted, and as noted above, the comparison of factor occupancy in distinct systems represents an instructive approach, which yielded unexpected mechanistic insights. Upon rewriting components of the manuscript to more accurately describe what is known about TAL1 chromatin occupancy (e.g. TAL1 - GATA chromatin chromatin site selection, cis-elements mediating TAL1-GATA chromatin binding, functional consequences of TAL1-GATA co-occupancy etc.), refining (truncating) the exceptionally long Discussion, and implementing the following recommendations, the manuscript shall be further strengthened.

Additional Comments:

1) Page 3, middle - The authors refer to TAL1 as a "double-edge" transcription factor, since it functions differently in distinct contexts. However, this seems to be the rule for many transcription factors, and I do not see how TAL1 may be somehow unique in this regard.

2) page 10, middle - The authors describe the major difference between TAL1 occupancy sites in primary erythroid versus the malignant T-cell line. This gross difference raises the question as to whether it reflects the comparison of a transformed cell with a normal cell or erythroid with T-cell. It would be important to further address this important issue by discussion and/or associated experimentation. Would one expect overt differences between TAL1 occupancy in normal versus transformed erythroid cells?

3) While the authors focused on ETS1, it would be instructive to assess whether other ETS factors functionally interact with TAL1 in the context of malignant T-cells. For example, incorporating data on which ETS factors are expressed in this system, which are the most abundant, and comparing ETS1 to another well expressed ETS factor would be particularly informative.

4) The mechanistic analysis involving co-IP in Fig. 6 has potential to be important. I do not see details on the antibodies used in the experimental methods section, nor are molecular weight markers shown to allow other investigators to evaluate/reproduce this data.

5) Large tracts of text are presented without paragraphs, which requires revision.

Referee #2 (Remarks to the Author):

This manuscript identifies the complement of DNA sequences bound by the transcription factor and oncoprotein TAL1 in normal primary erythroid cells and leukemic T cells (T-ALL) which were identified by ChIP-sequencing analyses. This data was combined with gene expression profiling to focus on gene sets that could account for the distinct functions of TAL1 in red cells and T-ALL. Motif search together with additional functional and biochemical experiments provide convincing support for the proposal that TAL1 is associated with distinct transcriptional networks in the two lineages.
The work is conducted carefully and presented in a clear and concise manner. The focus on composite motifs is of particular interest. This is the first report of a genome-wide analysis of TAL1's targets in leukemic cells.

Specific points

1. Description of TAL1 KD erythroid cells (Figure 1) must include cell morphology (MGG/benzidine staining), immunophenotype (CD36, CD71, GPA) and colony-forming capacity (there is a suggestion that colony assays were done in the Supplementary Material - but the data is not mentioned). This would document the differentiation block in TAL1 KD primary human erythroid cells, which has never been fully reported.
Separately, for global gene expression profiling, it is critical that WT and KD populations are closely matched to minimise secondary changes in gene expression. The authors need to clarify exactly how well matched the KD and WT cell populations were. The additional data required on the KD cells will help.

2. The authors must mention whether erythroid cell populations used in ChIP-seq studies are the same as those analysed for expression profiling. The stage and days of differentiation must be indicated. The two studies should use the same populations (day 12).

3. For Jurkat KD cells, microarray analysis must be done before KD cells all become apoptotic. Was this the case?

4. To support the hypothesis that TAL1 disrupts E-protein homodimers at cis-elements of genes repressed in T-ALL, the authors need to show that TAL1 targets in Jurkat cells are bound by E-protein homodimers in primary T cells or KD Jurkat cells.

5. A number of recent papers present ChIP-sequencing analyses on TAL1 and GATA1 in erythroid cells. The authors should refer to the data from TAL1 ChIP-seq analyses from primary mouse erythroid cells (Kassouf et al, Gen Research, 2010; the motif analyses in the two studies show comparable data) and mention that a strong correlation between SCL and GATA motifs in red cells was reported in recent GATA1 ChIP-seq analyses (Cheng et al, Gen Research 2009; Fujiwara et al, Mol Cell, 2009).

6. The microarray data have been deposited into GEO, but not the ChIP-seq data. These data have to be submitted before publication of the paper. They should have been made available to the reviewers.

Minor points

7. The time point at which Jurkat cells are analysed after addition of doxycyclin should be indicated.
8. The reference Fang et al, 2009 has been misplaced.

Referee #3 (Remarks to the Author):

Comments to authors: The goal of this study was to shed light on how the transcription factor TAL1 contributes to acute T cell leukemia. The authors used siRNA technology to knockdown (KD) TAL1 expression in erythroid (normal) and T cell (Jurkat leukemic cell line) contexts to discover basis of leukogenic role in T cells. The KD cells were analyzed both by phenotypic studies and gene

expression profiling. Findings suggest that TAL1 functions in erythroid differentiation, but retains T cells in an early developmental state, characteristic of leukemia. Genome-wide occupancy data, obtained by ChIP-SEQ, suggests that different genes are bound by TAL1 in the two cell types. An interesting set of TAL1-regulated and TAL1 bound genes (with broad 50kb cutoff for gene proximity to bound region) were derived from the combination of approaches to implicate five possible pathways of TAL1's role in leukemogenesis. All represent known general mechanisms of leukemogenesis-repression of pro-apoptotic, tumor suppressor and T-cell differentiation genes as well as activation of anti-apoptotic and anti-T cell differentiation genes. Direct ChIP confirmed ChIP-SEQ data at interesting genes in Jurkat, other leukemic cell lines, and patient's cells. In addition, bioinformatics approaches found E-boxes consistent with TAL1 binding. Additional overrepresented motifs (GATA, ETS1, RUNX) were also discovered. A genomic picture of a previous in vitro-characterized E-box/GATA composite elements was discovered. Also, a new E-box/ETS element was found. These factors are known regulators of T cell genes, but the overlap with TAL1 is of interest as this potentially implicates a new set of transcription factors in TAL1's mechanism of oncogenesis. Some occupancy dependency was implicated by KD experiments followed by ChIP. The new findings demonstrate the value of genome-wide approaches that combine gene profiling, genomic occupancy and bioinformatics. No major mechanistic insight was obtained but the combined approaches shed valuable insight into the possible pleiotrophic effect of overexpression of an oncogenic transcription factor.

Specific comments.

1-There is some concern that up and down-regulated genes are not clearly defining the phenomenon of TAL1 arresting differentiation of T cells. Effects on genes related to cell growth, cell cycle and differentiation differ in two cell types. This may reflect more the state of the two biological systems rather than erythroid vs. T cell function. Also, it appears that there are up and down-regulated genes in both cell models-so not a real difference-and likely not all are direct targets, as shown later. Also, it is poorly justified why two different significant cutoffs were used for sets of genes studied. The ontology analyses found cell type genes at low significance and perhaps are only discovered due to large numbers of genes put into the GO search rather than valid functional categories. Authors should make it more that the partial list represented in FigureS2 are cherry-picked set-even to the degree that in Supplemental table II the T cell differentiation genes are not even listed due to low significance. The weakness of these data is somewhat countered by the lists of genes already implicated in T cell function that were analyzed further and discussed.

2-KD of TAL1-cells die by apoptosis, thus compromising the rigor of the analysis of changes in gene expression in this arm of the experimental design. Selected transcription factors increase as a sign of differentiation, yet Runx and Ets1 which regulate T cell specific genes found in differentiatied cells are down-regulated. Again, the gene expression changes have limitations.

3-Authors propose that TAL1 genomic sites are highly restrictive compared to MyoD as judged by number of total bound regions. It is difficult, and thus inappropriate, to compare overall number of peaks from one protein to other protein in ChIP-SEQ experiments, especially if the data analysis and read depth are not the same. Also, antibody variability is likely to contribute to differences without biological significance. 1000's of occupied "targets" is in the range of published data for various transcription factors, so this point should be downplayed or even deleted.

4-Figure 4A GO terms in this figure need p values.

5-p.10-The stated bias towards TSS in Figure 4B represents 50% of the peaks with >50% at greater distance. This should be stated more carefully to more accurately reflect the data.

6-Figure 6 A Score for motifs is defined as z-value of motif, yet there is no definition of z-value in methods. See further comments below. What criteria are used on Figure 6 A to name Runx, GATA, etc.? I strongly recommend that MEME analysis and a conventional PWM pix be included for the E-Box preference, the E-GATA composite and the E-ETS composite. Also, the actual numbers of findings of these composite elements with constrained spacing should be mentioned explicitly in the text. The Hollenhorst et al. cited paper reported an ETS-RUNX composite element. It would be of interest whether this element is overlapping with the TAL1 bound regions.

7-Figure 6 E-Co-IP experiments need to be performed with an ethidium bromide control to eliminate the possible linkage through DNA-a common artifact of co-IP for two DNA binding proteins. There is no methods section for this experiment.

8-page 17-The authors use published Hollenhorst et al genome-wide occupancy data for ETS1 and RUNX in Jurkat cells to report overlap with TAL1 bound regions. It is not clear why a range of percentages are given. It is not clear whether the proximity of TAL1 and ETS1, RUNX sites was analyzed by a program (such as InterceptRegions in the USEQ suite) that set an overlap interval or whether just linkage to same nearest gene was used. This would be very valuable in considering contention of direct functional consequence of the co-occupancy of factors. Also, a test for statistical significance, such as a Fisher's exact test, should be used on the frequency of dual occurrence TAL1 and RUNX1/3 or TAL1 and ETS1. It should be noted that the ETS1 bound regions represent over 10,000 genes and many T cell genes, thus, the overlap may not be functional relevant in many cases.

9-ChIP SEQ data analysis has not been performed with current methodologies. Software and versions used have been upgraded in the field, such as the MACS program or the USEQ suite. These methods make simple local comparisons of reads from experimental vs. input that is superior to the global expectation methods cited. The gold standard for reads is 10 million whereas the reported data are based on 5-6 million reads. Although this has perhaps not seriously hampered the study, the investigators are encouraged to consider this upgrading.

10-Bioinformatics to find motifs under occupancy peaks was performed by older-style methodology that is somewhat word processing based. The Score outcome is not well described with the Z-value not being defined. In particular is there a contribution from length, complexity or strength of dinucleotide conservation? Also, the use of non-conventional motif presentation method is difficult to read compared to well-established methods and does not add significant information to the report. The new method is described awkwardly. The use of "depth" of edge colors is not clear-perhaps "darkness" in scale of black to grey would make sense. But where is the quantification of this variable? Newer more robust methods such as MEME with standard motif PWM format output should be used, at least for some critical observations (see above).

Minor:

Jurkat cell line should be capitalized throughout.

1st Revision - authors' response 12 November 2010

Referee #1 (Remarks to the Author):

The manuscript by Palii et al. describes a comparative analysis of chromatin occupancy by the critical regulator of hematopoiesis TAL1 in erythroid versus malignant T-cells. As prior studies have described TAL1 chromatin occupancy genome-wide (and relevant non-genome-wide studies) in erythroid and T-cells, and also mechanistic studies on how TAL1 occupies chromatin, this reduces the potential novelty of the current body of work. Importantly, the authors have not appropriately presented their work in the context of what was known previously from studies of TAL1 function in erythroid cells, including how TAL1 interfaces with GATA-1 on chromatin. Nevertheless, no prior studies compared TAL1 occupancy and targets in erythroid versus malignant T-cell, and assessing how trans-acting factors select chromatin targets and function through cis-elements at the targets in distinct cellular contexts represents an important area of investigation in which many unknowns exist. The authors noted that numerous TAL1-occupied sites differed between the two systems and used bioinformatics and molecular approaches to evaluate the potential contribution of other factors to the distinct chromatin binding patterns. This analysis implicated ETS and RUNX factors as being determinants for TAL1 function in malignant T-cells, which represents a significant advance. Overall, the work appears to have been carefully conducted, and as noted above, the comparison of factor occupancy in distinct systems represents an instructive approach, which yielded unexpected mechanistic insights. Upon rewriting components of the manuscript to more accurately describe what is known about TAL1 chromatin occupancy (e.g. TAL1 - GATA chromatin chromatin site selection, cis-elements mediating TAL1-

GATA chromatin binding, functional consequences of TAL1-GATA co-occupancy etc.), refining (truncating) the exceptionally long Discussion, and implementing the following recommendations, the manuscript shall be further strengthened.

We thank the reviewer for the comments and suggestions.

We agree that in the previous version of the manuscript we have devised our introduction with a particular focus on the role of TAL1 in T-ALL to the detriment of its function in erythroid cells. We apologize for this. We have now revised the text to address this point, including a description of what was known previously about TAL1-GATA1 occupancy in erythroid cells.

Regarding the novelty of our work, we are not aware of any previous study describing TAL1 genome-wide binding in the T-cell lineage (i.e. T-cell progenitors or T-ALL cell lines or primary blasts). We also note that our manuscript represents the first study to assess the genome-wide binding of human TAL1 in primary erythroid cells as opposed to previous studies using transformed human erythroid cell lines (i.e. K562). Using a transformed cell line presents some caveats. For example even though K562 cells were established from a 53-year-old woman with CML, this cell line expresses embryonic and fetal, but not adult, ß-globin genes. Therefore the stage of development/erythroid differentiation represented by the K562 cell line is not entirely clear. In contrast, in our study, we have used non-transformed adult primary erythroid cells differentiated ex vivo from adult CD34+ hematopoietic progenitors. This is a significant advantage in that: 1) this is a physiologically relevant cell type at a defined stage of differentiation (in our case adult pro-erythroblasts); and 2) it eliminates problems associated with cell transformation affecting specificity of transcription factor binding. We note that the effects of cell transformation are less relevant to our study in Jurkat cells (originally derived from a T-ALL patient) because in T-ALL, cell transformation is a consequence of TAL1 expression. Furthermore we have validated our findings from the Jurkat cells in 4 additional T-ALL cell lines and in primary blasts from two T-ALL patients.

Additional Comments:

1) Page 3, middle - The authors refer to TAL1 as a "double-edge" transcription factor, since it functions differently in distinct contexts. However, this seems to be the rule for many transcription factors, and I do not see how TAL1 may be somehow unique in this regard.

We agree with the reviewer that most transcription factors function differently in distinct contexts. What we were trying to imply with the wording “double-edge” is that TAL1 has opposite functions in 2 cell lineages: it promotes cell differentiation in one lineage (i.e. erythroid), while it blocks differentiation in another lineage (i.e. T-cell). Whereas many transcription factors regulate different genes in different cell types, TAL1 is quite unique in its capacity to promote opposite cell fates depending on the lineage. Nevertheless, we agree that our wording could be confusing and we have removed the word “double-edge” while simply describing TAL1’s differential functions.

2) page 10, middle - The authors describe the major difference between TAL1 occupancy sites in primary erythroid versus the malignant T-cell line. This gross difference raises the question as to whether it reflects the comparison of a transformed cell with a normal cell or erythroid with T-cell. It would be important to further address this important issue by discussion and/or associated experimentation. Would one expect overt differences between TAL1 occupancy in normal versus transformed erythroid cells?

An important point to consider regarding the respective contributions of cell-transformation *vs* cell-lineage in the differences we observe for TAL1 in erythroid *vs* T-ALL is that TAL1 is a causing factor for malignant transformation in the T-cell lineage (Condorelli et al. Cancer Res (56), 1996, Kelliher et al. EMBO J. (15), 1996). Therefore, our study should be perceived as a comparison between 2 existing cell types that express TAL1: erythroid cells (that happen to differentiate) and T-cell (that happen not to differentiate but to proliferate) (this is explained on p.6 of the manuscript). Clearly, the differences we observed for TAL1 between the 2 cell contexts reflect both the cell-type (erythroid *vs* T-cell) and the transformation status (normal *vs* malignant). We have now clarified this point at the beginning of the Discussion (p.21). But importantly, both

contributing factors are physiologically relevant, especially since we have confirmed our results from Jurkat cells in 4 additional T-ALL cell lines and in primary blasts from 2 T-ALL patients.

While distinguishing the respective contributions of cell transformation status *vs* cell-type for TAL1 function would be very valuable, it remains a technical challenge since expression of TAL1 in T-cell would cause malignant transformation. Distinguishing TAL1 function between normal and transformed erythroid cells would require us to compare TAL1 binding in our primary human erythroid cells *vs* primary human erythroid cells which have been transformed at the same stage of differentiation. These experiments are beyond the scope of this manuscript and will be the focus of future studies. As highlighted above (in the response to the main Point), K562 cells are not suitable for such a comparison.

3) While the authors focused on ETS1, it would be instructive to assess whether other ETS factors functionally interact with TAL1 in the context of malignant T-cells. For example, incorporating data on which ETS factors are expressed in this system, which are the most abundant, and comparing ETS1 to another well expressed ETS factor would be particularly informative.

ETS family members demonstrate extensive co-expression, with at least 17 members expressed in Jurkat cells. Based on gene expression studies (Hollenhorst et al. NAR (32), 2004), ETS1 is the most abundant ETS family member expressed in Jurkat cells (213 mRNA copies per cell), followed by GABPA (146 mRNA copies per cell). As requested we have tested whether GABPA interacts with TAL1 by reciprocal co-immunoprecipitations of endogenous proteins in Jurkat nuclear extracts. We found that TAL1 does not interact with GABPA confirming the specificity of its interaction with ETS1 (amended Figure 6E). This result, together with the fact that the ETS motif we have identified by de novo search in Jurkat cells (Figure 6A) resembles a recently identified ETS1-specific motif (i.e. bound by ETS1 but not GABPA) with a T-cell specific function (Hollenhorst et al. PLoS Genetics 2009 (5)) strongly support our hypothesis that TAL1 acts on T-cell specific genes in T-ALL. Furthermore this specificity might at least partially be mediated through TAL1 interaction with ETS1 that displays a T-cell specific function. This new data and conclusion have been added to the manuscript (Figure 6E and p.18, 23)

4) The mechanistic analysis involving co-IP in Fig. 6 has potential to be important. I do not see details on the antibodies used in the experimental methods section, nor are molecular weight markers shown to allow other investigators to evaluate/reproduce this data.

Details on the antibodies (including their catalog numbers) used in the co-IPs and western blots are provided in the Supplemental Methods (see Antibodies section). As requested, we have now added molecular weight markers on Fig. 6 as well as on all the other figures displaying western blots. Please note that Figure 6E presented in this version of the manuscript displays the results of new co-IP experiments that have been repeated in the presence of ethidium bromide (as requested by Reviewer 3).

5) Large tracts of text are presented without paragraphs, which requires revision.

We have revised the text to address this point.

Referee #2 (Remarks to the Author):

This manuscript identifies the complement of DNA sequences bound by the transcription factor and oncoprotein TAL1 in normal primary erythroid cells and leukemic T cells (T-ALL) which were identified by ChIP-sequencing analyses. This data was combined with gene expression profiling to focus on gene sets that could account for the distinct functions of TAL1 in red cells and T-ALL. Motif search together with additional functional and biochemical experiments provide convincing support for the proposal that TAL1 is associated with distinct transcriptional networks in the two lineages.
The work is conducted carefully and presented in a clear and concise manner. The focus on composite motifs is of particular interest. This is the first report of a genome-wide analysis of TAL1's targets in leukemic cells.

Specific points

1. Description of TAL1 KD erythroid cells (Figure 1) must include cell morphology (MGG/benzidine staining), immunophenotype (CD36, CD71, GPA) and colony-forming capacity (there is a suggestion that colony assays were done in the Supplementary Material - but the data is not mentioned). This would document the differentiation block in TAL1 KD primary human erythroid cells, which has never been fully reported.

As requested we have performed additional assays to better document the defect in differentiation upon TAL1 KD in primary human erythroid cells (see new Figure S2 and Figure 1H). While MGG staining showed minimal differences in morphology upon TAL1 KD (Figure S2A), benzidine staining clearly shows that there is a strong decrease in hemoglobin synthesis (Figure 1H and Figure S2B) resulting at least in part from decreased ß-globin transcripts upon TAL1 KD (Figure 4C). Furthermore, TAL1 KD leads to a decrease in CD36, CD71 and GPA cell surface markers (Figure S2C). Please, note that the decrease is more pronounced for the late erythroid marker GPA, which is consistent with the proerythroblast stage of our cells. Finally we did not see colony formation in the scrambled control or TAL1 KD cells (data not shown), which again is consistent with the proerythroblast stage where cells have lost their colony forming capacity

Separately, for global gene expression profiling, it is critical that WT and KD populations are closely matched to minimise secondary changes in gene expression. The authors need to clarify exactly how well matched the KD and WT cell populations were. The additional data required on the KD cells will help.

WT and TAL1 KD cell populations are globally very similar according to their gene expression profiles. Indeed gene expression profiling indicate that from the 28,869 annotated genes interrogated on our microarray platform, only 2% changed upon KD of TAL1 in erythroid cells and 2.22% changed upon TAL1 KD in Jurkat cells. As an example we show a plot of the normalized expression values for one replicate of each cell type (erythroid and Jurkat). Please see Figure 1 for Reviewer only *(at the end of this file)*.

2. The authors must mention whether erythroid cell populations used in ChIP-seq studies are the same as those analysed for expression profiling. The stage and days of differentiation must be indicated. The two studies should use the same populations (day 12).

We confirm that ChIP-seq studies in erythroid cells were done using cells at Day 12 of differentiation (i.e. the same stage used for gene expression profiling). We have clarified this point in the manuscript (p.7-8 and p.9-10).

3. For Jurkat KD cells, microarray analysis must be done before KD cells all become apoptotic. Was this the case?

Microarray analysis of Jurkat KD cells was done 72h after Dox treatment. All phenotypic analyses shown on Figure 2 have also been performed at this time point. As shown on Figure 2E, approximately 70% of the cells are positive for Annexin V at this time-point. However it should be noted that AnnexinV is an early marker of apoptosis (J Exp Med. 1995, 182(5):1545-56) and the level of necrotic cells (i.e. 7AAD positive) is minimal at this time point (see Figure 2 for reviewer only) showing that the cells are still intact. Importantly, evidence presented below indicates that working on a background of early stages of apoptosis has not hampered our gene expression analysis. Indeed:

1) Only 2.22% of the 28,869 annotated genes interrogated on our microarray platform change upon TAL1 KD (see Figure 1 for reviewer only)

2) Among changing genes, both upregulated (370) and downregulated (249) genes were identified upon TAL1 KD (Figures S3B,C). Furthermore, these results have been confirmed by RT-qPCR for 44 tested genes (both up- or down-regulated) (Figure S4B and data not shown).

3) Changes in gene expression (both up and downregulated) upon TAL1 KD have now also been validated at the protein level for important T-cell specific proteins (Figure 3D).

In conclusion, our Jurkat knocked down cells were analyzed at a time-point that is appropriate to identify TAL1-dependent genes important for T-cell differentiation against a background of apoptosis.

4. To support the hypothesis that TAL1 disrupts E-protein homodimers at cis-elements of genes repressed in T-ALL, the authors need to show that TAL1 targets in Jurkat cells are bound by E-protein homodimers in primary T cells or KD Jurkat cells.

The model that TAL1 disrupts E-protein homodimers is not stemming from our findings, but has been proposed previously by others (reviewed in Exp Hematol. 2004 32(1):11-24). The findings of our paper do not imply that this hypothesis is correct. Instead we only make the remark that the E-box enrichment in Jurkat cells is consistent with this proposed model. Testing of such a model remains difficult to do as the ChIP method does not allow one to assess protein homo- or hetero- dimerization (i.e. even a positive signal in a ChIP-reChIP experiment cannot distinguish between binding proximity and dimerization). Furthermore testing homodimerization of endogenous proteins in vivo would require the development of an Ab that recognizes the homodimer but not its monomeric counterpart.

To ensure that the reader does not believe we have confirmed the E-protein homodimer disruption model we have modified the wording of the text, which now reads: "The finding that TAL1 binds to the E-protein homodimer's preferred E-box more frequently in Jurkat than erythroid cells is consistent with the long-standing model in which TAL1 deregulates E-protein homodimers' function in T-ALL (Begley & Green, 1999; O'Neil & Look, 2007)." In addition, we have modified our model on Figure 8A to remove any reference to the E-protein homodimer disruption model.

5. A number of recent papers present ChIP-sequencing analyses on TAL1 and GATA1 in erythroid cells. The authors should refer to the data from TAL1 ChIP-seq analyses from primary mouse erythroid cells (Kassouf et al, Gen Research, 2010; the motif analyses in the two studies show comparable data) and mention that a strong correlation between SCL and GATA motifs in red cells was reported in recent GATA1 ChIP-seq analyses (Cheng et al, Gen Research 2009; Fujiwara et al, Mol Cell, 2009).

We have now better incorporated the results from these studies throughout our manuscript (Introduction, Results and Discussion).

6. The microarray data have been deposited into GEO, but not the ChIP-seq data. These data have to be submitted before publication of the paper. They should have been made available to the reviewers.

The ChIP-seq data is now deposited into GEO, under the accession number GSE25000.

Minor points

7. The time point at which Jurkat cells are analysed after addition of doxycyclin should be indicated.

We have added this information in the legend of Figure 2 and in the manuscript (p.7).

8. The reference Fang et al, 2009 has been misplaced.

This mistake has been corrected.

Referee #3 (Remarks to the Author):

Comments to authors: The goal of this study was to shed light on how the transcription factor TAL1 contributes to acute T cell leukemia. The authors used siRNA technology to knockdown (KD) TAL1 expression in erythroid (normal) and T cell (Jurkat leukemic cell line) contexts to discover basis of leukogenic role in T cells. The KD cells were analyzed both by phenotypic studies and gene expression profiling. Findings suggest that TAL1 functions in erythroid differentiation, but retains T cells in an early developmental state, characteristic of leukemia. Genome-wide occupancy data,

obtained by ChIP-SEQ, suggests that different genes are bound by TAL1 in the two cell types. An interesting set of TAL1-regulated and TAL1 bound genes (with broad 50kb cutoff for gene proximity to bound region) were derived from the combination of approaches to implicate five possible pathways of TAL1's role in leukemogenesis. All represent known general mechanisms of leukemogenesis-repression of pro-apoptotic, tumor suppressor and T-cell differentiation genes as well as activation of anti-apoptotic and anti-T cell differentiation genes. Direct ChIP confirmed ChIP-SEQ data at interesting genes in Jurkat, other leukemic cell lines, and patient's cells. In addition, bioinformatics approaches found E-boxes consistent with TAL1 binding. Additional overrepresented motifs (GATA, ETS1, RUNX) were also discovered. A genomic picture of a previous in vitro-characterized E-box/GATA composite elements was discovered. Also, a new E-box/ETS element was found. These factors are known regulators of T cell genes, but the overlap with TAL1 is of interest as this potentially implicates a new set of transcription factors in TAL1's mechanism of oncogenesis. Some occupancy dependency was implicated by KD experiments followed by ChIP. The new findings demonstrate the value of genome-wide approaches that combine gene profiling, genomic occupancy and bioinformatics. No major mechanistic insight was obtained but the combined approaches shed valuable insight into the possible pleiotrophic effect of overexpression of an oncogenic transcription factor.

We thank the reviewer for these comments and agree that our experiments provide valuable insight into how aberrant expression of TAL1 could lead to oncogenesis in T-cells. In addition, we feel that the finding that TAL1 binding to specific genes is regulated through RUNX and ETS1 factors provide mechanistic insight into how transcription factors are targeted to specific sites within the genome.

Specific comments.

1-There is some concern that up and down-regulated genes are not clearly defining the phenomenon of TAL1 arresting differentiation of T cells. Effects on genes related to cell growth, cell cycle and differentiation differ in two cell types. This may reflect more the state of the two biological systems rather than erythroid vs. T cell function.

Our unbiased GSEA analysis (Figure 3A-C) clearly shows that there is an enrichment of early stage T-cell genes in normal Jurkat cells. Furthermore, the GSEA analysis complemented with RTqPCR and Western blot shows that upon TAL1 knockdown, there is re-activation of several T-cell markers including the master regulators of T-cell differentiation GATA3 and Aiolos at the transcript (Figure S4B) and protein (Figure 3D) levels, as well as SOX4, TOX, CD69, CCR9, SH2D3C, IL27RA , DUSP10, DUSP1, BCL6 and SLAMF1 (Figure S4B and data not shown). Among these 12 genes, 9 have also been identified as direct targets of TAL1 (i.e. Aiolos, TOX, CD69, DUSP10, DUSP1, CCR9, SH2D3C, IL27RA and BCL6) by ChIP-seq analysis (See Supplementary Table VI and Figure 8B), a result confirmed by ChIP-qPCR (Figures 4, S7 and data not shown). Taken together, these results strongly support our conclusion that the transcriptional program controlled by TAL1 in jurkat cells involves an arrest of T-cell differentiation. Also, we do not mean to imply that knocking down of TAL1 transform Jurkat cells back into fully functional T-cells, a result that cannot be expected by knocking down a single transcription factor in a transformed cell line.

RE the comment that “Effects [of TAL1 KD] on genes related to cell growth, cell cycle and differentiation differ in two cell types”, we note that our results and that of others (reviewed in Lecuyer et al. 2004 (32) Exp Hematol) have shown that TAL1 binding specificity can be modulated by its interacting partners (i.e. GATA, ETS, RUNX). In addition, different chromatin states between cell types are expected to strongly affect the accessibility of genes that are targeted to and regulated by TAL1. Therefore, our results do indeed reflect the role of TAL1 in 2 biological systems: one where it causes differentiation (erythroid), and one where it leads to oncogenesis (T-cell).

Also, it appears that there are up and down-regulated genes in both cell models-so not a real difference-and likely not all are direct targets, as shown later.

Indeed, it has been shown previously that TAL1 can be both a repressor and an activator in erythroid cells (Mol Cell Biol. 2000 Mar;20(6):2248-59; EMBO J 2000 19:6792–6803; EMBO J

2006 25*:*357–366). We (Figure S2) and others (Blood 2006 108*:*986–992) have also observed this dual function in Jurkat cells. The interesting difference between these two cell types is that TAL1 acts as an activator more frequently in erythroid cells and a repressor more frequently in Jurkat cells. The mechanism which gives rise to this difference between cell types will be the subject of future studies.

Also, it is poorly justified why two different significant cutoffs were used for sets of genes studied.

For knocking down TAL1 we have used 2 different delivery systems for the same shRNA (i.e. Dox-inducible for Jurkat cell line, and lentivirus-mediated for primary erythroid cells). For our microarray studies, the choice of alternate FDR is based on the fact that in the Jurkat cell line, we were able to obtain a better knockdown and generated enough material to perform 6 independent replicates. In the case of the primary erythroid cells, the knockdown was less efficient (but still highly significant) and we only performed 4 replicates due to the cost associated in obtaining primary human cells. Thus the FDR cutoff chosen for erythroid cells was at a lower threshold to obtain a similar number of changing genes as indicated in the Methods section. While the FDRs are not the same, both are significant. In addition, all genes discussed in the paper have been confirmed by RT-qPCR.

The ontology analyses found cell type genes at low significance and perhaps are only discovered due to large numbers of genes put into the GO search rather than valid functional categories.

The term "significance" of Gene Ontology analyses has to be interpreted carefully. The p-value computation in the context of GO analysis cannot be taken at face value because the dependency structure of the multiple hypotheses being tested is too complex to be modeled. In fact, the rank order of the GO term is more relevant than the pvalue. Because of these limitations GO analyses are usually used as a guide for interpreting high-throughput data, and they must be corroborated by experimental evidence.

With this in mind, the role of these analyses in our paper has been to assist us in interpreting and presenting our gene lists through summarization and to support other pieces of evidence through the overrepresented themes. We would like to highlight that in no part of our manuscript, have we used the GO analysis results as the unique or hard evidence of a particular TAL1 function. Instead, we used it as a basis for exploring potential roles of TAL1 using additional experimentation.

We do provide p-values for the GO terms as we obtained them but again we don't make hard claims based on them. Aside from this issue, we can confirm that the top enriched terms are supportive of the observed phenotypes and our experimental findings.

Authors should make it more that the partial list represented in FigureS2 are cherry-picked set-even to the degree that in Supplemental table II the T cell differentiation genes are not even listed due to low significance.

The legend of this Figure (i.e. Figure S3 in the new version of our paper) states: "The most relevant GO categories […] are presented." This sentence is immediately followed by a reference to Supplementary Table II so that the reader can access all the categories if he wishes (i.e."for a full list of GO terms, see Supplementary Tables I and II"). We have modified the text that now reads: " The most relevant GO categories were selected and are presented".

In the previous version of the manuscript, some categories did not appear in the Supplemental Table II because we had used a p value cutoff of 1E-02 for this Table (and the other GO-containing tables). We have now used a pvalue cutoff of 5E-02 for Supplementary Table II (as well as the other Tables containing GO categories).

The weakness of these data is somewhat countered by the lists of genes already implicated in T cell function that were analyzed further and discussed.

2-KD of TAL1-cells die by apoptosis, thus compromising the rigor of the analysis of changes in gene expression in this arm of the experimental design. Selected transcription factors increase as a sign of differentiation, yet Runx and Ets1 which regulate T cell specific genes found in differentiatied cells are down-regulated. Again, the gene expression changes have limitations.

Since several processes (i.e. apoptosis and reactivation of the T-cell differentiation transcriptional program) are taking place simultaneously upon TAL1 knockdown in Jurkat cells, the analysis of changes in gene expression has to be performed carefully. Accordingly our interpretation of these data has been performed in the most rigorous manner and this has led to meaningful conclusions, many of which we have independently validated. Therefore, it is clear from our data that the apoptotic phenotype has not compromised the rigor of the analysis of changes in gene expression. Points to consider are the following:

1) Our conclusion regarding the reactivation T-cell differentiation is not based on a few selected transcription factors but on an unbiased comparative analysis of gene expression changes that occur upon TAL1 knockdown in Jurkat cells *vs* differentiating primary human T-cells (described in Figure 3A-C). This GSEA revealed a significant re-activation (both up and downregulated genes) of the T-cell transcriptional program upon TAL1 knockdown in Jurkat cells.

2) AnnexinV is an early marker of apoptosis (J Exp Med. 1995, 182(5):1545-56), and FACS analysis indicates that the level of necrotic cells (7AAD positive) is very low at the time-point used for microarray after TAL1 knockdown (see Figure 2 for reviewer only, at the end of this file).

3) Most (98%) genes do not change upon TAL1 KD in Jurkat cells at the time-point used for microarray after TAL1 knockdown (see Figure 1 for Reviewer only, at the end of this file).

4) We have validated by RT-qPCR the re-activation of 12 genes selected for their previously documented involvement on T-cell differentiation (i.e. GATA3, Aiolos/IKZF3, SOX4, TOX, CD69, DUSP10, DUSP1, CCR9, SH2D3C, IL27RA, SLAMF1and BCL6) (Figures 4C, S4B and data not shown). Furthermore, we now provide new data confirming the increase at the protein level of two master regulators of T-cell differentiation (i.e. the transcription factors GATA3 and Aiolos) (Figure 3D).

Taken together, these results show that our Jurkat knocked down cells were analyzed at a time-point that is appropriate to identify TAL1-dependent genes important for T-cell differentiation against a background of apoptosis. Other evidence that further sustain our hypothesis that TAL1 disrupts the T-cell differentiation program in Jurkat cells include 1) the identification, using a de novo search, of an ETS motif that resembles the ETS motif associated to T-cell-specific enhancers (Hollenhorst PLoS Genetics 2009 (5)) (Figure 6A); and 2) the demonstration that TAL1 interacts with ETS1 (which binds to T-cell-specific enhancers) but not with GABPA, another ETS family member which is expressed in Jurkat cells but is targeted to promoters of housekeeping genes (Hollenhorst PLoS Genetics 2009 (5)) (Figure 6E).

Regarding the comment that RUNX1 and ETS1 which regulate T-cell genes are downregulated upon TAL1 KD, we note that overexpression of RUNX1 (also called AML1) is a signature of TAL1-expressing leukemic T-ALL cells (Ferrando et al. 2002 Cancer Cell (1)). In addition ETS1 is also overexpressed in human T-ALL (Sacchi et al. Leukemia, 1988 (2) 12-18). As such we propose that the decrease of RUNX1 and ETS1 that we observe (both at the transcript and protein levels) upon TAL1 knockdown may represent a return to a steady state necessary for T-cell differentiation.

3-Authors propose that TAL1 genomic sites are highly restrictive compared to MyoD as judged by number of total bound regions. It is difficult, and thus inappropriate, to compare overall number of peaks from one protein to other protein in ChIP-SEQ experiments, especially if the data analysis and read depth are not the same. Also, antibody variability is likely to contribute to differences without biological significance. 1000's of occupied "targets" is in the range of published data for

various transcription factors, so this point should be downplayed or even deleted.

We agree with the reviewer that a more detailed analysis and comparison would be necessary to support the biological significance of this initial observation and have deleted the subclause of the sentence referring to MyoD from the revised text.

4-Figure 4A GO terms in this figure need p values.

We have added the p values to the figure as requested.

5-p.10-The stated bias towards TSS in Figure 4B represents 50% of the peaks with >50% at greater distance. This should be stated more carefully to more accurately reflect the data.

We have modified the text to better reflect the data on Figure 4B. New text reads: "We also observed that while there is a higher local density of TAL1 peaks near TSSs in both erythroid and Jurkat cells (Figure 4B), the vast majority of TAL1 peaks are located away from promoter regions of known genes, mostly within introns and intergenic regions (Figure 4A)."

6-Figure 6 A Score for motifs is defined as z-value of motif, yet there is no definition of z-value in methods. See further comments below. What criteria are used on Figure 6 A to name Runx, GATA, etc.?

We have modestly changed the legend to Figure 6 and added more detailed Supplemental Methods to better describe the motif discovery and the assignment of the z-score. Motifs were given factor-binding specificity based on the Jasper and Transfac motif databases.

I strongly recommend that MEME analysis and a conventional PWM pix be included for the E-Box preference, the E-GATA composite and the E-ETS composite.

We understand the reviewer's concern about not using a "standard" motif discovery tool, however, we feel that MEME is not the best tool for our application. Some of the reasons to develop other motif analysis tools include:

1. MEME is simply not scalable enough for application on a large sequence dataset. One needs to cut the dataset to the order of hundreds of peaks instead of using the complete peak set because of memory and running time requirements.
2. We are interested in the problem of discriminative motif discovery, i.e., detecting motifs that best discriminate between two datasets. MEME is designed to find motifs that discriminate one dataset from a naïve background. As a result, MEME is susceptible to various biases in the dataset. For example, transcription factor ChIP-seq data tends to be enriched in the promoter region; therefore, motifs common to promoter regions are likely to be identified and might potentially disguise the motifs that distinguish bound and non-bound promoters. Under or over correction of GC background distribution also affects the output significantly, and repeats that are carried by a subset of sequences are likely to be reported as significant motifs. Such biases make it challenging to select the most relevant motifs post processing.
3. MEME tends to find long motifs that are overly specific because longer motifs are more informative and the longer motif increases the log likelihood ratio relative to the background. However, the specific longer motif may be appropriate for only a small subset of sequences relative to a more general form of the motif that might be shorter.

Instead of using one of the many motif discovery tools available today, we built our own tool specifically suited to the datasets we have to work with. We have now described our motif discovery in more detail in the revised Supplemental Methods. We apologize for not providing a more detailed description of the method in the original submission.

Also, the actual numbers of findings of these composite elements with constrained spacing should be mentioned explicitly in the text.

We have now added these numbers to Figure S5D and in the text of the manuscript (p.18)

The Hollenhorst et al. cited paper reported an ETS-RUNX composite element. It would be of interest whether this element is overlapping with the TAL1 bound regions.

As shown on Figure 6A this composite element was not identified using our de novo motif analysis of genomic regions under TAL1 peaks. When performing a guided motif search using the pattern "VGGAWGTGG" as a seed to find the ETS-RUNX composite motif, we find this motif to be present within TAL1 peaks that overlap with ETS1 and RUNX. However it is also enriched within TAL1 peaks that overlap with ETS1 but not RUNX. In conclusion, even though the composite "ETS-RUNX" motif is present in a subset of our TAL1/RUNX/ETS1 shared peaks, we cannot conclude that it is significantly enriched in TAL1/RUNX/ETS1 co-bound sites compared to TAL1/ETS1 (no RUNX) co-bound sites. In addition, we found non-overlapping ETS1 and RUNX sites for most of the ETS1/RUNX overlapping peaks, therefore, we hypothesize that the composite sites account for a small subset of the overlapping peaks.

7-Figure 6 E-Co-IP experiments need to be performed with an ethidium bromide control to eliminate the possible linkage through DNA-a common artifact of co-IP for two DNA binding proteins. There is no methods section for this experiment.

We have repeated all our co-IPs (and performed the new co-IP required by Reviewer 1) in the presence of ethidium bromide as described (Lai et al. PNAS 1992 (89) 6958). These experiments confirmed all the interactions we have shown in the previous version of this manuscript. Results of the co-IPs in the presence of ethidium bromide are now displayed on Figure 6E. In addition, we have added a new section on co-IP in the Supplemental Methods section.

8-page 17-The authors use published Hollenhorst et al genome-wide occupancy data for ETS1 and RUNX in Jurkat cells to report overlap with TAL1 bound regions. It is not clear why a range of percentages are given. It is not clear whether the proximity of TAL1 and ETS1, RUNX sites was analyzed by a program (such as InterceptRegions in the USEQ suite) that set an overlap interval or whether just linkage to same nearest gene was used. This would be very valuable in considering contention of direct functional consequence of the co-occupancy of factors. Also, a test for statistical significance, such as a Fisher's exact test, should be used on the frequency of dual occurrence TAL1 and RUNX1/3 or TAL1 and ETS1. It should be noted that the ETS1 bound regions represent over 10,000 genes and many T cell genes, thus, the overlap may not be functional relevant in many cases.

The range of percentages reflects the range of FDRs used in the following sentences of the first version of the manuscript: "Furthermore, genome-wide comparisons indicate that 50 to 61% of TAL1 peaks containing a Runx motif are bound by RUNX1/3 in jurkat cells (for FDRs ranging from 0.001 to 0.02 for the RUNX data) while 65 to 83% of TAL1 peaks containing an Ets motif are bound by ETS1 (for FDRs ranging from 0.0001 to 0.03 for the ETS DATA)." To compare TAL1 peaks identified by us with RUNX1/3 and ETS1 peaks identified by Hollenhorst et al. (PLoS Genetics 2009 (5)), we have processed the raw data from Hollenhorst et al. using our own pipeline (i.e. the same pipeline used to process TAL1 data). This was useful to examine the overlap at different stringency levels. In the first version of the manuscript, we had provided the percentage of overlap at different stringency levels (i.e. different FDRs), hence the range of percentages of overlap given. To avoid confusion in the new version of the manuscript, we provide the percentage of overlap only for the highest stringency level.

The overlap between TAL1 and RUNX or ETS binding was determined based on the interval +/- 500 nt from the peak of the TAL1 site, not by association with the nearest gene. Therefore, this is a very stringent association of the presence of sites and the factor binding. We have now described this in the Supplemental Methods section.

We agree that it is tempting to provide a number associated with the statistical significance of this association. However, it is difficult to know how to compare the observed binding to the null hypothesis. Assuming a random genome-wide distribution of binding events would yield a statistical

probability of near zero that these were chance co-occurences. On the other hand, it is clear that transcription factor binding sites do not distribute randomly on the genome, and it is therefore unclear to us how to realistically model this without bias. Therefore, we would prefer not to assign a significance value when we do not know the underlying distributions.
ß

Importantly, we have shown that knocking down RUNX1/3 or ETS1 leads to decrease of TAL1 binding (Figure 7D and Figure S8). This is an independent demonstration that the overlap of TAL1 with RUNX1/3 and ETS1 is functionally relevant.

9-ChIP SEQ data analysis has not been performed with current methodologies. Software and versions used have been upgraded in the field, such as the MACS program or the USEQ suite. These methods make simple local comparisons of reads from experimental vs. input that is superior to the global expectation methods cited. The gold standard for reads is 10 million whereas the reported data are based on 5-6 million reads. Although this has perhaps not seriously hampered the study, the investigators are encouraged to consider this upgrading.

As indicated in Figure S5A, there are over 10 million quality reads for each experimental group, with approximately 8 million aligned reads for each experimental group.

We agree with the reviewer's suggestion that using background local estimation will be helpful, but it remains an open question how to effectively use local background for peak-calling due to the fact that the background estimation may vary significantly based on the size of the local context. In this revision, we have re-calculated the FDR and P-values using a negative binomial distribution, which can be viewed as a Gamma-Poisson mixture and is described in the revised Supplemental Methods. This model provides us a more realistic P-value and FDR estimate based on control data set tests.

To further address the reviewers concerns, we have re-analyzed our data using MACS and derived essentially the same set of peaks. At a stringent P-value (10^{-10}), the MACS analysis identified 1890 peaks in JURKAT compared to the 2238 peaks in our analysis, with an overlap of 1705 peaks; and 5415 peaks in erythroid cells compared to the 5707 peaks in our analysis, with an overlap of 4873 peaks. Reducing the MACS P-value to 10^{-5} overlaps with nearly all of the peaks identified in our analysis (only 17 of the JURKAT peaks do not overlap and only 8 of the erythroid peaks do not overlap). Therefore, our analysis yields similar results to the MACS analysis.

Finally, we respectfully take exception to the suggestion that we are not using current methodologies. Several different approaches exist for ChIP-Seq data analysis, including the packages we have developed as open-source software for Bioconductor. For simple peak calling, many of these packages are relatively equivalent, as evidenced by our comparison to MACS, above. Our detailed understanding of the underlying process in our package makes this more suitable to us for more complex data comparisons.

10-Bioinformatics to find motifs under occupancy peaks was performed by older-style methodology that is somewhat word processing based. The Score outcome is not well described with the Z-value not being defined. In particular is there a contribution from length, complexity or strength of dinucleotide conservation? Also, the use of non-conventional motif presentation method is difficult to read compared to well-established methods and does not add significant information to the report. The new method is described awkwardly. The use of "depth" of edge colors is not clear-perhaps "darkness" in scale of black to grey would make sense. But where is the quantification of this variable? Newer more robust methods such as MEME with standard motif PWM format output should be used, at least for some critical observations (see above).

Please see response to Point 6 above for discussion of the choice of motif finding tools. As noted above, we have now more carefully described our motif identification in the revised Supplemental Methods, which we believe is an improvement over the existing MEME motif analysis. In response to the reviewer's questions, there is no specific contribution of the length, complexity, or dinucleotide conservation in the motif identification. Regarding the visual

presentation of dinucleotide preference, we agree that in these specific instances, this graphic presentation does not add, and might detract, from the visual image, therefore we now present the motifs in standard PWM form (see modified Figure S5D), as requested by the reviewer.

Minor:
Jurkat cell line should be capitalized throughout.

We have capitalized Jurkat throughout the manuscript.

2nd Editorial Decision 30 November 2010

I have received the comments on your revised manuscript from two of the referees that originally assessed the study. Both referees find that you have addressed all the previous concerns and now support publication in The EMBO Journal. I am therefore happy to accept the manuscript for publication, it will make a nice contribution to the Journal. You will receive the official acceptance letter in the next couple of days.

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #2
The authors have addressed the issues raised by the reviewers in a detailed and satisfactory way. The modifications have strengthened the manuscript that is now suitable for publication.

Referee#3
This manuscript provides novel insight into the mechanism by which TAL1 acts in leukemiogenesis by identification of direct targets. In addition, convergence of TAL1 targets with those of other T cell factors will be of general interest. The study nicely illustrates the value of combined genomic approaches including gene expression profiling, factor occupancy via ChIP-SEQ, and bioinformatics. The revised manuscript addresses all major concerns raised in my review. In particular, greater clarity of methodologies, reanalysis of ChIP-SEQ data and additional co-IP experiments improved the manuscript. Use of more conventional presentation format for binding site motifs will help communicate the findings to both genomics and transcription audiences.

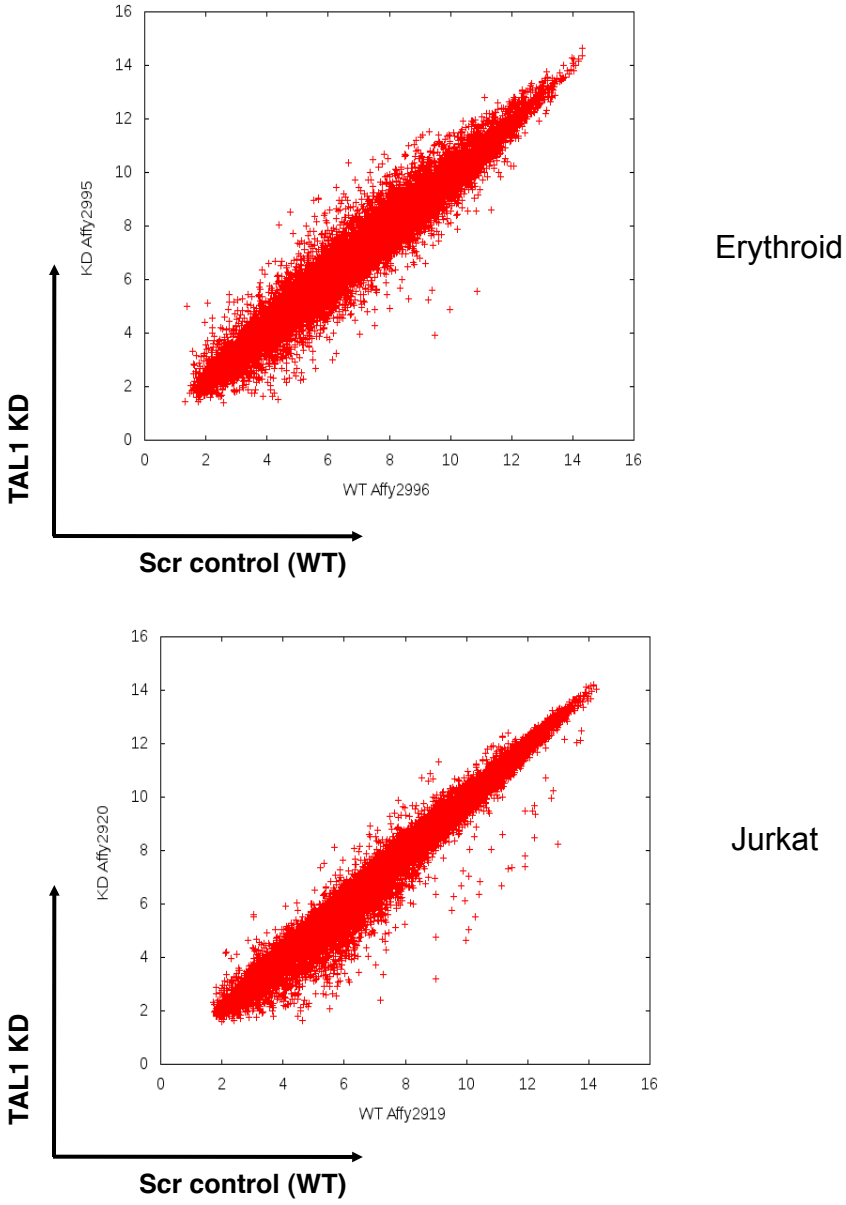

Figure 1 for Reviewer only
Normalized expression values on GeneChip® Human Gene 1.0 ST microarray
for WT vs TAL1 KD in erythroid and Jurkat cells. One replicate is shown for each cell type.

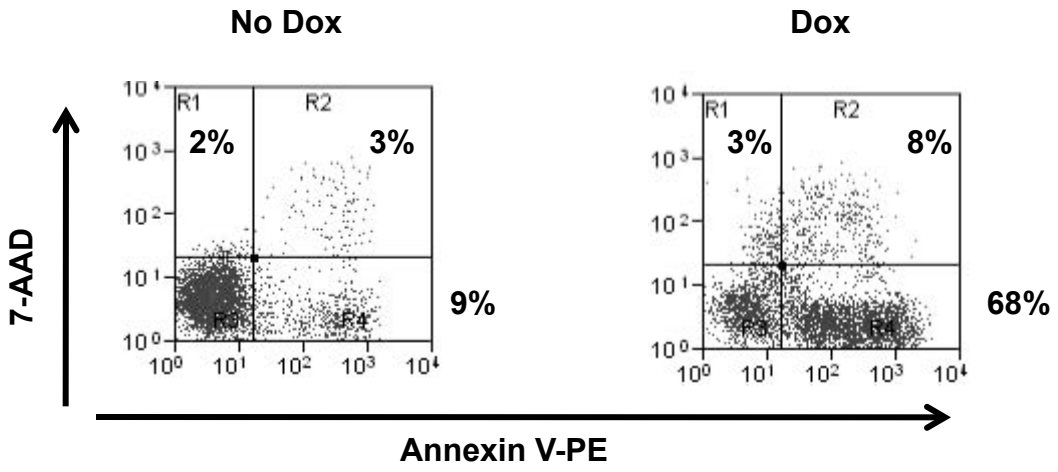

Figure 2 for Reviewer only
TAL1 KD leads to an increase in apoptosis in Jurkat cells

Jurkat cells were treated with Dox to induce the KD of TAL1. Apoptosis was measured by AnnexinV staining using FACS. Necrotic cells are stained with 7AAD. Numbers on the right of the histograms indicate the percentage of AnnexinV positive cells.

Note the low levels of necrotic cells (on the y-axis).