THE EMBO JOURNAL

Identification of a dynamic gene regulatory network required for pluripotency factor-induced reprogramming of mouse fibroblasts and hepatocytes

Dimitris Thanos, Maria Papathanasiou, Stefanos Tsiftsoglou, Alexander Polyzos, Deppie Papadopoulou, Panagiota Karagianni, Maria Pliatska, Iannis Talianidis, Marios Agelopoulos, Dimitrios Valakos, and Eleftheria Klagkou
DOI: 10.15252/embj.2019102236

Corresponding author(s): Dimitris Thanos (thanos@bioacademy.gr) , Marios Agelopoulos (magelo@bioacademy.gr)

Review Timeline:

Submission Date:	12th Apr 19
Editorial Decision:	14th May 19
Revision Received:	7th Jul 20
Editorial Decision:	13th Aug 20
Revision Received:	29th Aug 20
Accepted:	7th Sep 20

Editor: Stefanie Boehm

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. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Thank you for submitting your study on regulatory networks required for cellular reprogramming for consideration by The EMBO Journal. We have now received three referee reports on your study, which are included below for your information.

As you will see, the reviewers express an overall interest in the study, but they also raise several major concerns. In addition to several specific technical issues and ones regarding sufficient statistical testing, the referees find that the proposed 9 TF network should be explained and analyzed in further detail, for example by addressing the phenotype of knockdowns of all 9 factors (referee #1), or if these factors can replace OSKM during reprogramming (referee #2).

Should you be able fully address all issues raised by the referees, then we would like to invite you to prepare and submit a revised manuscript.

REFEREE REPORTS

Referee #1:

In their manuscript entitled "Construction of a dynamic gene regulatory network required for cellular reprogramming", Thanos and colleagues conduct transcriptional profiling to better understand the molecular logic underlying the formation of induced pluripotent stem cells (iPSCs). By comparing transcription factors (TFs) upregulated in both mouse embryonic fibroblasts (MEFs) and mouse embryonic hepatocytes (Heps) they identify a set of 9 shared TFs. Several of these (including the pluripotency regulators Nanog) have been previously studied in the context of induced pluripotency, while others such as Rcan1, Taf1c, Tead4 and Irf6 have not. Using a combination of bioinformatics, chromatin analysis and genetic (largely knockdown) approaches, they provide evidence of specific roles of these TFs in early to intermediate stages of iPSC formation. Overall, this is an interesting study with some noteworthy findings. Several observations of the authors with regards to reprogramming phases largely confirm prior studies, but such a confirmation is important and also supports the suitability of the authors' experimental approach. In addition, Thanos and colleagues provide several new lines of evidence for previously unknown roles and interactions of TFs, which should be of general interest. Therefore, there work seems in principle well-suited for publication in The EMBO Journal. Below are some points that I believe should be addressed.

Main concerns

1. It is an attractive feature of this manuscript that the authors do not only use MEFs for their studies. However, a satisfying characterization of the Heps used for reprogramming (morphology, marker genes) is missing. As fibroblastic cells are present in all preparations of embryonic tissues, the authors should provide evidence that indeed endodermal liver cells served as input for their experiments.

2. Why don't the authors take advantage of the fact that they have two different input cell types and investigate unique aspects of each? There seems to be a plethora of interesting questions that should be accessible in their data. For example, what are TFs upregulated only in one cell type (if there are any)? Do Heps have to undergo MET?

3. Reprogramming efficiencies (# of dox-independent iPSC colonies per input cells) should be provided for both MEFs and Heps.

4. The authors should provide a better description of their genetic experiments. What is the efficiency of KD for each gene? Does KD in Hep have the same consequences (if this cannot be done for any reason, the authors should at least be clear about it and discuss this)? What is the consequence of 9 TF KD in MEFs and established iPSCs?

5. No statistical tests are applied to the overexpression experiments in Suppl Figure 5B/C. These gain-of-function experiments appear central to the authors' conclusions and should ideally be shown in a main figure. It would also strengthen the manuscript to include OE experiments for all of the "new" TFs (or at least discuss this omission).

6. I am puzzled by Figure 1D. Does double KD of Rcan1 and Taf1c or of Nanog and Irf6 indeed not reduce reprogramming efficiencies (while single KD of these factors does)? The authors should clarify what is the reference point of their scale ("% reprogramming efficiency") and ideally include scrambled vector into this figure.

Additional comments

- The significance of the authors findings should be discussed in the context of several well-described approaches to achieve high efficient iPSC formation (Zviran et al., PMID 30554962; Vidal et al., PMID 25358786; Di Stefano et al., PMID 2433620).

- The authors should consider using a different phrase than "reprogrammable MEFs" for their virally transduced cells as this term usually indicates the use of transgenic (i.e. genetically homogenous) system.

- The purpose of the entire first results section (which only refers to supplemental data and does not directly relate to the central point of the role of the role of the 9 TFs) is not entirely clear. This could be shortened.

- Fig. 2F is not very intuitive. Axis and other labeling would help the reader understand what is shown here.

- Page 13 - the author refer to Klf4 as "constitutively expressed". Can KLF4 protein indeed be detected in MEFs? Does KLF4 KD in MEFs lowers expression of presumptive targets?

Referee #2:

In this manuscript, the authors investigated the early transcriptional events of cellular reprogramming by the Yamanaka factors OKSM(Oct4,Sox2,Klf4 and c-Myc) in MEFs and Hepatocytes, and identified a gene regulatory network composed of 9 transcriptional regulators(Cbfa2t3, Gli2, Irf6, Nanog, Ovol1, Rcan1, Taf1c, Tead4 and Tfap4) , which are the early binding targets for OKSM, marking the transition of a reprogrammable subpopulation cells from somatic cell state to pluripotency.

In general, the authors attempt to answer an important and unresolved problem and performed relevant experiments and detailed bioinformatics analysis. However, some issues need to be addressed before publication.

Major concerns:

1 The authors constructed a transcriptional factor regulatory network consisting of 9 transcription regulators, the reviewer wants to know whether those 9 factors can reprogram the MEFs or Hepatocytes into to iPSC instead of four Yamanaka factors(OKSM) or can substitute any or some of them? If the 9 factors are really the key direct downstream targets for OKSM in the early stage of reprogramming, maybe some of them can replace the OKSM totally or partially?

2 It's hard to understand what's the meaning in figure 2D. Is it a scatter plot for the expression of single gene in a population? But the Y-axis is note as a ratio of expression cell. The reviewer was confused by this and please clarify.

3 In the identified gene regulated network of the 9 TRs in Figure 4A, the binding of Oct4 to those genes seems not so clear, is not consistent with the center role of Oct4 in pluripotency induction, the authors should check the ChIP-seq data, whether or not Oct4 is binding to any among those 9 TRs.

4 The authors mainly use knocking down (KD) strategy to do functional test, the authors need to do at least one rescue experiment, for example, overexpressing Cbfa2t2, to confirm their conclusion.

5 The item "pre-iPSC" need to be defined in this manuscript, as other papers have published quite a lot on pre-iPSC, whether the "pre-IPSC " mentioned in this manuscript is as same as in the published ones?

Minor concern:

1 In figure 3F□G and I, the Sox2 ChIP-seq peaks seem quite different between the iPSC and ESCs, It's truth or somehow flexible of the ChIP-seq data? How many duplicates for this experiment?

Referee #3:

General summary and opinion about the principle significance of the study, its questions and findings:
The manuscript "Construction of a dynamic gene regulatory network required for cellular reprogramming" (EMBOJ-2019-102236) by Papathanasiou et al. describes the temporal analysis of gene expression during IPSC reprogramming in two experimental regimes. Concentrating on genes upregulated during this process the authors analyze the role of 9 transcriptional regulators. Although the scientific logic and the general approaches are plausible and the topic is interesting, the study is not able to substantiate the conclusions in its current form.
Specific major concerns essential to be addressed to support the conclusions:
- The manuscript at hand is not the first to analyze gene expression during IPSC reprogramming and/or determine transcription factor networks in this process. Although this is not on its own an issue, the authors should invest in comparing and discussing their data in this context (e.g. how do these TFs map onto reprogramming intermediates published previously (Lujan, O'Malley, Polo, Schiebinger etc.)).

- It is unclear for several experiments if conclusions can be drawn, since effects are minor and statistical tests are missing. This is found on several locations, but exemplified best in: "Supplementary Fig S5B shows that high-level expression of Tfap4 and Irf6 increased reprogramming efficiency". Supplementary Fig.5B suggests an improvement of reprogramming 0.3% over control which is hardly notable, the data has no error bar or standard deviation. Classic experiments, such as substitution of reprogramming TFs with any of their candidates, etc. could help.

- The TR model is problematic in several ways, it is not immediately apparent what data went into it and what and whether it can predict more than that. But also the underlying basic- and follow up

validation-experiments are problematic in their current representation. This has several reasons:
o The authors seem to lack a general concept for base-line/negative expression in fibroblasts and proof of strong/functional expression in Reprogramming/IPSCs for these 9 factors. In the microarrays relative (background) expression (d0) of the different factors varies from 0 to 800 and "strong" expression from 4 to 3000 (d6).
This is especially relevant,
• because in the model and sc-qPCR they are classified digitally (as expressed/present and not expressed/absent), in SF1 as high and low expressed and in Figure 4 as % expressed. The authors have to clarify this and show that similar mRNA levels are not depicted as absent, lowly expressed or 500% induced depending on the gene.
• because the induction is often minor (TAF1C, TEAD4);
• Because the normalization is unclear
(What does "plotted according to the median expression value" mean?
Are values shown e.g. Figure 1C normalized to the expression minus Dox?
Does "Signal intensity at time point T dox+/ Signal intensity
at time point 0 {MEFs} >=2" & "Signal intensity at time point T dox-/ Signal intensity at time point 0 {MEFs} <2" mean you included genes induced 2.1 fold with DOX and 1.9 fold without?)
• because multiple probes corresponding to the same gene on the microarray were filtered out "and only those with the highest signal intensity across all time points were kept for further analysis". (Why did you not use the average, how would it change your data?)
• and because protein expression (western blot/ Chip) is not shown.
o Including up and down-regulation to different levels in both directions between all gene pairs makes the model visually very messy and uninformative.

Minor concerns that should be addressed:
- What happens to the KD cells? Do they die, do they retain MEF characteristics, do alternative reprogramming cell fates become upregulated - in other words is this a specific phenotype for impaired reprogramming or some sort of general defect? Sequencing or expression profiling (other than Lin28 and EpCam) may be helpful.
- Are the phases derived from transcriptomes normalized to the -Dox control? Do you see similar phases in the transcriptome of those cells?
- How do you define transcriptional regulators? What is the source for this gene list?
- Is the Reprogramming block of Pycard reproducible and significant?
- Figure 2E, label is misleading
- Why is the scaling in the ChIP changing in each panel (3-40, 2-50, 3-70 etc). Although this is overall convincing data it is suboptimal for analysis.

Response to Reviewers

Reviewer #1:

Overall, this is an interesting study with some noteworthy findings. Several observations of the authors with regards to reprogramming phases largely confirm prior studies, but such a confirmation is important and also supports the suitability of the authors' experimental approach. In addition, Thanos and colleagues provide several new lines of evidence for previously unknown roles and interactions of TFs, which should be of general interest. Therefore, their work seems in principle well-suited for publication in The EMBO Journal.

We thank the reviewer for his/her comments and the overall positive feedback that we received in order to improve our study. We appreciate the fact that Reviewer#1 highlights the value of our work towards a better understanding of the molecular mechanisms controlling cellular reprogramming.

Main concerns

1. It is an attractive feature of this manuscript that the authors do not only use MEFs for their studies. However, a satisfying characterization of the Heps used for reprogramming (morphology, marker genes) is missing. As fibroblastic cells are present in all preparations of embryonic tissues, the authors should provide evidence that indeed endodermal liver cells served as input for their experiments

The reviewer asks a very important question regarding the purity of the mouse hepatocytes (mHeps-endodermal nature) used in our experiments, since it is well known that embryonic tissues are heterogeneous regarding the types and developmental stages of the cells present. To address this important question, we have carried out a series of additional experiments to characterize the mHeps preparations used in our reprogramming experiments. We combined the results from different assays (Gene Annotation analysis of transcriptomics experiments, immunofluorescence and morphological analysis) and have confirmed the true endodermal identity of the mHeps cultures used in the experiments. More specifically:

1) We carried out a detailed Gene Annotation meta-analysis of our mHeps transcriptomics data and verified that the top category of the most highly expressed genes corresponds to liver-specific genes **(Figure RB1 and Table RB1)**. More specifically, we found that 772 out of the 2.500 most highly expressed genes in mHeps are liver-specific genes. Therefore, the gene signature of our mHeps preparations correlates with this of bona-fide liver cells, thus confirming the molecular identity of the mHeps cultures utilized in this study.

2) Next, we examined our mHeps cultures by immunofluorescence in order to visualize the spatial expression of known Hepatocyte markers. As shown in **Figure RB2**, our mHeps cultures are homogeneous and the expression of hepatic markers, such as HNF4α and Albumin, was the prominent feature of these cells, thus verifying their hepatic origin and the purity of the cultures. Finally, it is important to mention that all mHeps cultures used in our reprogramming studies were isolated and cultured by using the same methodology without any prior storage. The detailed protocol is described in the **Materials and Methods** section of the original manuscript under the heading: *Isolation of mouse Hepatocytes (mHeps)*.

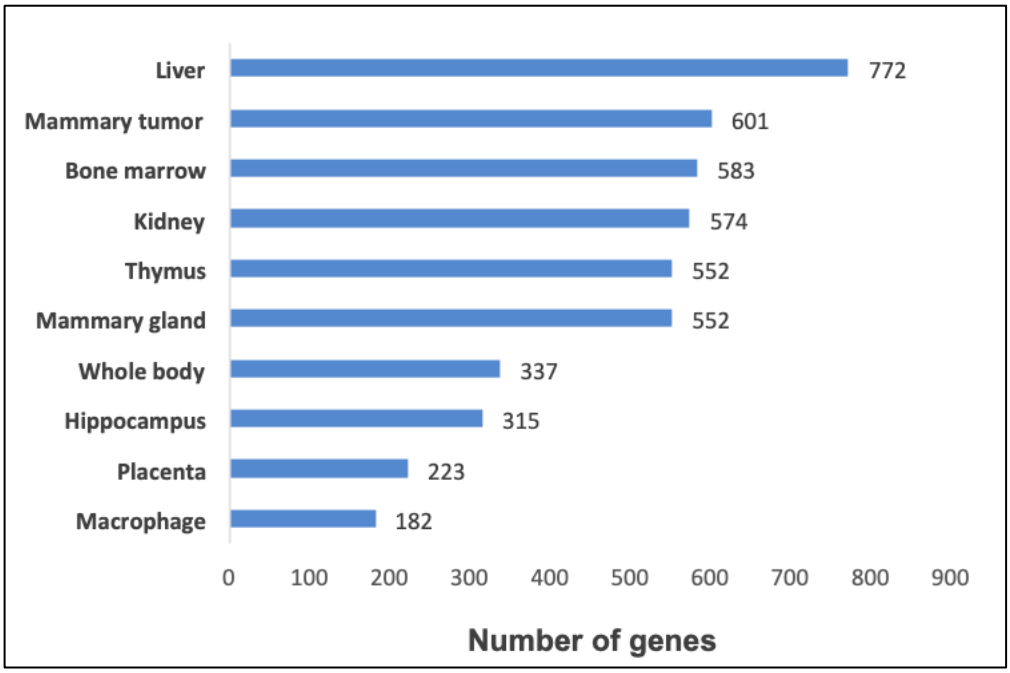


Figure RB1: Gene Annotation analysis of the 2.500 most highly expressed genes in mHeps as revealed by our transcriptomics studies using the "DAVID knowledgebase". Shown are the top 10 most significantly enriched tissue terms (p-value< 1e-3). The Gene Annotation panel depicts that the dominant group of genes expressed in our mHeps cell preparation corresponds to liver-specific genes

Liver-specific genes		
ACAA2	FN1	NDRG1
ACADM	GALK1	NID1
ANXA5	GM2A	PSAP
BNIP3	GNB1	RPL36
BNIP3L	H6PD	RPS12
C3	HMGCS1	SAP30
COL1A1	KRT8	SERPINF1
COX7A2	LPL	ST6GAL1
EI24	LRRFIP1	TAGLN
EXOC4	MAN2A1	TAPBP

Table RB1: Shown are the most highly expressed genes identified in our mHeps cultures that are classified as liver-specific genes.

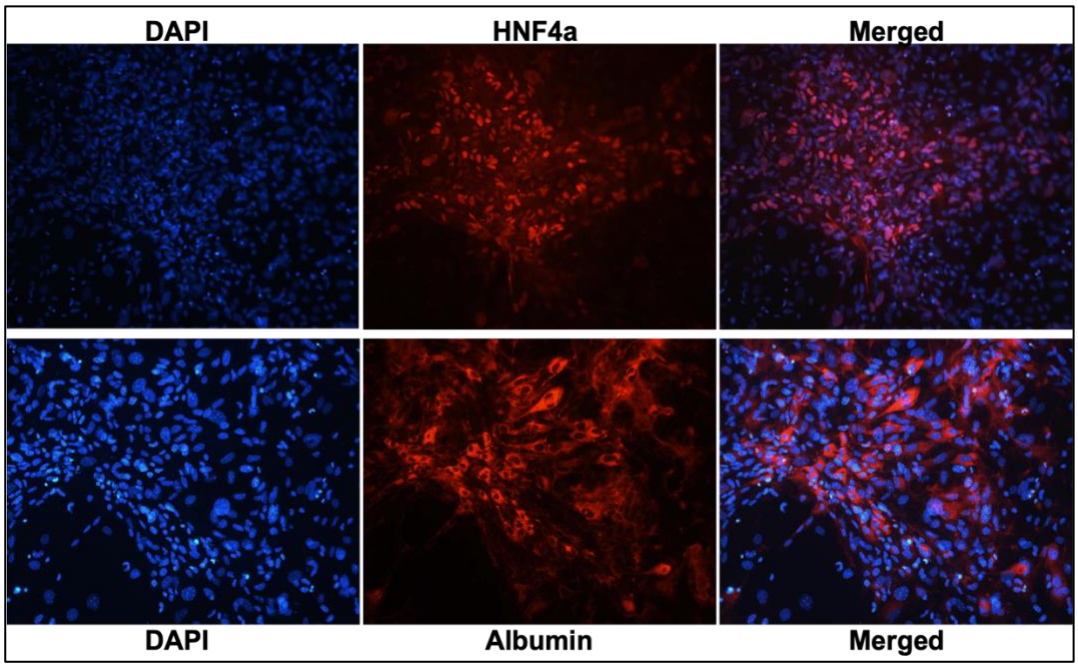


Figure RB2: Shown are immunofluorescence assays using antibodies against HNF4a and Albumin in mHeps cultures confirming the homogeneity of the isolated hepatocytes used in cellular reprogramming assays.

Taken together the above experiments strongly suggest that our mHeps reprogramming experiments were carried out using pure endodermal liver cells as judged by the transcriptomics, immunofluorescence and the morphological data presented above. We have included this information in the revised version of the manuscript as "unpublished data".

2. Why don't the authors take advantage of the fact that they have two different input cell types and investigate unique aspects of each? There seems to be a plethora of interesting questions that should be accessible in their data. For example, what are TFs upregulated only in one cell type (if there are any)? Do Heps have to undergo MET?

The reviewer raises a number of important and interesting points regarding the existence of cell-type-specific reprogramming pathways. He/she proposes that since we have carried out side-by-side reprogramming experiments using two different starting cell types (MEFs and mHeps), we are in a unique position to address this question. We agree with the reviewer, but we feel that these experiments fall beyond the scope of this manuscript. Herein, we have discovered and characterized a Gene Regulatory Network (GRN) that can be reconstructed in at least two different somatic cell types in order to achieve pluripotency. Our hypothesis was based on the very important fact that since nearly every cell type can be converted to iPSCs, there should exist common GRNs in different cell types, without of course excluding the existence and/or involvement of cell type-specific pathways. Indeed, the experiments presented in this manuscript suggest the existence of such "universal Gene Regulatory Networks" built by transcriptional regulators. **Nevertheless, we appreciate the reviewer's genuine interest regarding cell-specific transcriptional regulatory networks. We describe below some of our initial findings regarding the "*existence of TFs upregulated in one cell type*".**

Figure RB3 depicts comparisons between MEFs and mHeps of (A) all Differentially Expressed Genes (DEGs) and (B) Transcriptional Regulators only (TRs) (Figure RB3A and RB3B, respectively), as captured from Day 0 to Day 6 of reprogramming. As shown in Figure RB3A, the number of DEGs in MEFs (1,507) is ~2 times higher than in mHEPs (747). This finding is consistent with the fact that

mHEPs exhibit epithelial characteristics, and thus are “closer” to the pluripotent cell state phenotype as compared to MEFs (see also below).

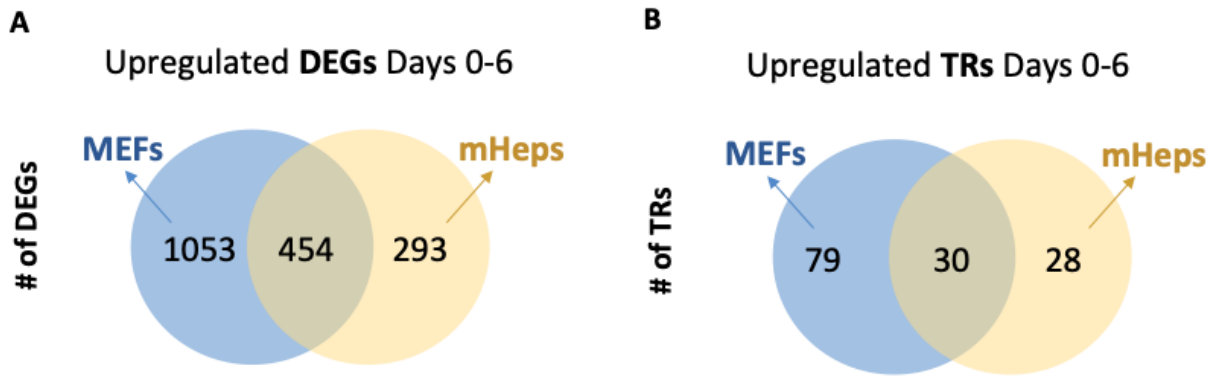


Figure RB3: Venn diagrams showing the number of cell type-specific and common upregulated DEGs (A) and TRs (B) between MEFs and HEPs during the first 6 days of the reprogramming process.

Interestingly, Gene Ontology (GO) and signal transduction pathway analyses of the cell type-specific DEGs revealed that biological processes related to cell cycle and NOTCH signaling pathways are enriched in MEFs, whereas in mHeps metabolic-related processes and cell fate determination pathways are predominant (**Table RB2)**.

MEF specific Genes		
GO Term	**#Genes**	**Representative genes**
GO:0007049~cell cycle	94	BRCA1, MAP3K8, PRC1, MYBL2, RAD21, AURKA, MITD1, AURKB, TIMELESS, CDT1
GO:0045746~negative regulation of Notch signaling pathway	7	DLX2, SLC35C1, HEY1, NRARP, NOTCH4, DLK2, RFNG
GO:0008593~regulation of Notch signaling pathway	4	DTX1, NOTCH4, LFNG, LLGL2
GO:0000082~G1/S transition of mitotic cell cycle	12	CCNE2, CCNE1, CCND1, CDKN2D, POLE, SKP2, IQGAP3, RPS6, RCC1, RHOU, CDCA5, CABLES1
mHEP specific Genes		
GO Term	**#Genes**	**Representative genes**
GO:0006629~lipid metabolic process	18	SCD3, CPT2, PLBD1, GDPD3, LPIN1, GDPD2, APOA1, HMGCS2, APOC3, LIPG
GO:0001709~cell fate determination	3	HOXA2, CHRDL1, DLL3
GO:0008202~steroid metabolic process	5	APOA1, HMGCS2, HSD11B1, SRD5A2, DHCR24
mmu00480:Glutathione metabolism	6	LAP3, GPX2, OPLAH, GSTO1, G6PD2, GSTM7

Table RB2: Table showing representative genes of pathways or biological processes that are enriched during MEFs and mHeps reprogramming to pluripotency.

Interestingly, we found that although the TGF-beta and WNT pathways are downregulated during reprogramming in both cell types, different genes in MEFs and mHeps from each pathway are affected, a result consistent with the operation of cell type- and gene-specific molecular pathways converging to similar biological processes. Of note, apart of the 30 TRs shared between MEFs and mHeps described in the original version of the manuscript, there are 79 TRs upregulated in MEFs only **(Figure RB3B and Table RB3A)** and 28 upregulated specifically in mHeps **(Figure RB3B and Table RB3B)**. In the revised version of the manuscript we have discussed the implications of these findings.

A

MEF uniquely upregulated genes			
4933413G19 Rik	Ezh2	Notch4	Taf13
Ankrd10	Gli1	Nrarp	Taf1b
Arnt2	Gm98	Patz1	Taf4a
Arntl2	Gmnn	Pcbd1	Taf7
Atf3	Grhl1	Phf5a	Tcf7
Bach2	Grhl2	Pmf1	Tcfap2a
Bcl3	Gtf3c4	Pold3	Tcfap2c
Bnc1	Hey1	Psmg4	Tcfcp2l1
Brca1	Hmx1	Rbpms	Tcfeb
Ccne1	Hoxa4	Rcor2	Tle4
Cdkn2d	Ilf2	Rfc1	Trip13
Cdyl	Klf10	Runx3	Tub
Cep290	Klf16	Ruvbl2	Uhrf1
Crebbp	Lmo1	Sap30	Wdr77
Dek	Med30	Sim1	Wwc1
Depdc1a	Med4	Sox13	Zc3h8
Dlx2	Mybbp1a	Sox3	Zfp277
Dtx1	Mybl2	Ssbp2	Zfp57
Erf	Myef2	Supt16h	Zfp593
Etv1	Nfe2l3	Suv39h2	

B

mHEP uniquely upregulated genes	
Arnt	Smarca2
Baz2a	Stat5a
Cbx2	Tob1
Dbp	Trps1
Egr3	Tshz1
Elf3	Zfp105
Elf5	Zfp219
Fos	Zmiz2
Fosl2	
Hoxa2	
Id4	
Irx3	
Irx5	
Jund	
LOC100044751	
Mll3	
Mxd3	
Neurog3	
Rel	
Runx1t1	

Table RB3: Table showing TR genes that are specifically upregulated during the first 6 days of reprogramming in MEFs (**A)** and mHeps (**B)**.

Regarding the reviewer's important question whether the mHeps undergo MET to reach pluripotency, we note that hepatocytes are epithelial cells of the liver (Choi and Diehl, 2009) and therefore, they don't have to undergo MET to achieve pluripotency. This information has been included in the revised version of the manuscript.

3. Reprogramming efficiencies (# of dox-independent iPSC colonies per input cells) should be provided for both MEFs and Heps.

The relevant Reprogramming Efficiencies are provided in the **Materials and Methods** section under the heading "*RNAi functional assays in MEFs undergoing reprogramming"* in the revised version of the manuscript.

4. The authors should provide a better description of their genetic experiments. What is the efficiency of KD for each gene? Does KD in Heps have the same consequences (if this cannot be done for any reason, the authors should at least be clear about it and discuss this)? What is the consequence of 9 TF KD in MEFs and established iPSCs?

The reviewer makes a valid point requesting a better description of the knockdown (KD) experiments described in the manuscript. In the revised version of the manuscript, we describe in detail the experimental procedures to generate KD cells in the **Materials and Methods** section under the heading "*RNAi functional assays in MEFs undergoing reprogramming"*. Briefly, we evaluated the KD efficiency for each of the Transcriptional Regulators (TRs) by qPCR to detect the abundance of the corresponding transcript in comparison to the levels of a KD-control sample bearing a scramble sequence (% expression). These data were plotted side-by-side with the resulting Reprogramming Efficiencies (REs) of the corresponding KD cells. REs (%) were calculated upon normalization to the RE (%) of the control KD vector bearing a scramble sequence. This additional information has been incorporated in the **new Figure 1B** of the revised manuscript. Briefly, the data clearly demonstrate that the efficiency of each TR's KD does not correlate with its effect in reprogramming efficiency (i.e. the reprogramming efficiency is not proportional to the KD efficiency). This observation highlights the true impact of the TRs in the reprogramming process.

In addition, the reviewer raises a very important point, which has been clarified also in the revised version of the manuscript regarding the effect of the KDs in mHeps. Our 9TRs KD experiments in mHeps produced qualitatively similar results with those in MEFs. It should be mentioned, however, that accurate measurements of the Reprogramming Efficiencies (REs) in mHeps couldn't be routinely performed by Alkaline Phosphatase (AP) staining, due to high endogenous levels of the enzyme in these cells. However, we have readily observed iPSCs colonies in mHeps on Day 21, as we did in MEFs. Thus, the results obtained from the KD experiments in mHeps are in perfect correlation with those in MEFs underscoring a universal role for the 9TRs in cellular reprogramming.

Regarding the question related to the consequences of the 9TRs' KD in naïve MEFs, we carried out KD experiments for the 4 TRs (Rcan1, Taf1c, Tead4, and Tfap4), which are expressed in naïve MEFs at basal levels. Our data showed that none of these KDs significantly affected cell proliferation **(Fig. RB4)** or caused cell apoptosis **(Fig. RB5)**.

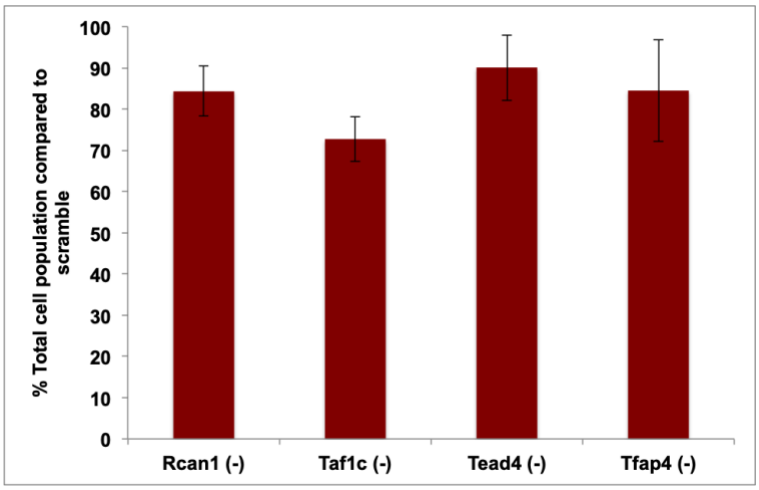


Figure RB4: Diagram depicting the total MEFs population after KD of the indicated TRs.

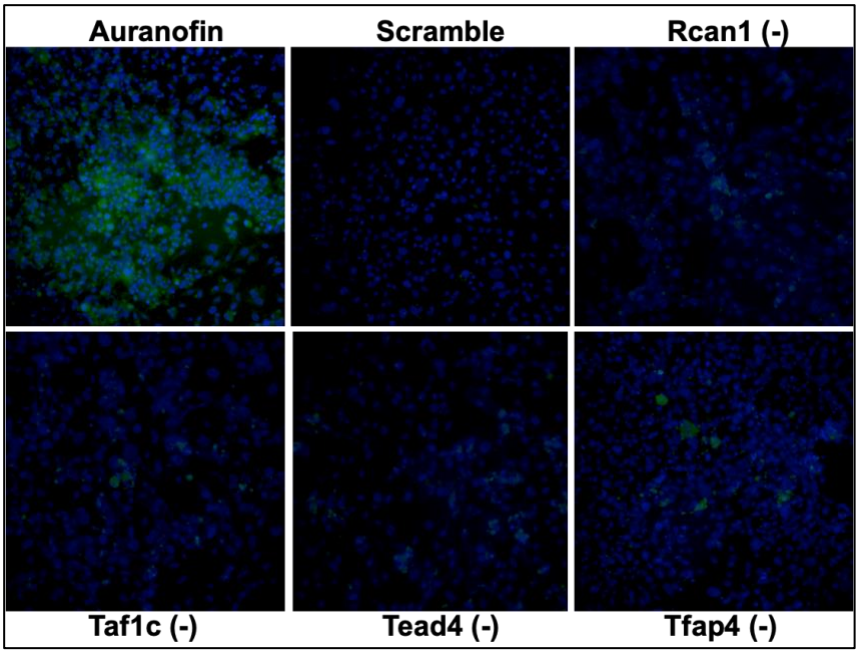


Figure RB5: Test for apoptosis by cleaved Caspase-3 immunostaining in KD-MEFs. The green signal corresponds to apoptotic cells. Treatment of cells with Auranofin induces cell apoptosis (Marzano C et al., 2007; Park N and Chun YJ, 2014) and it was used as a positive control in our assays.

We have also extended our analysis of the effects of the 9TRs KD on cell proliferation and viability using MEFs undergoing reprogramming (Day 6), which express all 9TRs. As seen in **Figure RB6** none of the 9TRs affected cell proliferation or viability.

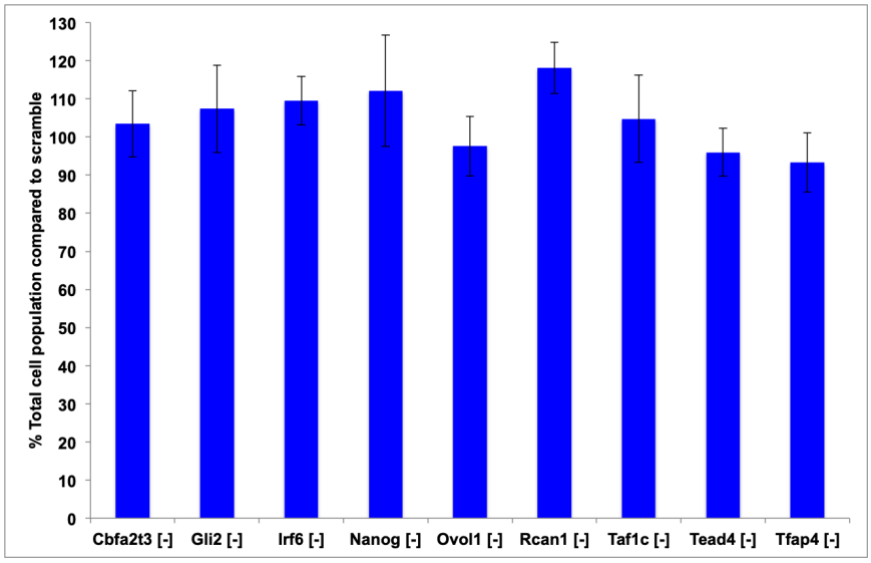


Figure RB6: Total number of cells on Day 6 of reprogramming as transduced with the indicated shRNAs of 9TRs.

Taken together, our experiments exclude any non-specific effects and further underscore that the cell reprogramming phenotypes obtained from the 9TRs KDs are due to the dysregulation of the relevant gene expression programs. We have included this information in the revised version of the manuscript as “unpublished data”.

We did not carry out similar KD experiments in iPSCs or ESCs for the following reasons: first, 2 out of the 9TRs (Ovol1 and Cbfa2t3) are expressed at practically undetectable levels in iPSCs or ESCs and second, we believe that these experiments, although very interesting, are related to the role(s) of the TRs in maintaining and not acquiring the pluripotent cell state, a question that is clearly beyond the scope of our manuscript. Previous studies in ESCs have demonstrated that the core pluripotency network is robustly stable and in many instances can be guarded by the compensatory overlapping activities of other regional TR subnetworks (Jaenisch and Young, 2008; Ng and Surani, 2011; Young, 2011; Nishiyama et al., 2013).

5. No statistical tests are applied to the overexpression experiments in Suppl Figure 5B/C. These gain-of-function experiments appear central to the authors' conclusions and should ideally be shown in a main figure. It would also strengthen the manuscript to include OE experiments for all of the "new" TFs (or at least discuss this omission).

As per reviewer’s request, we have modified old Supplemental Figure 5, which is shown as the new main Figure 5 in the revised version of the manuscript.

6. I am puzzled by Figure 1D. Does double KD of Rcan1 and Taf1c or of Nanog and Irf6 indeed not reduce reprogramming efficiencies (while single KD of these factors does)? The authors should clarify what is the reference point of their scale ("% reprogramming efficiency") and ideally include scrambled vector into this figure.

The reviewer raises a beautiful question: he/she asks how it is possible that the single KDs of Rcan1 and Taf1c or Nanog and Irf6 reduce reprogramming efficiency, whereas their double KD has practically, no effect. This is a truly important question applying to many genetic experiments for

the analysis of gene regulatory networks. An answer to this question should consider that the total transcriptomic output of gene regulatory networks, and thus their overall biological function (in our case reprogramming efficiency), signifies the integration of the individual effects (positive and negative) of the various connections of genes, involving a functional interplay between their outputs. These networks are stable in a given cell type but they could undergo rewiring under certain conditions, like cell transformation, differentiation etc. As gene regulatory networks are divided into functional sub-circuits and consist of multiple layers of regulatory interactions, it is expected that the hierarchy of their overall structure and internal interactions would define their functional properties. Indeed, to directly answer the question raised by the reviewer, we considered the complexity of the interactions between different TRs in the 9TR GRN and noted that Rcan1 and Taf1c separately interact with Irf6 eliciting opposite effects on its expression. This implies the deployment of antagonistic actions on Irf6, whereas Irf6 itself, being at the heart of the GRN, establishes a multitude of additional positive and negative interactions by receiving inputs from nearly all TRs in the GRN.

It is also important to emphasize that since the 9TR GRN is constructed in a stepwise manner during the first 6 days of reprogramming, the effects of our KD experiments should be interpreted not in the context of the already established 9TR GRN, but most importantly in the context of the dynamic construction of the network. Thus, the loss of any GRN component(s) by KD in the beginning (Rcan1 and Taf1c) or later (Irf6 and Nanog) during the dynamic assembly of the 9TR GRN, could alter the balanced expression of the other TRs of the network, and as such the reprogramming efficiency. In addition, the network could become stabilized in a different balanced state by rewiring its connections. Evidently, our experiments did not address the robustness of the 9TR GRN, because these experiments are beyond the scope of this manuscript. However, we feel that this point as raised by the reviewer is very important and have modified the relevant sections of the manuscript to address this issue.

Regarding the question for the reference point used to prepare the heatmap in Figure 1D, we note that all Reprogramming Efficiencies (REs) were calculated against the REs obtained for the scramble vector for the single KDs, whereas for double KDs, the REs were evaluated against the RE for the corresponding single KDs.

Additional comments

- The significance of the authors' findings should be discussed in the context of several well-described approaches to achieve high efficient iPSC formation (Zviran et al., PMID 30554962; Vidal et al., PMID 25358786; Di Stefano et al., PMID 2433620).

As per reviewer's request, we performed additional comparative meta-analyses on our mHeps transcriptomics data and found a considerable downregulation of both the TGF-beta and Wnt signaling pathways **(Table RB4)**. This finding is consistent with the published data of Vidal et al (Vidal et al., 2014) suggested by the reviewer, in which hepatoblast reprogramming is greatly facilitated upon inhibition of the TGF-beta signaling pathway. Thus, these data confirmed the validity of our reprogramming platform (related to our answer to **Point 2,** see above**)**.

	GOTERM/PATHWAY	-log10(pvalue)	pvalue	Genes
Day 3	mmu04350:TGF-beta signaling pathway	3.245200597	5.69E-04	BMP4, INHBA, SMAD7, SMAD6, FST, THBS1, TGFB2, BMP6, PITX2
	GO:0090263~positive regulation of canonical Wnt signaling pathway	1.480668655	0.033062169	WNT2, WNT4, ILK, LRRK2, GPRC5B
	mmu04310:Wnt signaling pathway	1.027260349	0.093916014	WNT2, PRKCA, WNT4, PLCB4, WNT11, FZD4, PRKCB
Day 6	GO:0035567~non-canonical Wnt signaling pathway	2.966930472	0.001079119	FZD8, VANGL2, FRZB, FZD6
	GO:0060071~Wnt signaling pathway, planar cell polarity pathway	1.851969968	0.014061448	VANGL2, FZD3, FZD6
	GO:0060070~canonical Wnt signaling pathway	1.625870244	0.023666267	FZD8, SOX4, SMAD3, FZD3, FZD6
	GO:0090090~negative regulation of canonical Wnt signaling pathway	1.43768173	0.036502135	NKD1, GPC3, CDH2, FRZB, FZD6

Table RB4: Gene Ontology (GO) and Pathway Analysis of genes that are downregulated during reprogramming of mHeps showing enrichment of the Wnt and TGF-beta pathways.

-The authors should consider using a different phrase than "reprogrammable MEFs" for their virally transduced cells as this term usually indicates the use of transgenic (i.e. genetically homogenous) system.
The reviewer is right. We have changed the wording to: *"MEFs undergoing reprogramming"* in the revised version of the manuscript.

- The purpose of the entire first results section (which only refers to supplemental data and does not directly relate to the central point of the role of the role of the 9 TFs) is not entirely clear. This could be shortened.
As per reviewer's advice, we have shortened this section in the revised version of the manuscript.

- Fig. 2F is not very intuitive. Axis and other labeling would help the reader understand what is shown here.
The reviewer is right. In the revised version of the manuscript we provide a better description of old Figure 2F, which has become Figure 2G.

- Page 13 - the author refer to Klf4 as "constitutively expressed". Can KLF4 protein indeed be detected in MEFs? Does KLF4 KD in MEFs lowers expression of presumptive targets?
As per reviewer's request, we show that the Klf4 protein is indeed expressed at low levels in naïve MEFs and, as expected, c-Myc protein is also constitutively expressed. Figure EV2B of the revised version of the manuscript clearly depicts that the protein expression data are in agreement with the ChIP-seq data showing that Klf4 binds to the Tfap4 regulatory region in naïve MEFs (Figure 3I), forming the groundwork of the 9TR GRN. Our observations are also in agreement with two previous publications showing that Klf4 is indeed expressed in MEFs (Polo et al., 2012) and is capable of DNA binding *in vivo* (Chronis et al., 2017).

In addition, the reviewer asks: what is the effect on Klf4 KD in naïve MEFs? We are currently addressing this question by performing RNA-seq experiments in Klf4 KD cells. These transcriptomics data will be integrated/ to Klf4 ChIP-seq data to identify genes bound and regulated by Klf4. However, these experiments are far beyond the scope of our manuscript.

Referee #2:

In this manuscript, the authors investigated the early transcriptional events of cellular reprogramming by the Yamanaka factors OKSM (Oct4,Sox2,Klf4 and c-Myc) in MEFs and Hepatocytes, and identified a gene regulatory network composed of 9 transcriptional regulators(Cbfa2t3, Gli2, Irf6, Nanog, Ovol1, Rcan1, Taf1c, Tead4 and Tfap4), which are the early binding targets for OKSM, marking the transition of a reprogrammable subpopulation cells from somatic cell state to pluripotency. In general, the authors attempt to answer an important and unresolved problem and performed relevant experiments and detailed bioinformatics analysis. However, some issues need to be addressed before publication.
We thank the reviewer for his/her fair and positive evaluation of our work and for the overall constructive criticism. We found the concerns and critique raised by the reviewer valid, useful and important. Therefore, we have addressed them comprehensively by performing additional experiments and have modified the manuscript accordingly.

Major concerns:

1. The authors constructed a transcriptional factor regulatory network consisting of 9 transcription regulators, the reviewer wants to know whether those 9 factors can reprogram the MEFs or Hepatocytes into to iPSC instead of four Yamanaka factors (OKSM) or can substitute any or some of

them? If the 9 factors are really the key direct downstream targets for OKSM in the early stage of reprogramming, maybe some of them can replace the OKSM totally or partially?
The reviewer makes a valid and important point. He/she wants to know whether the transcription factors constituting the 9TR GRN can substitute for any or some of the 4 Yamanaka factors (OSKM) in inducing cellular reprogramming. A positive answer to this question will further validate the main finding of our work, that is, for a critical role of the 9TR GRN in cellular reprogramming. Although the assumption behind the proposed experiment is straightforward, it should be kept in mind that the 9TR GRN, as it is the case for any GRN, is not a simple flow diagram representing just epistatic relationships between OSKM and the implicated transcription factors. It is well established that OSKM have additional gene targets, which together with the 9TRs and other factors affect cellular reprogramming. Rather, the 9TR GRN represents the appropriate integration of components sharing one common characteristic, that is, they are direct downstream targets of OSKM. The dense network of positive and negative interactions between different TRs indicate the existence of distinct thresholds of TR activity. This fine balance determines cellular reprogramming depending on multiple layers of molecular mechanisms of genetic and epigenetic regulation.

However, we took the reviewer's proposal to heart and carried out new reprogramming experiments to test whether any of the 9TRs can substitute for c-Myc, since Oct4, Sox2 and Klf4 have well-defined genomic targets previously known to be critical for reprogramming. The new data are shown in Figure EV2A of the revised version of the manuscript and have been appropriately discussed. Briefly, we showed that although the absence of c-Myc causes a significant delay in the reprogramming course and a decrease of the reprogramming efficiency as it has been previously described (Wernig et al., 2008; Nakagawa et al., 2008), overexpression of some of the 9TRs such as Cbfa2t3, Ovol1, and Gli2 together with O/S/K results in the re-establishment of the kinetics of the process

2. It's hard to understand what's the meaning in figure 2D. Is it a scatter plot for the expression of single gene in a population? But the Y-axis is note as a ratio of expression cell. The reviewer was confused by this and please clarify.
The reviewer was confused regarding the meaning of Figure 2D. He/she is absolutely right. We have rewritten the text in the main manuscript and modified the figure legend to clarify the meaning and significance of the data shown.

3. In the identified gene regulated network of the 9 TRs in Figure 4A, the binding of Oct4 to those genes seems not so clear, is not consistent with the center role of Oct4 in pluripotency induction, the authors should check the ChIP-seq data, whether or not Oct4 is binding to any among those 9 TRs.
The reviewer makes again a valid point that we have addressed in the revised version of the manuscript. He/she wonders for the role of Oct4 in regulating the expression of the 9TRs. This question emerged because Oct4 DNA binding to these genes was not prevalent in the original version of Fig 4A. However, in the modified Figure 4A, new Figure 4B of the revised manuscript we have added the ChIP-seq data from the 18 hour time point, where it is clearly depicted the robust Oct4 binding to the genes of the 9TR GRN from the beginning of reprogramming. These new results, underscore the central role of Oct4 in regulating TRs expression, as pointed out by the reviewer. As seen in the figure, at this early time point, Oct4 binds not only to the Cbfa2t3, Gli2, Rcan1, Tead4 and Tfap4, but also to the promoters of all 4 Yamanaka factors (Figure 3 and Figure EV4A of the main manuscript).

4. The authors mainly use knocking down (KD) strategy to do functional test, the authors need to do at least one rescue experiment, for example, overexpressing Cbfa2t2, to confirm their conclusion.
The reviewer proposed a beautiful, but technical challenging experiment. To address his/her request and to provide definitive and conclusive answers regarding the role of the 9TR network, we

have carried out two rescue experiments in which the corresponding endogenous mouse gene was knocked down, and at the same time in the same cells an shRNA-resistant human homologue of the corresponding knocked down TR gene was overexpressed. These data are shown in the new Figure 5C of the revised version of the manuscript. Briefly, we showed that overexpression of the human Nanog or human Irf6 genes in Nanog-KD-MEFs or Irf6-KD-MEFs respectively, rescues the reprogramming potential.

5. The item "pre-iPSC" need to be defined in this manuscript, as other papers have published quite a lot on pre-iPSC, whether the "pre-IPSC " mentioned in this manuscript is as same as in the published ones?
In the revised version of the manuscript we term these formations "**early** iPSCs colonies".

Minor concern:

1 In figure 3F, G and I, the Sox2 ChIP-seq peaks seem quite different between the iPSC and ESCs. It's truth or somehow flexible of the ChIP-seq data? How many duplicates for this experiment?
The reviewer asks a reasonable and important question: Why there are differences between the iPSCs and ESCs for the Sox2 ChIP-seq peaks at the promoters of Rcan1, Taf1c and Tfap4 genes? Our current understanding is that although iPSCs and ESCs share extensive similarities, these two types of the pluripotent cell states are not identical, thus exhibiting molecular differences and, in many instances, display clonal differences (Chin et al., 2009). In addition, we note that a common property of the three genes (Rcan1, Taf1c and Tfap4) is that they are all expressed at significant levels in naïve MEFs (prior to the initiation of cellular reprogramming). When this notion is taken into consideration with the fact that the expression of these genes is generally independent of Sox2 binding (as they are expressed at basal levels in naïve MEFs) and with the fact that our experiments have been carried out using one iPSC and one ESC clone, strongly suggest that the observed differences in the distribution of Sox2 binding are most likely clone-specific or naturally stochastic.
However, despite the above-mentioned reasonable possibilities, we were intrigued by the reviewer's point and have carried out additional analyses to investigate the genome-wide correlation of Sox2 binding between mESCs and miPSCs. The new analyses are shown in the new Figure EV5 of the revised version of the manuscript. Briefly, the correlation analysis showed that the DNA binding patterns of OSKM between iPSCs and ESCs are very similar. Sporadic differences in the DNA binding patterns identified at individual genomic locations are either due to the dynamic nature of transcription factor DNA binding and/or to the physiological differences between the two pluripotent states (ESC and iPSC) as well as to the natural clonal variability. Regarding the last point, our ChIP-seq experiments have been performed using biological replicates for mESCs and for MEFs at 18 hours of reprogramming (Pearson correlation >0.8).

Reviewer #3:

This reviewer is less enthusiastic about our manuscript and after a short introductory summary statement he/she concludes: *Although the scientific logic and the general approaches are plausible and the topic is interesting, the study is not able to substantiate the conclusions in its current form.*

We thank the reviewer for appreciating the scientific logic and the general interest of the topic under investigation in the manuscript. However, we don't agree with his/her concluding remark. Below we address his/her critique in detail.
Specific major concerns essential to be addressed to support the conclusions:

-The manuscript at hand is not the first to analyze gene expression during IPSC reprogramming and/or determine transcription factor networks in this process. Although this is not on its own an

issue, the authors should invest in comparing and discussing their data in this context (e.g. how do these TFs map onto reprogramming intermediates published previously (Lujan, O'Malley, Polo, Schiebinger etc.)).

Although it is not entirely clear to us what kind of data (transcriptomics, ChIP-seq etc) should be compared to the previous publications mentioned by the reviewer (Polo et al., 2012; O'Malley et al., 2013), we suspect that he/she requests a meta-analysis of our time course OSKM ChIP-seq data in the context of previously published transcriptomics experiments of cellular reprogramming. However, it is important to mention that although this analysis is remarkably interesting, it falls outside the scope of this manuscript. Herein, we used ChIP-seq to demonstrate that the OSKM factors target directly the genes encoding the transcription factors that reconstruct the 9TR network in order to provide additional insights for the combinatorial role of these factors in cellular reprogramming. We did not intend to study the general dynamic nature of OSKM interactions with the mouse genome and/or the consequences of these interactions in activation or repression of global transcription during reprogramming.

However, in order to satisfy the reviewer's scientific curiosity, we performed new analyses to test whether there is a general correlation between our OSKM ChIP-seq data and the transcriptomics data described in previous publications (Polo et al., 2012; O'Malley et al., 2013, Samavarchi-Tehrani et al., 2010). Importantly, in these previous publications there is a complete lack of information correlating OSKM DNA binding with transcriptional activation/repression during reprogramming.

As shown in **Figure RB7** below, we performed initial enrichment analyses of our identified OSKM binding sites in the promoter regions (+/- 5kb from TSS) of the different gene clusters as annotated in the following **publications**:
PMID: 23728301 O'Malley et al., 2013 **(Figure RB7A)**
PMID: 20621051 Samavarchi-Tehrani et al., 2010 (re-analyzed data in O'Malley et al., 2013 publication) **(Figure RB7B)**
PMID: 23260147 Polo et al., 2012 **(Figure RB7C)**

Figures RB7A, B, and C depict Dot Plots representing the correlation between our OSKM binding sites at the promoter regions of the gene clusters annotated in the publications above. The P-value is represented by the dot size and the ratio of Observed (Obs) vs Expected (Exp) by the color. The **red color** indicates correlation with OSKM enrichment at the corresponding promoter, whereas the **blue color** depicts anti-correlation (absence of OSKM enrichment). The significance of the findings was statistically evaluated by applying the two-tailed Fisher's exact test. Our conclusions are as follows:

I) Comparison of our OSKM-ChIP-seq experiments to O'Malley et al., 2013 data:

We found a strong correlation between our OSKM binding sites at the promoters of genes in clusters A, C, D and the gene expression data described in the above publication.

Cluster A (loss of expression and loss of OSKM binding during reprogramming),
Cluster C (early activation and early OSKM binding during reprogramming),
Cluster D (late activation and late OSKM binding during reprogramming)

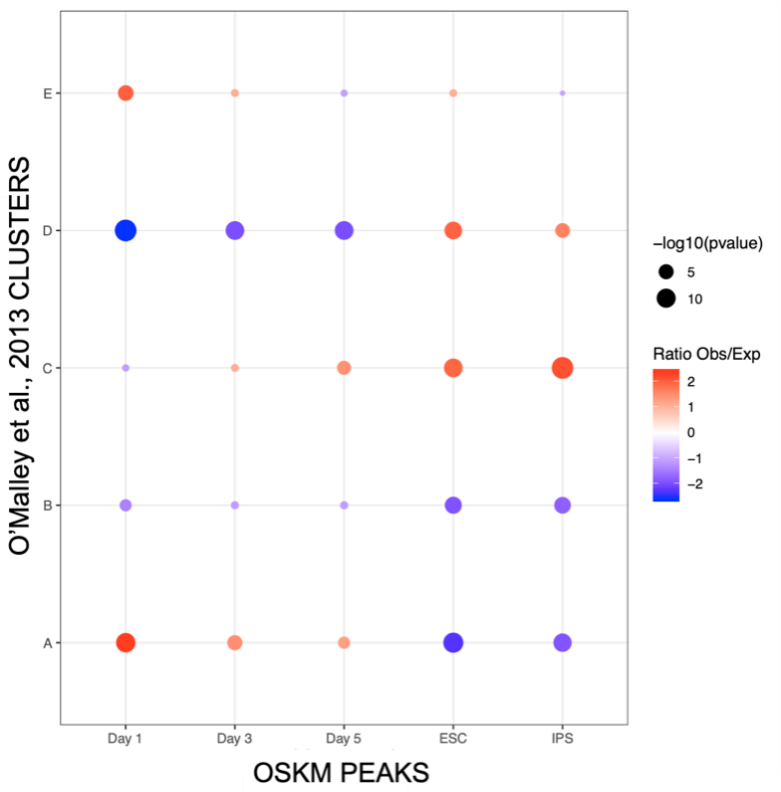


Figure RB7A: Correlation between our OSKM DNA binding patterns and the expression pattern of the gene clusters as described in O'Malley et al., 2013.

More specifically, we found that:

- There is no OSKM-binding to the genes of Cluster A during the late phases of reprogramming, a finding that correlates with the **downregulation of these genes.** Turning off these genes is a requirement for the transition from MEFs to iPSCs.
- There is significant enrichment of OSKM binding to the genes of Cluster D in ESCs/iPSCs, a result consistent with the late **activation** of these genes during reprogramming (correlates with OSKM binding at Day 3 and Day 5).

II) Comparison of our OSKM-ChIP-seq experiments to Samavarchi-Tehrani et al., 2010 data re-analyzed by O'Malley et al., 2013 data:

We found a strong enrichment of OSKM binding to the promoters of genes in the following clusters and correlation with the gene expression data described in the above publication.
Cluster tA (downregulation of genes on Day 2 and loss of OSKM binding),
Cluster tC (late activation and OSKM binding throughout reprogramming),
Cluster tD (gene activation in iPSCs and OSKM binding throughout reprogramming).

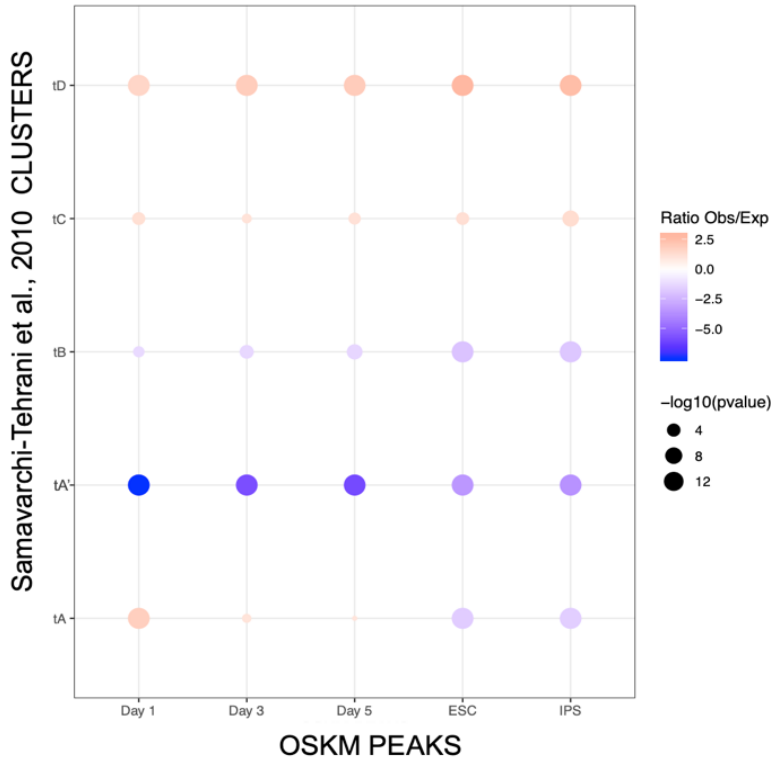


Figure RB7B: Correlation between our OSKM DNA binding patterns and the expression pattern of the gene clusters as described in Samavarchi-Tehrani et al., 2010.

III) Comparison of our OSKM-ChIP-seq experiments to Polo et al., 2012:

We found strong enrichment of OSKM binding to the promoters of the genes in the following clusters/gene expression data described in the above publication.
Cluster I (early activation and early OSKM binding)
Cluster II (early activation and early OSKM binding)
Cluster III (late activation and late OSKM binding)
Cluster V (late inhibition and loss of OSKM binding in ESCs/iPSCs)
Cluster VI (early inhibition and loss of OSKM binding from Day3 to ESCs/iPSCs)

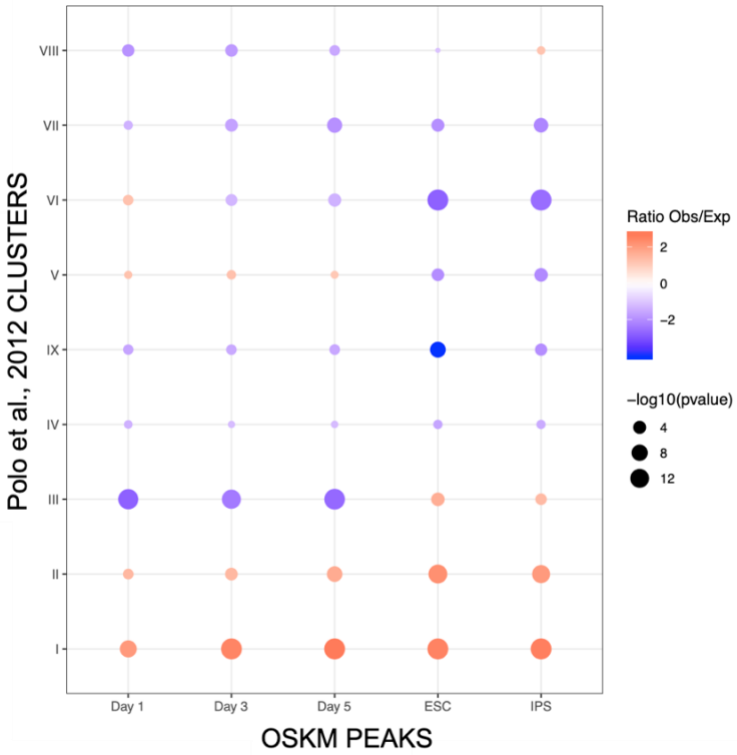


Figure RB7C: Correlation between our OSKM DNA binding patterns and the expression pattern of gene clusters as described in Polo et al., 2012

- Clusters I and II contain more than 1.000 gradually activated genes, a result consistent with the enrichment of OSKM binding throughout reprogramming (starting from Day 1 until the formation of iPSCs). Thus, the stable binding of OSKM is necessary for the **constant upregulation** of these genes (involved in cell cycle and nucleus function).
- Cluster III contains stem-cell-related genes expressed in ESCs/iPSCs only **(late activation)** consistent with our OSKM DNA binding data
- Cluster V contains 232 genes regulating "Organismal development" processes, which are **downregulated** at Day 9. The finding that OSKM bind to these genes until Day 5 of reprogramming correlates with the downregulation of Cluster V upon Day 5.

Taken together, the above comparisons strongly suggest that activation of gene expression during reprogramming correlates with OSKM binding to the corresponding target-gene(s), whereas gene repression correlates with the absence of OSKM binding in the gene clusters described in the suggested publications. As mentioned above, although there are some interesting observations regarding the patterns of OSKM DNA binding and gene activation/repression during reprogramming, these data are definitely beyond the scope of our manuscript focusing primarily on the identification of a specific gene regulatory network.

- It is unclear for several experiments if conclusions can be drawn, since effects are minor and statistical tests are missing. This is found on several locations, but exemplified best in:

"Supplementary Fig S5B shows that high-level expression of Tfap4 and Irf6 increased reprogramming efficiency". Supplementary Fig.5B suggests an improvement of reprogramming 0.3% over control which is hardly notable, the data has no error bar or standard deviation. Classic experiments, such as substitution of reprogramming TFs with any of their candidates, etc could help.

The reviewer may be confused and as a result he/she became unnecessary harsh by stating that *"It is unclear for several experiments if conclusions can be drawn, since effects are minor and statistical tests are missing".* This strong statement has no basis, because in contrast to the reviewer's conclusion, the overexpression of Tfap4 and Irf6 increased reprogramming efficiency not by 0.3% but by 29%. We have carried out this experiment several times and the results shown in the new Figure 5B of the revised manuscript clearly indicated that the simultaneous overexpression of Irf6 and Tfap4 caused a 42% increase in reprogramming efficiency. This is a significant effect given that our assay measures complex biological readouts (reprogramming efficiency) and not simple molecular effects like the levels of expression of a single gene. Regarding his/her comment about the lack of statistical tests, we have added the error bars in the figures and all relevant information is clearly described in the Materials and Method section of the revised manuscript under the heading: *Statistical Analyses*.

Regarding the substitution experiments, we have responded above (Reviewer's #2 Major concern 1) and have added a new figure (Figure EV2) in the revised version of the manuscript.

- The TR model is problematic in several ways: it is not immediately apparent what data went into it and what and whether it can predict more than that. This has several reasons:
We do not fully understand his/her point about "problems" with the 9TR network and *"whether it can predict more than that"*. The reviewer should have clarified his/her critique by specifically pointing out and criticizing both the experimental logic and the data to avoid generalizations. In the revised version of the manuscript, we rewrote parts of the relevant section to avoid confusion and improve clarity and have also included a new diagram (new Figure 4A) depicting the types of data used to reconstruct the 9TR GRN.

- *The authors seem to lack a general concept for base-line/negative expression in fibroblasts and proof of strong/functional expression in Reprogramming/IPSCs for these 9 factors. In the microarrays relative (background) expression (d0) of the different factors varies from 0 to 800 and "strong" expression from 4 to 3000 (d6).*

We feel that the reviewer ignored the fact that each of the 9TRs is expressed at varying levels in MEFs (or not expressed at all). For example, as shown in **Figure RB8**, Rcan1, Taf1c, Tead4 and Tfap4 display significant levels of basal level expression in naïve MEFs, whereas the other factors are either not expressed, or they are expressed at very low levels. The data shown in **Figure 1C** of the manuscript and in the figure below, display in comparison the relative levels of expression after normalization with the endogenous GAPDH, as determined by qPCR, and not by the Microarray experiments mentioned by the reviewer. Some genes display high and some others lower induction index during reprogramming (i.e. fold induction). High induction index is characteristic of genes with low or undetectable basal levels of expression in MEFs such as Irf6, whereas the opposite is true for genes expressed at significant levels prior to reprogramming, like the constitutively expressed Taf1c and Tead4 genes mentioned by the reviewer.

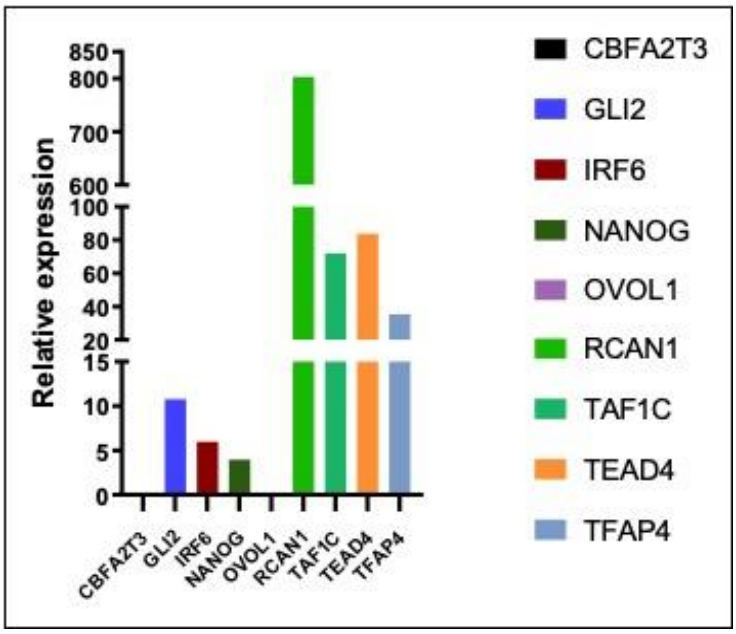


Figure RB8: Bargraph showing the basal relative expression levels of 9TRs in naïve (non-transduced) MEFs. The data were plotted upon normalization to GAPDH expression levels.

This is especially relevant,

- *because in the model and sc-qPCR they are classified digitally (as expressed/present and not expressed/absent), in SF1 as high and low expressed and in Figure 4 as % expressed. The authors have to clarify this and show that similar mRNA levels are not depicted as absent, lowly expressed or 500% induced depending on the gene.*

The reviewer requests that we should use a universal way to depict the relative and absolute levels of gene expression of the 9TRs. Although, we do not fully understand his/her point, the results of experiments using different techniques are universally presented in an optimum manner in order to be comprehensible.

because the induction is often minor (TAF1C, TEAD4)

We have addressed this point immediately above.

Because the normalization is unclear. (What does "plotted according to the median expression value" mean? Are values shown e.g. Figure 1C normalized to the expression minus Dox?

We have clarified this point by stating that the graphs depict relative levels of expression after GAPDH normalization. More specifically, we used the standard **ΔCt method** in order to calculate the relative expression levels of the target genes normalized to GAPDH expression. This was already described in the **Materials and Methods** section of the original manuscript under the heading: *cDNA synthesis and Real-time qPCR*.

Does "Signal intensity at time point T dox+/ Signal intensity at time point 0 {MEFs} >=2" & "Signal intensity at time point T dox-/ Signal intensity at time point 0 {MEFs} <2" mean you included genes induced 2.1 fold with DOX and 1.9 fold without?)

The reviewer points out that many of the genes identified as DEGs could have been the result of minor differences in their expression between + and - Dox treatment. We reanalyzed the dataset including an additional cut off of **1.5 fold change** between + and - Dox treatment at each time point (fold of gene expression of +Dox at time point X / fold of gene expression of -Dox at time point X).

Our analysis revealed that:

a) 890 out of the 4083 (21%) genes that we previously characterized as DEGs did not exceed the 1.5 fold cutoff. Nevertheless, our major findings do not seem to be affected by this additional filter.
b) The two transcriptional waves are still observed.
c) The Gene Ontology terms remain the same as originally described.
d) 30 out of 30 commonly upregulated TRs were identified.
Taken together, our original approach for the analysis of the Microarray data is valid.

- *because multiple probes corresponding to the same gene on the microarray were filtered out "and only those with the highest signal intensity across all time points were kept for further analysis". (Why did you not use the average, how would it change your data?)*

The reviewer made a valid point. We chose to carry out the analysis by filtering out all probes without signal, because these probes either map to regions that correspond to alternatively spliced, initiated or terminated transcripts, and/or contain sequences bearing many mismatches with the relevant transcript. However, despite this, we repeated the analyses of the Microarray data by using the average probe signal for each gene, exactly as proposed by the reviewer. **Figure RB9** shows that the data shown in the main manuscript **(Figure EV1)** are remarkably similar to those derived using the proposed (by the reviewer) method of analysis. These observations strongly highlight the validity and the accuracy of the transcriptomics analyses presented in the main manuscript.

The new analysis, as proposed by the reviewer, revealed 26 instead of 30 commonly upregulated TRs at Day 6 of reprogramming in mHEPs and MEFs. Of these, 24 are common with the original analysis, but 2 are new **(Figure RB10,** far right column**)**. All 9TRs, except Cbfa2t3 were also detected by using the method proposed by the reviewer. However, Cbfa2t3 is bona-fide reprogramming-inducible TR as verified by RT-qPCR experiments and Microarray assays **(Fig 1C and EV1I)**. In other words, independently of the method used for the analysis of the transcriptomics data, we detect a very similar spectrum of commonly upregulated TRs in MEFS and mHeps. Most importantly, the 9TRs that constitute our GRN are identified by both analytical approaches. Taken together, these observations further underscore the validity of our experiments.

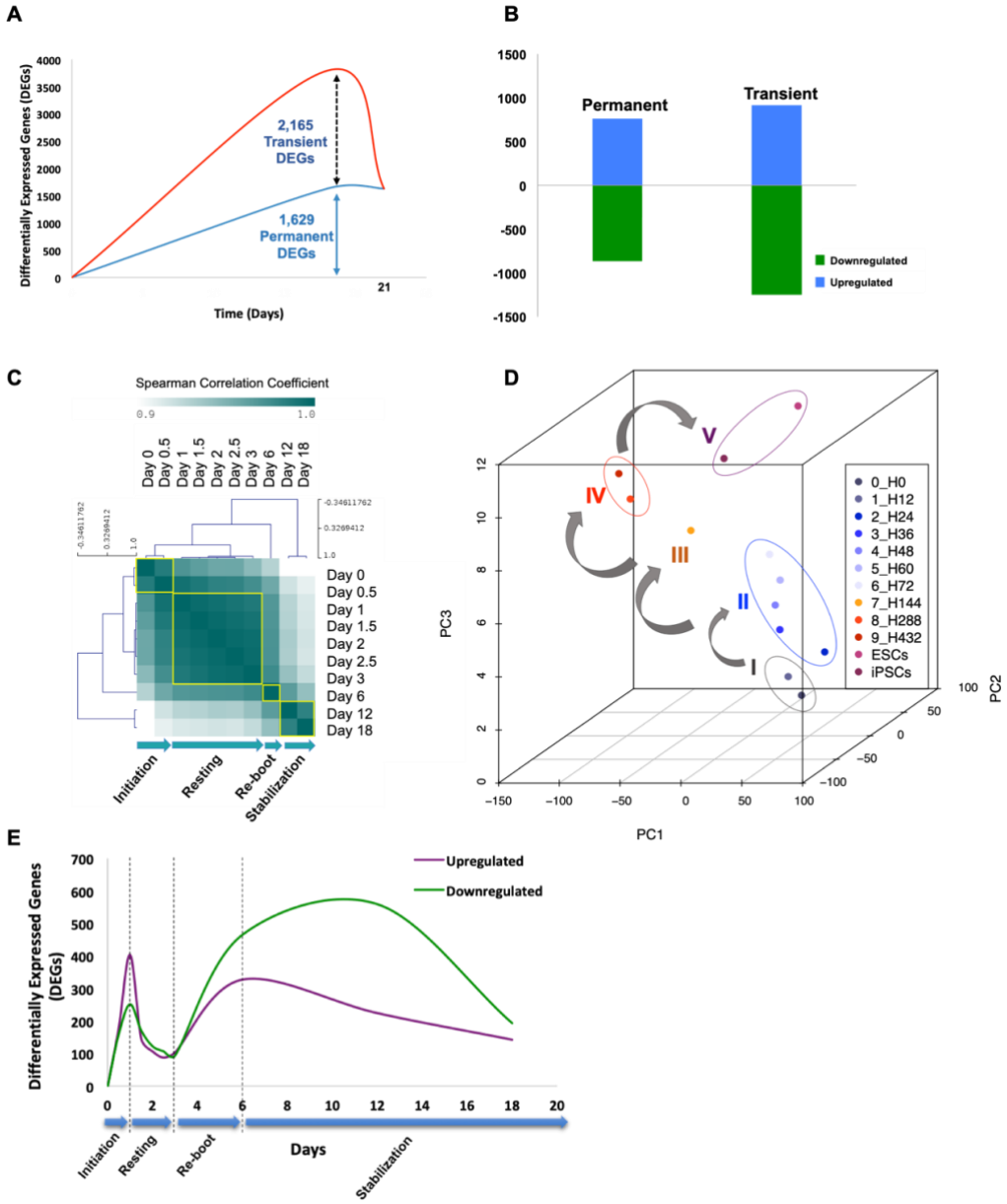
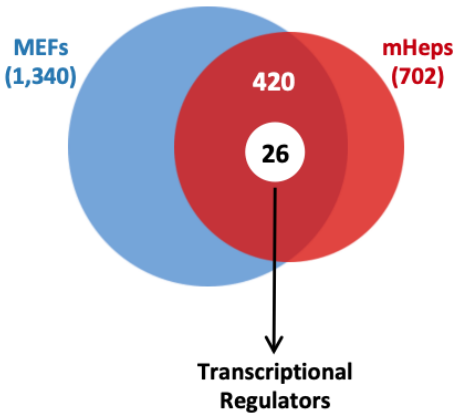


Figure RB9: A) Line plot depicting the number of DEGs identified until Day 18 of reprogramming, which either change their expression transiently or permanently at least once during reprogramming. **B)** Bar graph showing the number of up- and downregulated DEGs in transiently and permanently changed gene groups during reprogramming. **C)** Hierarchical clustering of the gene expression profiles of MEFs undergoing reprogramming from Day 0 to Day 18. Clustering revealed four distinct phases of transcriptional changes. **D)** PCA analysis of the gene expression profiles described in **(C)**. **E)** Line plot depicting the number of upregulated (purple line) and downregulated DEGs (green line) during MEF reprogramming, occurring in two separate waves.



Common upregulated TRs			
BRCA2	NPAS2	TFAP4	ID4
EHF	PHOX2A	TAF1C	MXD3
FOXD1	PYCARD	TEAD4	
FOXK1	RB1	GLI2	
HOPX	TCFL5	IRF6	
KLF3	ZBTB32	NANOG	
MEN1	ZFP423	OVOL1	
MXI1	ZIC3	RCAN1	

*Common with previous analysis (24)
*New TRs (2)

Figure RB10: Venn diagram derived from our transcriptomics analysis applying the average signal of probes depicts the total number of genes upregulated in MEFs and mHeps undergoing reprogramming. Shown in the Figure are the commonly upregulated genes (420) on Day 6 of reprogramming. Of these, 26 genes, listed in the adjacent table, encode for Transcriptional Regulators (TRs).

- *and because protein expression (western blot/ Chip) is not shown.*

In the revised version of the manuscript we have included western blots **(Fig 1C)** depicting the inducible expression of several TRs reconstructing the 9TR network, as suggested by the reviewer.

- *Including up and down-regulation to different levels in both directions between all gene pairs makes the model visually very messy and uninformative.*

We disagree with the reviewer. We strongly believe that the 9TR network model as presented provides an accurate picture of the dynamic interactions between all 9TRs reconstructing this regulatory network.

Minor concerns that should be addressed:

- What happens to the KD cells? Do they die, do they retain MEF characteristics, do alternative reprogramming cell fates become upregulated – in other words is this a specific phenotype for impaired reprogramming or some sort of general defect? Sequencing or expression profiling (other than Lin28 and EpCam) may be helpful.

The reviewer wonders what happens to the KD cells by asking a large number of questions. Most of these questions are beyond the scope of this manuscript. As we mentioned in our response to Reviewer's 1 main concern #4, naïve MEFs bearing KDs for the 4 TRs do not change their proliferation rate or show apoptotic effects **(Figure RB4** and **RB5)**. We have also carried out KD experiments for the 9TRs measuring the viability of cells at Day 6 of reprogramming **(Figure RB6)**. As we showed in **Figure RB6** above, our results highlight that the effect of 9TRs KD in reprogramming efficiency is not due to effects in cell proliferation or viability, but to the direct involvement of these genes in the reprogramming process.

- Are the phases derived from transcriptomes normalized to the –Dox control? Do you see similar phases in the transcriptome of those cells?

Yes, the data have been normalized to the -Dox control.

- How do you define transcriptional regulators? What is the source for this gene list?

TRs are annotated according to the Ingenuity Pathway Analysis software (IPA 2011 version). The accuracy of these annotations was verified by literature searching.

- Is the Reprogramming block of Pycard reproducible and significant?
This experiment has been performed twice and the results were reproducible and statistically significant according to unpaired Student's two-tailed t-test.

- Figure 2E, label is misleading
We don't agree with the reviewer on the use of the word "misleading". We have clarified the text to provide more information regarding the method used to derive the data shown.

- Why is the scaling in the ChIP changing in each panel (3-40, 2-50, 3-70 etc). Although this is overall convincing data it is suboptimal for analysis.
The Bedgraph and BigWig files generated for Figure 3 and Figure EV4 were normalized by scaling down all ChIP-seq experiments to the sample with the smallest number of reads. The selected cut-offs separate the true ChIP signal from the noise and depict the true peaks around the selected promoter regions of the indicated genes.

REFERENCES

Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegrini M, Plath K, Lowry WE (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5 (1):111-23.

Choi SS and Diehl AM (2009) Epithelial-to-mesenchymal transitions in the liver. *Hepatology* 50 (6):2007-13.

Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K (2017) Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* 168(3):442-459.e20.

Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132 (4):567-82.

Marzano C, Gandin V, Folda A, Scutari G, Bindoli A, Rigobello MP (2007) Inhibition of thioredoxin reductase by Auranofin induces apoptosis in cisplatin-resistant human ovarian cancer cells. *Free Radic Biol Med* 42(6):872-81.

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 26(1):101-6.

Nishiyama A, Sharov AA, Piao Y, Amano M, Amano T, Hoang HG, Binder BY, Tapnio R, Bassey U, Malinou JN, Correa-Cerro LS, Yu H, Xin L, Meyers E, Zalzman M, Nakatake Y, Stagg C, Sharova L, Qian Y, Dudekula D, Sheer S, Cadet JS, Hirata T, Yang HT, Goldberg I, Evans MK, Longo DL, Schlessinger D, Ko MSH (2013) Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. *Sci Rep* 3:1390.

Ng HH, Surani MA (2011) The transcriptional and signalling networks of pluripotency. *Nat Cell Biol* 13(5):490-6.

O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, Kaji K (2013) High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* 499(7456):88–91.

Park N, Chun YJ (2014) Auranofin promotes mitochondrial apoptosis by inducing annexin A5 expression and translocation in human prostate cancer cells. *J Toxicol Environ Health A*. 77(22-24):1467-76.

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alai S, Cloutier J, Bar-Nur O, Cheloufi S, Stadtfeld M, Figueroa ME, Robinton D, Natesan S, Melnick A, Zhu J, Ramaswamy S, Hochedlinger K (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151(7):1617–1632.

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL (2010) Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* (7) 1:64-77.

Vidal SE, Amlani B, Chen T, Tsirigos A, Stadtfeld M (2014) Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. *Stem Cell Reports* 3(4):574-84.

Wernig M, Meissner A, Cassady JP, Jaenisch R (2008) c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* 2 (1):10-2.

Young R (2011) Control of the embryonic stem cell state. *Cell* 144(6):940-54.

2nd Editorial Decision 13th Aug 2020

Thank you for submitting your revised manuscript for our consideration. Please apologize the delay in communicating this decision to you, which was due to a delayed referee report on account of the current pandemic situation. We now have two reports from the original referees (see comments below) and I am pleased to say that these referees overall now support publication. However, referee #3 had raised several issues regarding the way the data is presented and the statistical analyses performed. We have again looked at these points, also asking the other two referees to comment on them. Thus I would like to ask you to respond to referee #2's remaining concerns and address the following points in a final revised version of the manuscript

1) EMBO Journal policy recommends that the actual individual data from each experiment should be plotted if n < 5, alongside an error bar (please also see: https://www.embopress.org/page/journal/14602075/authorguide#statisticalanalysis). Please update accordingly, in particular for Figure 5. For this you do not need to repeat experiments, but plot the individual replicates you have in the figures.

2) The representation of figure 5B differs from 5C and it would be more clear to the reader if both were displayed as % reprogramming efficiency, while including a bar for the empty vector. If this is not possible, please briefly discuss this point. In addition, please mention in the figure legend that AP+ colonies are scored as iPSCs.

In this final revised version, I would also ask you to please address a number of editorial issues that are listed in detail below.

REFEREE REPORTS

Referee #1:

Thanos and colleagues have submitted a revised version of their manuscript entitled "Construction of a dynamic gene regulatory network required for cellular reprogramming" in which they have addressed all points that I raised as potential weaknesses of their original work. Their responses are satisfying and the changes they made to the manuscript have improved the its clarity and quality. In particular, my concerns about the identity of input cells used for reprogramming and about the efficiency of their genetic interference experiments have been fully addressed. I have no further concerns and recommend this manuscript for publication.

Referee #2:

Based on the feedback from the author, it will be suitable candidate for publication by revising the points listed below.
1 There were 30 upregulated TRs identified based on the analysis (Fig 1A) but only 28 TRs were tested (Fig 1B), will the missing two TRs (Cphx and Phox2a) function as positive , negative or without effect during reprogramming?
2 Based on the model in Fig 6, did the cells that co-expression of two or more identified TRs have higher reprogramming efficiency than the rest cells? Could the author identify this by immunofluorescence or FACS?
3 Is MET or other biological process were infected after KD TRs?

Response to reviewers

Reviewer #1

Thanos and colleagues have submitted a revised version of their manuscript entitled "Construction of a dynamic gene regulatory network required for cellular reprogramming" in which they have addressed all points that I raised as potential weaknesses of their original work. Their responses are satisfying and the changes they made to the manuscript have improved the its clarity and quality. In particular, my concerns about the identity of input cells used for reprogramming and about the efficiency of their genetic interference experiments have been fully addressed. I have no further concerns and recommend this manuscript for publication.

We thank the reviewer for his/her constructive criticism to our approaches and experiments, which altogether have significantly improved the quality and clarity of our manuscript.

Reviewer #2

Based on the feedback from the author, it will be suitable candidate for publication by revising the points listed below.

We thank the reviewer for his/her constructive and positive evaluation of our manuscript. Below we provide answers to the reasonable remaining points raised by the reviewer and have modified the text of the manuscript accordingly.

1 There were 30 upregulated TRs identified based on the analysis (Fig 1A), but only 28 TRs were tested (Fig 1B), will the missing two TRs (Cphx and Phox2a) function as positive , negative or without effect during reprogramming?

The reviewer makes a reasonable and valid point asking why Figure 1B shows the results of 28 TR-knockdown experiments, whereas originally, we identified 30 TRs (Fig. 1A) that are activated during cellular reprogramming in both MEFs and mHeps. Indeed, we did not include the KD effects of Cphx and Phox2a in Figure 1B, because we could not succeed in obtaining efficient KD despite the fact that we carried out multiple shRNA trials for both transcription factors. This information was stated in the Material and Methods section of the originally revised manuscript. To avoid confusion we have included this information also in the legend of Figure 1B.

2 Based on the model in Fig 6, did the cells that co-expression of two or more identified TRs have higher reprogramming efficiency than the rest cells? Could the author identify this by immunofluorescence or FACS?

The reviewer has raised a beautiful question: do cells co-expressing two or more of the 9TRs have a higher probability to become iPSCs? Currently, we are carrying out experiments to address this remarkably interesting point by isolating and characterizing cells that express Irf6 or Ovol1 in a stochastic manner. Specifically, we generated reporters bearing the Irf6 and Ovol1 enhancers driving the expression of GFP. When these constructs were introduced to MEFs undergoing reprogramming recapitulated the stochastic expression of the endogenous genes. We isolated GFP positive cells by FACS and demonstrated that these cells give rise to iPSCs colonies with a 2-3 fold higher efficiency than the control GFP negative cells. These experiments are in progress and will be complemented with RNA-seq and other molecular analyses. Evidently, these data are beyond the scope of this manuscript.

3 Is MET or other biological process were infected after KD TRs?

The reviewer asks a very interesting question: Does the KD of the TRs affect MET and by doing so decrease reprogramming? The reviewer is right. Figure 4E shows that indeed, the 9TRs KD

significantly reduced the expression of the epithelial marker Epcam, whose increased expression highlights MET. As a result of MET inhibition, the subsequent expression of Nanog and Lin28a is dramatically reduced as compared to control cells (Fig. 4E).

Accepted **7th Sep 2020**

Thank you again for submitting the final revised version of your manuscript. I am pleased to inform you that we have now accepted it for publication in The EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Dimitris Thanos
Journal Submitted to: EMBO Journal
Manuscript Number:EMBOJ 2019-102236

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

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

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

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

2. Captions

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

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

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

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

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

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	All of our experiments have been performed in at least two biological replicates. This information is depicted at the corresponding figure legends.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Our data have been analyzed using F-test, unpaired two-tailed student t-test, Welch's test and ANOVA. The data are presented as mean +/- SEM
Is there an estimate of variation within each group of data?	Yes, +/- SEM

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumo

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	YES

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	This information is provided in the Material and Methods section of the paper. The novel antibodies prepared for Irf6, TEAD4, and Taf1c are described at the Materials and Methods section of the paper.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	HEK293T and mESCs (Bruce4) are commercially available.

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

D- Animal Models

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

E- Human Subjects

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

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	This informations has been included
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	YES
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

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