

# **Dynamics of alternative splicing during somatic cell reprogramming reveals functions for RNA-binding proteins CPSF3, hnRNP UL1 and TIA1**

Claudia Vivori<sup>1,2</sup>, Panagiotis Papasaikas<sup>1,#</sup>, Ralph Stadhouders<sup>1,##</sup>,  
Bruno Di Stefano<sup>1,%</sup>, Clara Berenguer Balaguer<sup>1,%%</sup>, Serena Generoso<sup>1,2</sup>,  
Anna Mallol<sup>1</sup>, José Luis Sardina<sup>1,%%</sup>, Bernhard Payer<sup>1,2</sup>, Thomas Graf<sup>1,2</sup>  
and Juan Valcárcel<sup>1,2,3\*</sup>

- 1 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology,  
Carrer del Dr. Aiguader 88, 08003 Barcelona, Spain
- 2 Universitat Pompeu Fabra (UPF), Carrer del Dr. Aiguader 88, 08003 Barcelona, Spain
- 3 Institució Catalana de Recerca i Estudis Avançats (ICREA),  
Passeig Lluís Companys 23, 08010 Barcelona, Spain

\* Correspondence to [juan.valcarcel@crg.eu](mailto:juan.valcarcel@crg.eu)

# Current address: Friedrich Miescher Institute for Biomedical Research,  
Maulbeerstrasse 66 / Swiss Institute of Bioinformatics, 4058 Basel, Switzerland

## Current address: Departments of Pulmonary Medicine and Cell Biology,  
Erasmus MC, Rotterdam, The Netherlands

% Current address: Department of Molecular Biology, Massachusetts General Hospital / Center for Regenerative Medicine / Center for Cancer Research, Massachusetts General Hospital / Harvard Medical School, Boston, MA, USA / Department of Stem Cell and Regenerative Biology / Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA

%% Current address: Josep Carreras Leukaemia Research Institute, Carretera de Can Ruti,  
Camí de les Escoles, s/n, 08916 Badalona, Spain

## Abstract

In contrast to the extensively studied rewiring of epigenetic and transcriptional programs required for cell reprogramming, the dynamics of post-transcriptional changes and their associated regulatory mechanisms remain poorly understood. Here we have studied the dynamics of alternative splicing (AS) changes occurring during efficient reprogramming of mouse B cells into induced pluripotent stem (iPS) cells. These changes, generally uncoupled from transcriptional regulation, significantly overlapped with splicing programs reported during reprogramming of mouse embryonic fibroblasts (MEFs). Correlation between gene expression of potential regulators and specific clusters of AS changes enabled the identification and subsequent validation of CPSF3 and hnRNP UL1 as facilitators, and TIA1 as repressor of MEFs reprogramming. These RNA-binding proteins control partially overlapping programs of splicing regulation affecting genes involved in developmental and morphogenetic processes. Our results reveal common programs of splicing regulation during reprogramming of different cell types and identify three novel regulators of this process.

## Keywords

Alternative splicing, Somatic cell reprogramming, CPSF3, hnRNP UL1, TIA1, Pluripotency

## Introduction

Alternative splicing (AS) is a widespread mechanism of gene regulation that generates multiple mRNA isoforms from a single gene, dramatically diversifying the transcriptome (and proteome) of eukaryotic cells. 95% of multi-exonic mammalian genes undergo AS, producing mRNA isoforms which often differ

in coding capacity, stability or translational efficiency, and that can be translated into proteins with distinct structural and functional properties (Pan et al. 2008; E. T. Wang et al. 2008). AS contributes to the regulation of many biological processes in multicellular eukaryotes, including embryonic development and tissue specification (reviewed in Baralle and Giudice, 2017). During the last decade, progress has been made to understand the role of post-transcriptional regulation (including AS) in the maintenance of cellular pluripotency and cell fate decisions, discovering genes differentially spliced between stem cells and differentiated cells and splicing regulators that control these choices (G. W. Yeo et al. 2009; Salomonis et al. 2010; Gabut et al. 2011; Han et al. 2013; Venables et al. 2013; Lu et al. 2014; Yamazaki et al. 2018; Solana et al. 2016a).

During reprogramming into induced pluripotent stem (iPS) cells, somatic cells revert to a pluripotent state after overexpression of the transcription factors OCT4, SOX2, KLF4 and MYC (OSKM) (Takahashi and Yamanaka 2006). Substantial progress has been made to understand the process at the transcriptional and epigenetic level, such as by identifying numerous roadblocks and some facilitators, but comparatively little is known about how post-transcriptional regulation impacts cell fate decisions. Recent work has revealed the functional relevance and conservation of splicing regulation during reprogramming (Han et al. 2013; Ohta et al. 2013; Toh et al. 2016; Kanitz et al. 2019; reviewed in Zavolan and Kanitz, 2018). Previously, a conserved functional splicing program associated with pluripotency and repressed in differentiated cells by the RNA-binding proteins MBNL1 and MBNL2 was previously reported (Han et al. 2013). This splicing program includes a mutually exclusive exon event in the transcription factor FOXP1: a switch in inclusion of *Foxp1* exons 16/16b modulates the functions of the transcription factor between pluripotent and differentiated

cells (Gabut et al. 2011). Illustrating the complexity of such splicing program, dynamic changes of AS during cell reprogramming of mouse embryonic fibroblasts (MEFs) revealed sequential waves of exon inclusion and skipping in reprogramming intermediates and the functional role of splicing regulators in modulating reprogramming, in particular during the initial mesenchymal-to-epithelial transition (MET) phase (Cieply et al. 2016). Given the limited efficiency of cell reprogramming in this system, subpopulations of reprogramming intermediates had to be isolated through the expression of a pluripotency marker, biasing studies to the most prevalent and dominant factors.

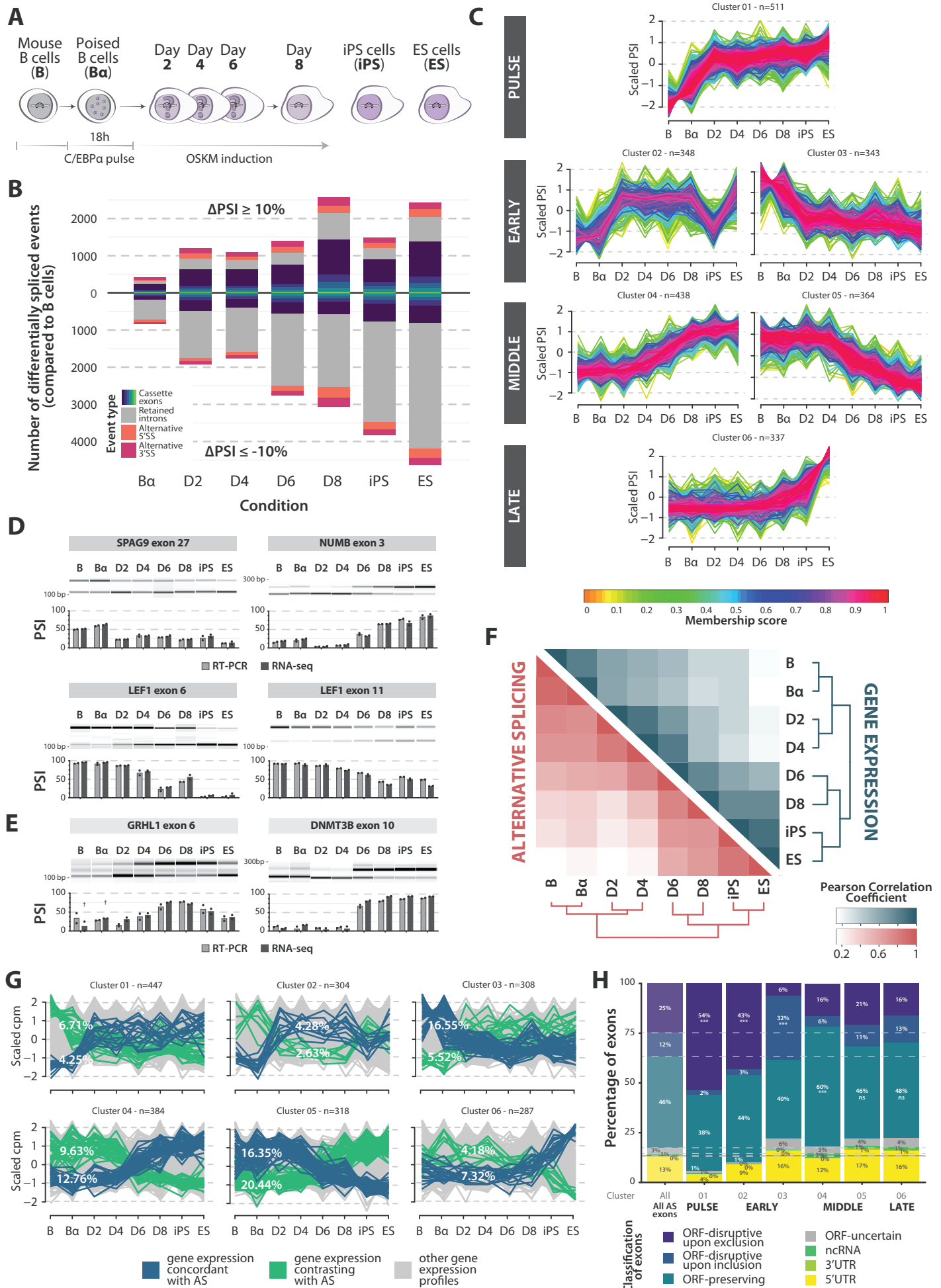
Here we took advantage of the rapid, highly efficient and largely synchronous reprogramming of pre-B cells (hereafter referred to as “B cell reprogramming”), obtained by a pulse of the transcription factor C/EBP $\alpha$  followed by induced OSKM expression (Di Stefano et al. 2014, 2016) to study the dynamics of AS during this transition. The essentially homogeneous reprogramming of the cells in this system allowed detailed temporal transcriptome analyses of the bulk population, without the need of selecting for reprogramming intermediates. We established clusters of temporal regulation and compared these changes with the ones differentially spliced in MEF reprogramming (Cieply et al. 2016). Analyzing the dynamic expression of RNA-binding proteins (RBPs) during reprogramming, we inferred potential AS regulators, of which three were studied in detail. Characterization of these factors, namely CPSF3, hnRNP UL1 and TIA1, by perturbation experiments demonstrated their role as AS regulators in the induction of pluripotency.

## Results

### **Dynamics of alternative splicing in C/EBP $\alpha$ -enhanced B cell reprogramming occur independently from gene expression changes**

To study the dynamics of changes in alternative splicing (AS) during cell reprogramming, primary mouse pre-B cells (hereafter referred to as “B cells”) were reprogrammed as previously described (Di Stefano et al. 2016; Stadhouders et al. 2018; Di Stefano et al. 2014). Briefly, B cells were isolated from bone marrow of reprogrammable mice, containing a doxycycline-inducible OSKM cassette and an OCT4-GFP reporter. These cells were infected with a retroviral construct containing an inducible version of C/EBP $\alpha$  fused to the estrogen receptor ligand-binding domain (ER). Infected cells were selected and re-plated on a feeder layer of inactivated mouse embryonic fibroblasts (MEFs). This was followed by a 18h-long pulse of  $\beta$ -estradiol, triggering the translocation of C/EBP $\alpha$ -ER to the nucleus and poisoning the B cells for efficient and homogeneous reprogramming (Di Stefano et al. 2016, 2014). After washout of  $\beta$ -estradiol, reprogramming was induced by growing the cells for 8 days in reprogramming medium containing doxycycline (see Methods and Stadhouders et al. 2018). RNA was isolated every other day from duplicates and subjected to paired-end sequencing (RNA-seq), resulting in high coverage (more than 100 million reads per sample) (**Figure 1A**). As positive controls, mouse embryonic stem (ES) cells and induced pluripotent stem (iPS) cells were also sequenced.

AS analysis was performed using *vast-tools* (Tapial et al. 2017), an event-based software that quantifies Percent Spliced-In (PSI) values of annotated AS events and constitutive exons in all samples. These analyses revealed more than 14000 AS changes during the entire reprogramming time course (for any possible



**Figure 1. Dynamics of alternative splicing and gene expression changes during B cell reprogramming** (legend on next page)

**Figure 1. Dynamics of alternative splicing and gene expression changes during B cell reprogramming.** (A) Schematic representation of C/EBP $\alpha$ -mediated B cell reprogramming time points and related controls analyzed by RNA-seq. B $\alpha$  cells: B cells treated for 18h with  $\beta$ -estradiol to activate C/EBP $\alpha$ . (B) Stacked bar plot representing cumulative number of events differentially spliced between B cells and subsequent reprogramming stages (x axis), as well as controls (iPS and ES cells). The y axis represents the number of differentially spliced events compared to B cells. The upper part corresponds to events with positive  $\Delta$ PSI values compared to B cells ( $>10\%$ ), the lower part to events with negative  $\Delta$ PSI values compared to B cells ( $<-10\%$ ). Red/orange areas: alternative 3'/5'splice sites (Alt3/Alt5) respectively; grey areas: retained introns; blue-green areas: cassette exons of increasing complexity (see Methods). See also Figure S1A. (C) Clusters of cassette exons displaying related profiles of inclusion level changes during B cell reprogramming. Six clusters (out of a total of 12 identified, see Figures S1B-C) are shown and classified into 4 categories, corresponding to the timing of the main shift observed (left). The y axis represents scaled Percent Spliced In (PSI) values. The color of each line corresponds to the *membership* score of each exon relative to the general trend of the cluster. n, The size (number of events) of each cluster is indicated (n). (D) RT-PCR validation of selected cassette exon changes inferred from RNA-seq analyses. AS events assigned to different clusters were analyzed by semi-quantitative RT-PCR and quantified by capillary electrophoresis. Each panel includes a gel representation of the inclusion (top) and skipping (bottom) amplification products of one of the replicates and the corresponding quantification of the duplicates (PSI = molarity of inclusion product / molarity of inclusion + skipping products). Light grey columns: PSI values quantified by RT-PCR; dark grey columns: PSI values quantified by RNA-seq analysis using *vast-tools* software (n=2). (E) Validation of *Grhl1* exon 6 and *Dnmt3b* exon 10 inclusion level changes, previously associated with reprogramming and pluripotency, performed as in panel D. Crosses indicate time points for which PSI values were calculated with low coverage (less than 10 actual reads). (F) Heatmap displaying correlations between B cell reprogramming stages according to gene expression (blue, top heatmap) and AS (red, bottom heatmap). Color scales represent Pearson correlation coefficient values calculated on the cpm values of the 25% most variably expressed genes or upon the PSI values of the 25% most variable cassette exons. (G) Gene expression patterns of genes containing the exons belonging to each of the AS clusters in panel C. Genes with expression changes correlating with the cluster centroid or its negative (*membership*  $> 0.3$ ) are highlighted in blue and green, respectively, while the grey portion represents the (majority of) genes which follow gene expression profiles that do not match the changes in inclusion patterns of their exon(s). Percentage of concordant/contrasting patterns are displayed for each cluster. See also Figure S1D. (H) Stacked bar plot representing the percentage of cassette exons in each of the AS clusters in panel C classified according to the following categories: disrupting the open reading frame (ORF) upon exclusion or inclusion, preserving the transcript ORF, mapping in non-coding RNAs, 3'UTRs or 5'UTRs or uncertain. The first column represents all exons differentially spliced between at least one pair of conditions, while the following ones represent the exons belonging to the each AS cluster (indicated below the bar). Where indicated, statistical significance was calculated by Fisher's exact test on the proportion of exons in the cluster compared to the general distribution of all AS exons (\*, \*\*, \*\*\* = p value  $< 0.05$ ,  $0.01$ ,  $0.001$  respectively).

pair of conditions: minimum absolute  $\Delta$ PSI of 10 between PSI averages and minimum difference of 5 between any individual replicates across conditions) and a gradual increase in the number of differentially spliced events when comparing B cells to progressive stages of somatic cell reprogramming (Figures 1B and S1A). Different classes of AS events were detected, with similar relative proportions at the various time points: 31-47% cassette exons, 5-11% alternative 3' splice sites and 6-11% alternative 5' splice sites, and 33-57% retained introns (Figure 1B).

To classify the dynamics of AS changes during reprogramming, we selected cassette exon (CE $\alpha$ ) events differentially spliced in at least one comparison (4556 exons) and performed a fuzzy c-means clustering analysis on their scaled PSI values (*Mfuzz*; Kumar and Futschik, 2007). This analysis revealed diverse kinetics of exon inclusion occurring during B

cell reprogramming (Figures 1C and S1B-C, Table S1). Six major clusters of AS dynamic profiles were detected: 1) exons that become included already after the C/EBP $\alpha$  pulse; 2) and 3) exons that are regulated either towards increased inclusion or skipping early after OSKM induction (day 2); 4) and 5) exons that display changes in inclusion at middle stages of reprogramming (day 4 onwards); 6) a cluster of exons only included at the latest steps of reprogramming (day 8 onwards) (Figure 1C). Each of these clusters consisted of 300-500 exons. The inclusion levels of examples of exons belonging to different cluster types were validated by semi-quantitative RT-PCR (Figure 1D). For reference, changes in the PSI values of *Grhl1* exon 6 and *Dnmt3B* exon 10, previously described to be associated with reprogramming and pluripotency (Gopalakrishnan et al. 2009; Gopalakrishna-Pillai and Iverson 2011; Cieply et al. 2016) were also quantified and found to follow similar inclusion patterns in B



cell reprogramming, compared to the ones previously described in other systems (**Figure 1E**).

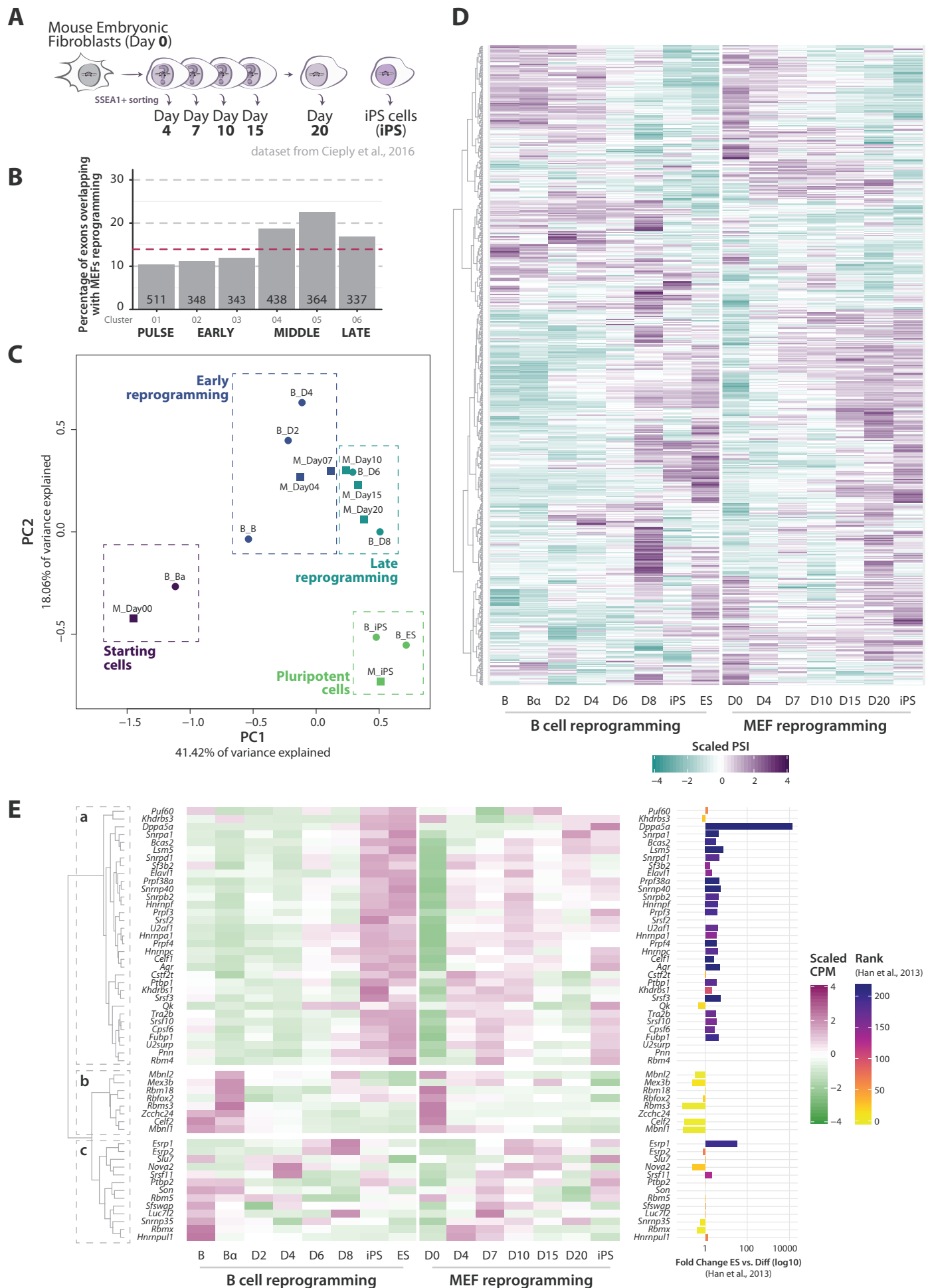
We next sought to compare general AS and gene expression dynamics during reprogramming. Gene expression was analyzed using the *edgeR* package (Robinson et al. 2009) and the level of similarity between each pair of conditions was estimated using a Pearson correlation coefficient on the cpm (counts per million) values of the most variable genes (3rd quartile coefficient of variation,  $n=2961$ ). In addition, Pearson correlation coefficient was calculated based on the PSI values from the most variable cassette exons (3rd quartile coefficient of variation,  $n=1140$ ). Both analyses showed pronounced switches between days 4 and 6 post-OSKM induction (**Figure 1F**). Overall, however, most clusters displayed matching profiles in gene expression and AS of any included exon in less than 10% of the genes, reaching a maximum of 20% in a subset of AS clusters (**Figures 1G** and **S1D-E**). These results argue, as observed before in a variety of other systems (e.g. Pan et al. 2004), that global programs of regulation of gene expression and AS are uncoupled from each other.

Interestingly, a larger proportion of exons included (or skipped) at early stages of reprogramming are predicted to disrupt the open reading frame (ORF) of the transcripts upon exon skipping (or inclusion, respectively), compared to middle/late exons and to the general distribution of mapped alternative exons (**Figure 1H**, classification as described in Tapial et al. 2017). This suggests a higher impact of AS-mediated on/off regulation of the corresponding protein expression via non-sense-mediated decay (NMD), and a switch to expression of full-length proteins during early steps of reprogramming. Middle clusters, instead, contain more exons predicted to preserve the coding potential of their transcripts, implying modulation of the functions of their encoded protein isoforms rather than on/off

switches (**Figure 1H**). Consistent with these concepts, while PSI values of cassette exons tend to increase throughout reprogramming (**Figure 1B**, blue bars), intron retention – generally leading to NMD – tends to decrease in the course of reprogramming (**Figure 1B**, grey bars).

### AS changes at intermediate reprogramming stages show commonalities with MEF reprogramming

As a first step to identify key AS events and potential regulators important for reprogramming, we compared our transcriptome analysis of B cell reprogramming with that of MEF reprogramming (Cieply *et al.* 2016). The two datasets differ in the experimental design and time frame (compare **Figures 1A** and **2A**). Specifically, the MEF system required sorting of cells undergoing reprogramming using the SSEA1 surface marker, while this was not necessary for B cell reprogramming due to its efficiency. To facilitate the comparison between the two transcriptome datasets, *vast-tools* analysis was applied to the MEF dataset (Cieply *et al.* 2016), which yielded 843 cassette exons differentially spliced ( $|\Delta\text{PSI}| \geq 10$ , range  $\geq 5$ ) between any pair of samples of the MEF reprogramming dataset. Despite differences in the experimental set up, 79% of these exons (669 out of 843) were also found to be differentially spliced in B cell reprogramming (**Figure S2A**, Table S2). A similar level of overlap was also observed at the level of other AS events (72% of AS events in general, **Figure S2A**). The overlap between differentially spliced exons in the two systems was higher for middle and late AS clusters than for early clusters of B cell reprogramming (**Figure 2B**), as might be expected from the convergence on a common program of AS related to pluripotency.



**Figure 2. B cell and MEF reprogramming systems share a program of AS changes** (legend on next page)

**Figure 2. B cell and MEF reprogramming systems share a program of AS changes.** (A) Schematic representation of MEF reprogramming time points analyzed by RNA-seq in (Cieply et al. 2016). (B) Percentage of exons in each B cell reprogramming AS cluster that are also detected as differentially spliced in the MEF reprogramming dataset of (Cieply et al. 2016). The number of events in each cluster is indicated at the bottom of the corresponding bar. The magenta dashed line indicates the average percentage for all 12 clusters. (C) PCA analysis of the 25% most variably expressed genes, segregated using *k*-means into 4 clusters: differentiated cells, early and late reprogramming and pluripotent cells, highlighted by colors and boxes. Circles: B cell reprogramming time points. Squares: MEF reprogramming time points. (D) Heatmap representing scaled PSI values (average between replicates) of exons differentially spliced in at least one time point in both B cell (left) and MEF reprogramming (right), with the corresponding hierarchical clustering. (E) Heatmap representing the expression of RNA-binding proteins known to be involved in pluripotency, somatic cell reprogramming and/or development. Scaled cpm values (average between replicates) of both datasets are shown, with the corresponding hierarchical clustering. The bar plot (right) represents (when available) the fold change expression between ES cells and differentiated mouse tissues calculated in (Han et al. 2013) and the corresponding ranking (color of the bar).

To further compare the two datasets, we performed a Principal Component Analysis (PCA) on the gene expression of the most variable genes in both datasets (3rd quartile coefficient of variation,  $n=2679$ ) separating the stages into four distinct groups by *k*-means clustering (**Figure 2C**). These groups clearly distinguish between starting cells, early and late stages of reprogramming and pluripotent cells. The PCA allowed us to outline a “reprogramming pseudotime” which was subsequently used to select AS exons and regulators for functional characterization. It also further highlighted the transition between days 4 and 6 in B cell reprogramming, juxtaposing them with days 7/10 in MEF reprogramming. A heatmap displaying the scaled PSI values of the 669 common differentially spliced exons of the two datasets at the different steps of the reprogramming process shows substantial similarities in exon inclusion (**Figure 2D**).

In contrast to the similarities in AS profiles between the two reprogramming systems, the observed that expression profiles of RNA-binding proteins (RBPs) differed significantly between the two datasets. Specifically, we analyzed the expression of RBPs previously associated with pluripotency, somatic cell reprogramming and/or development (Cieply et al. 2016; Han et al. 2013; Lu et al. 2014; Ohta et al. 2013) and presumably mechanistically linked to different aspects of post-transcriptional regulation (**Figure 2E**). Despite these more divergent profiles, hierarchical clustering captured three known functional groups, with

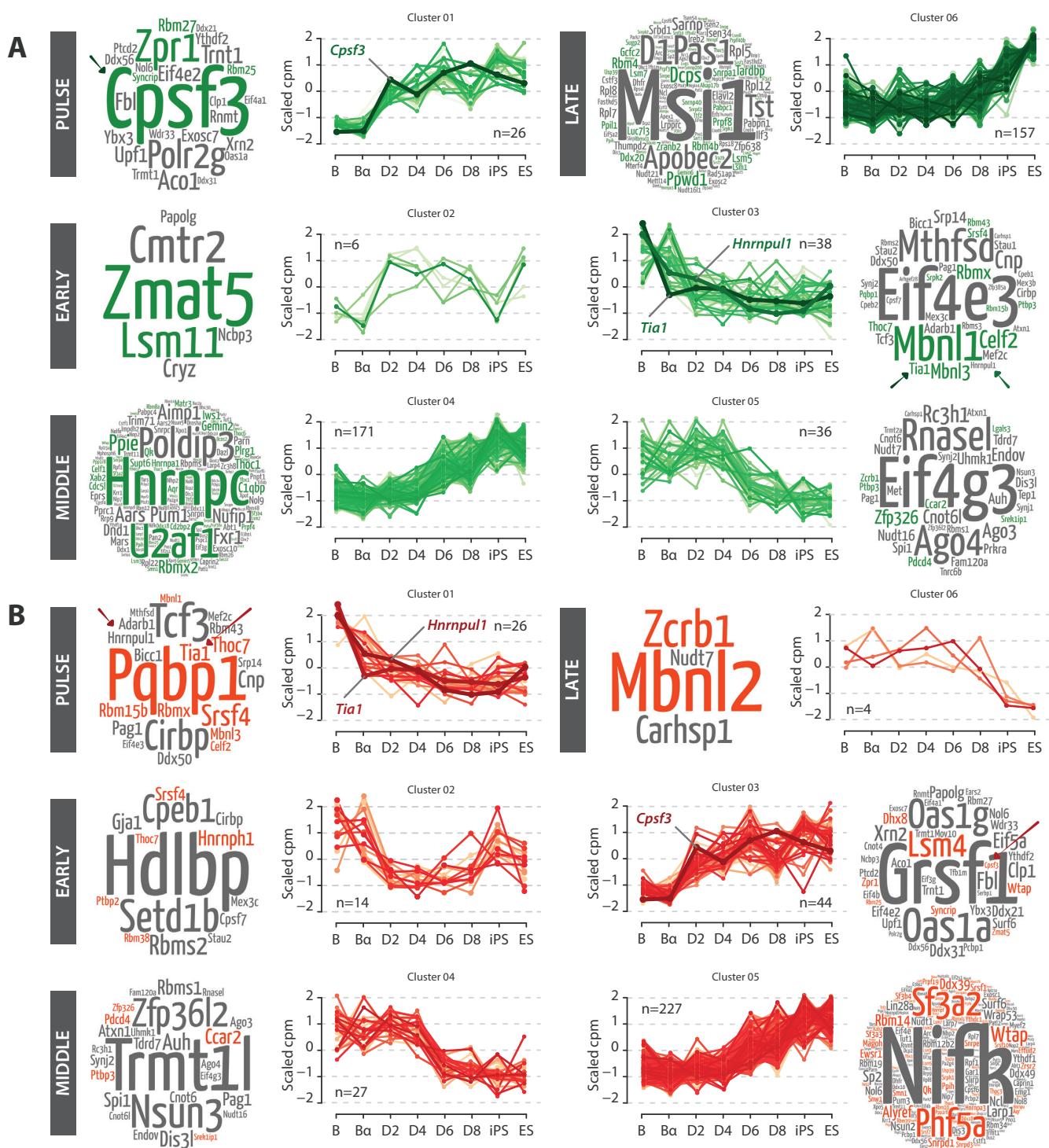
cluster *a*) containing factors with higher expression in iPS/ES samples (fold change > 0 and high rank score according to Han et al. 2013, right panel) and known to promote pluripotency/reprogramming, such as *U2af1* or *Srsf2/3* (Lu et al. 2014; Ohta et al. 2013). In contrast, cluster *b*) contains factors with higher expression in the starting somatic cells, which includes known repressors of reprogramming such as *Mbnl1/2*, *Celf2* and *Zcchc24* (Cieply et al. 2016; Han et al. 2013; Solana et al. 2016b). Finally, cluster *c*) contains factors with more variegated expression patterns at early and intermediate reprogramming steps (and mildly repressed in iPS/ES cells), including *Esrp1/2* (Cieply et al. 2016; Kanitz et al. 2019) (**Figure 2E**).

Taken together, our analyses revealed widespread AS changes during B cell reprogramming, which significantly overlapped with those of MEF reprogramming, especially at intermediate phases of the process.

### Predicted regulators of alternative splicing during somatic cell reprogramming

To infer potential regulators of exons differentially spliced during B cell reprogramming, we extracted gene expression profiles of 507 RBPs (as annotated in the *Uniprot* database), which were detectably expressed ( $\text{cpm} \geq 5$  in at least 33% of samples) and featured a minimum of variation across the B cell reprogramming dataset (coefficient of variation  $\geq 0.2$ ).





**Figure 3. Inferred regulators of AS changes during B cell reprogramming. (A)** Gene expression profiles of RNA-binding proteins (RBPs) correlating with the centroid of each AS cluster (positive regulators). Average scaled cpm values are represented by each line and the number of putative positive or negative regulators of each cluster is indicated (n). Wordclouds represent the regulators in each group, the size of the word being proportional to the correlation between gene expression and AS profiles (*membership* value). Colors indicate RBPs annotated as a splicing-related proteins. *Cpsf3*, *Hnrnpul1* and *Tia1* are highlighted in plots and wordclouds. See also Figure S3A. **(B)** Gene expression profiles of RBPs correlating with the negative of the centroid of each AS cluster (negative regulators). Displays as in panel A. See also Figure S3B.

Using the *membership* function of the *Mfuzz* package, we correlated (positively or negatively) the scaled gene expression profile of each

RBP to each AS cluster centroid. This allowed us to derive a list of potential regulators whose changes in levels of expression correlate (or

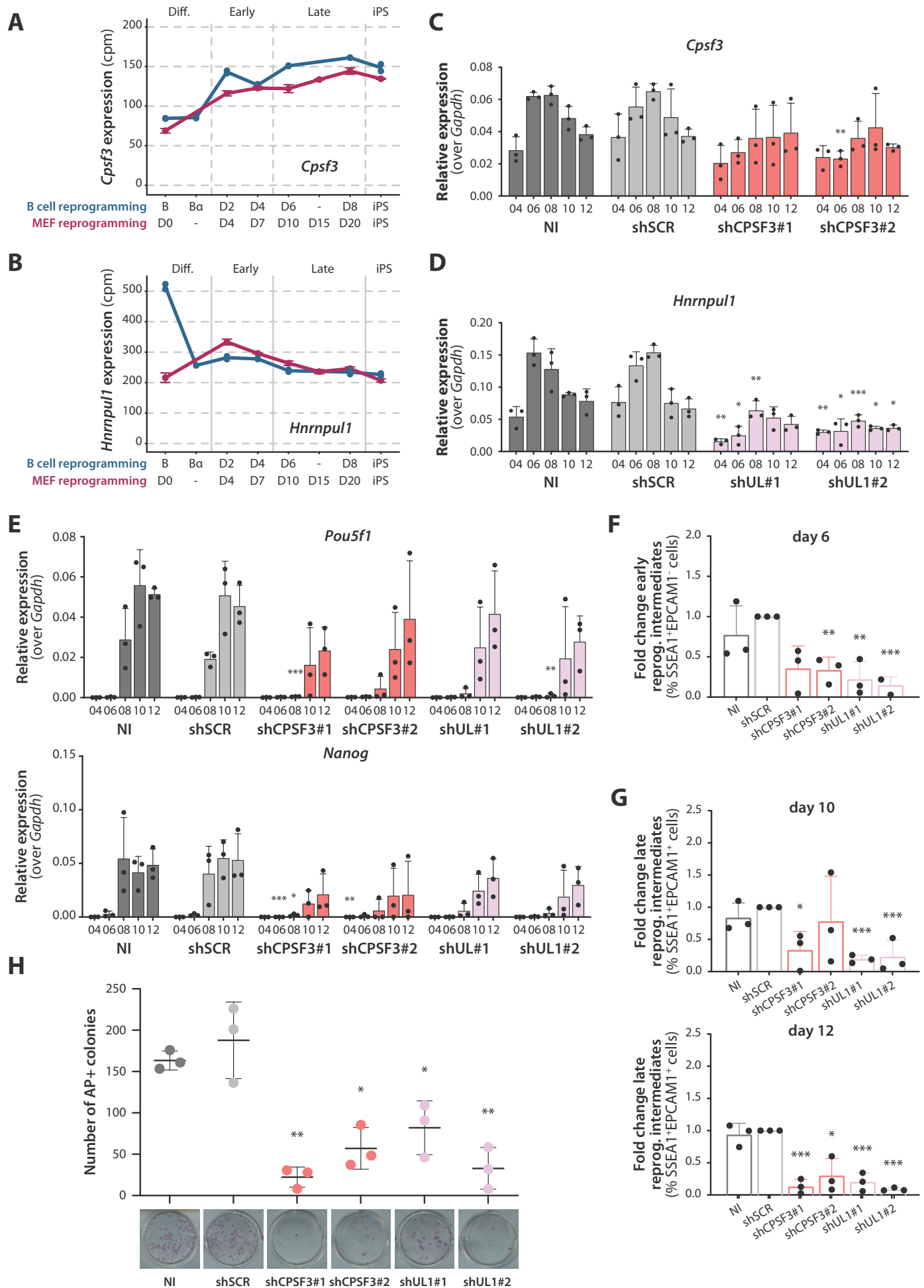
anti-correlate) with the profiles of AS changes in each cluster (membership > 0.3) (**Figures 3** and **S3**, Table S3). In line with previous work (Han et al. 2013; Ohta et al. 2013; Solana et al. 2016), our analysis identified known AS regulators involved in the induction or repression of pluripotency/developmental decisions, such as *Mbnl1/2*, *Celf2* (both potential negative regulators of pulse/late clusters 1/6) and *U2af1* (potential positive regulator of middle cluster 4) (**Figure 3**). Importantly, additional RBPs and splicing factors without previously known functions in reprogramming emerged as possible regulators.

To further define such regulators, we focused on RBPs that change their expression after the C/EBPα pulse or at early stages of reprogramming, as we speculated that these could drive the inclusion/skipping of both early and intermediate AS exons during the reprogramming of C/EBPα-poised B cells. We selected CPSF3, a top predicted positive regulator of very early events whose expression increases during B cell and MEF reprogramming, for functional validation during the induction of pluripotency (positive predicted regulator of cluster 1, **Figure 3A**). We also chose two predicted negative regulators of cluster 1, namely TIA1, a well-described AS regulator with roles in cell proliferation and development (see below), and hnRNP UL1, a member of the heterogeneous ribonucleoprotein (hnRNP) family whose involvement in splicing regulation is largely unexplored (**Figure 3B**). Due to experimental difficulties in genetic manipulation of the B cell reprogramming system (see Discussion), we modulated their expression at early stages of MEF reprogramming (Stadtfield et al. 2010) and examined the consequences on the dynamics of pluripotency induction and on relevant AS alterations.

## Knockdowns of CPSF3 or hnRNP UL1 repress MEF reprogramming

The Cleavage and Polyadenylation Specificity Factor (CPSF) complex is primarily involved in mRNA polyadenylation, but a number of its components, including CPSF3, have been shown to also play a role in splicing (Kyburz et al. 2006; Larson et al. 2016; Misra et al. 2014, 2015). *Cpsf3* expression increased early during reprogramming of both B cells and MEFs (**Figures 3A** and **4A**). The expression of *Hnrnpul1*, an hnRNP whose function in RNA metabolism is poorly understood, decreases in B cells after the C/EBPα pulse and then stabilizes at levels similar to those observed throughout MEF reprogramming (**Figures 3B** and **4B**).

To test their effects on MEF reprogramming, two different short hairpin RNAs (shRNAs) for each factor were cloned into lentiviral vectors containing a GFP reporter. The protocol used is summarized in **Figure S4A**. Briefly, early passage MEFs isolated from reprogrammable mice were transduced with constructs bearing the shRNAs and the cells treated with doxycycline to activate OSKM (day 0). GFP<sup>+</sup> cells were FACS-sorted 48 hours post-infection and seeded on inactivated MEFs serving as feeder layers, to proceed with reprogramming for up to 14 days. Cells were harvested every other day and samples analyzed by RT-qPCR for the expression of the shRNAs target and of pluripotency markers, and by flow cytometry to quantify the proportion of cells expressing the early pluripotency cell surface marker SSEA1 (appearing around day 5-6) and the later pluripotency marker EP-CAM1 (appearing around day 8, J. M. Polo et al. 2012). At day 14, two days after removing doxycycline, cultures were stained for alkaline phosphatase (AP) activity to identify iPS colonies and to assess the efficiency of reprogramming.



**Figure 4. Knockdowns of CPSF3 or hnRNP UL1 impair MEF reprogramming** (legend on next page)

**Figure 4. Knockdowns of CPSF3 or hnRNP UL1 impair MEF reprogramming.** (A) Gene expression profiles of *Cpsf3* in B cell reprogramming and MEF reprogramming (cpm values, blue and magenta lines, respectively). The x axis represents “reprogramming pseudotime” in both datasets, calculated through the PCA analysis of Figure 2B. (B) Gene expression levels of *Hnrnpul1* in B cell and MEF reprogramming (cpm values, blue and magenta lines respectively), as in panel A. (C) Relative expression levels of *Cpsf3* mRNA quantified by RT-qPCR in non-infected cells (NI), cells transduced with a scrambled control shRNA (shSCR) or one of two shRNAs specific for *Cpsf3*. The y axis represents the relative expression ( $2^{-\Delta\text{Ct}}$  value) of *Cpsf3* after normalization over *Gapdh*. (D) Relative expression levels of *Hnrnpul1* mRNA quantified by RT-qPCR as in panel C. (E) Relative expression levels of *Pou5f1* and *Nanog* quantified by RT-qPCR as in panels C and D. See also Figure S4B. (B-E) Average of biological triplicates and SD values are shown. Statistical significance was calculated by *t*-test on  $\Delta\text{Ct}$  values compared to the NI control (\*, \*\*, \*\*\* = *p* value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing using Holm-Sidak method). (F) Reduction of early reprogramming intermediates at day 6 post-OSKM induction upon knockdowns of *Cpsf3* and *Hnrnpul1*. Fold change was calculated from the percentage of SSEA1+EPCAM1<sup>+</sup> cells (of the total of alive cells) in every condition compared to the shSCR control using flow cytometry analysis. (G) Reduction of late reprogramming intermediates at day 10 and 12 post-OSKM induction upon knockdown of *Cpsf3* or *Hnrnpul1*. Fold change was calculated from the percentage of SSEA1+EPCAM1<sup>+</sup> cells (of the total of alive cells) in every condition compared to the shSCR control using flow cytometry analysis. See Figure S4C for examples of gates. (H) Number of colonies stained with alkaline phosphatase (AP) at day 14 post-OSKM induction upon knockdown of *Cpsf3* or *Hnrnpul1*. On the bottom images of representative wells are shown for every condition. (F,G,H) Average of biological triplicates and SD values are shown. Statistical significance was calculated by *t*-test comparing each condition to the shSCR control (\*, \*\*, \*\*\* = *p* value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Holm-Sidak method).

Cells infected with shRNAs targeting *Cpsf3* and *Hnrnpul1* were compared with non-infected (NI) cells, as well as with cells transduced with a scrambled control shRNA (shSCR). Knockdown efficiency, quantified by RT-qPCR (Figures 4C and D), showed a 51% and 58% reduction in *Cpsf3* at day 6 post-infection with shCPSF3#1 and #2, respectively, becoming slightly less efficient during reprogramming (Figure 4C). Similarly, knockdown of *Hnrnpul1* resulted in 81% and 76% reduction of mRNA levels at day 6 post-infection with shUL1#1 and #2, respectively (Figure 4D). The increase in the mRNA levels of endogenous *Pou5f1* (encoding OCT4) and *Nanog* pluripotency markers was delayed in both *Cpsf3* and *Hnrnpul1* knockdown cells (compare for example values at day 8 in Figures 4E and S4B), suggesting slower reprogramming kinetics. Consistent with these observations, knockdown of *Cpsf3* and *Hnrnpul1* reduced the percentage of cells expressing SSEA1 at day 6 (Figure 4F) and cells expressing both SSEA1 and EPCAM1 at days 10-12 (Figures 4G and S4C). Survival of cells during reprogramming did not seem to be affected in the knockdowns because no significant increase in the proportion of DAPI-stained cells was observed throughout reprogramming (Figure S4D). Finally, we counted the number of AP positive colonies at day 14 and found that the amount was significantly reduced in cells infected with either the

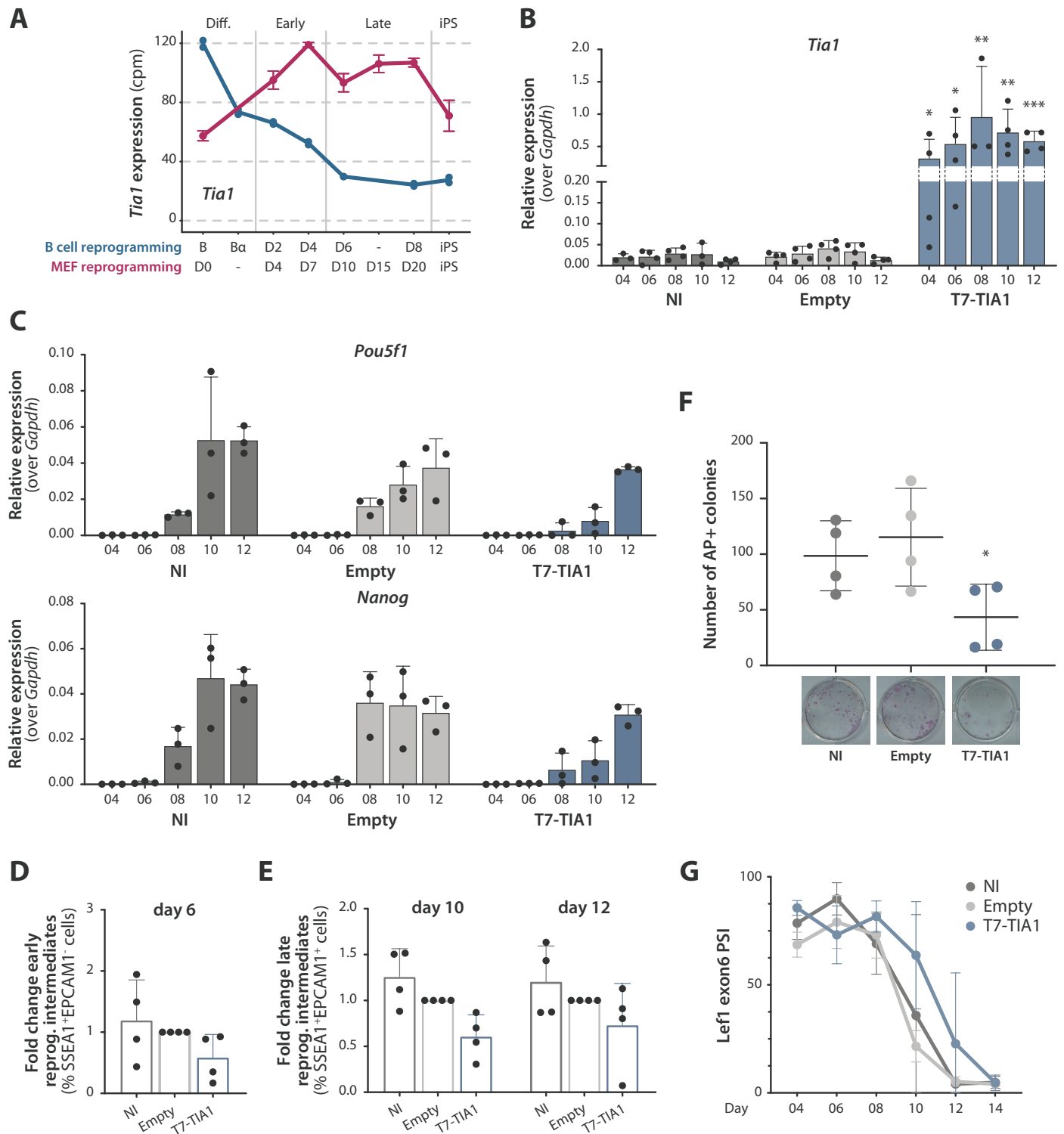
shRNAs targeting *Cpsf3* or those targeting *Hnrnpul1* (Figure 4H).

Taken together, these data show that the knockdown of *Cpsf3* or *Hnrnpul1* reduces the MEF to iPS reprogramming efficiency, therefore arguing that both RBPs contribute to the induction of pluripotency.

### Overexpression of TIA1 represses MEF reprogramming

TIA1 is an RBP and AS regulator (Förch *et al.* 2000, 2002; Izquierdo *et al.* 2005). *Tia1* mRNA levels decrease early during B cell reprogramming (Figures 3B and 5A), compatible with a potential role as a repressor of cell reprogramming in this system. Because depletion of TIA1 induces mouse embryonic lethality and the protein is important for MEF proliferation, cell cycle progression, autophagy and numerous signaling pathways (Sánchez-Jiménez and Izquierdo, 2013), we decided to test the effects of TIA1 overexpression during MEF reprogramming. For this purpose, primary MEFs were infected at day 0 (concomitantly with the induction of reprogramming), with retroviral constructs containing a T7-tagged *Tia1* cDNA and a GFP reporter to allow sorting of the transduced cells (Figure S4A). Levels of *Tia1* were quantified by RT-qPCR (24-fold increase compared to cells transduced with an empty





**Figure 5. Overexpression of TIA1 impairs MEF reprogramming** (legend on next page)

vector, **Figure 5B**). Consistent with our prediction from its expression profile in B cell reprogramming, overexpression of *Tia1* repressed the induction of endogenous *Pou5f1* and *Nanog* genes (**Figures 5C** and **S4E**) and led to a reduction of early SSEA1<sup>+</sup>EPCAM1<sup>-</sup> cells and of late reprogramming intermediates

(SSEA1<sup>+</sup>EPCAM1<sup>+</sup> cells) compared to empty vector and non-infected controls (**Figures 5D-E** and **S4F**). In addition, it significantly reduced the count of AP<sup>+</sup> colonies at day 14 post-OSKM induction compared to controls (**Figure 5F**), without substantially affecting the viability of reprogramming cells (**Figure S4G**). Of note,



**Figure 5. Overexpression of TIA1 impairs MEF reprogramming.** (A) Gene expression profiles of *Tia1* mRNA in B cell and MEF reprogramming (cpm values, blue and magenta lines respectively). The x axis represents the “reprogramming pseudotime” in both datasets, calculated through the PCA analysis of Figure 2B. (B) Expression levels of *Tia1* mRNA relative to *Gapdh*, quantified by RT-qPCR in non-infected cells (NI), cells transduced with an empty vector (Empty) or with T7-*Tia1* cDNA. (C) Expression levels of *Pou5f1* and *Nanog* relative to *Gapdh*, quantified by RT-qPCR as in panel B. See also Figure S4E. (B,C) Average of biological replicates and SD values are shown. Statistical significance was calculated by *t*-test on  $\Delta Ct$  values comparing to the Empty control (\*, \*\*, \*\*\* = *p* value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Holm-Sidak method). (D) Percentage of SSEA1+EPCAM1<sup>-</sup> early reprogramming intermediates (day 6 post-OSKM induction) upon *Tia1* overexpression determined by FACS (E) Percentage of SSEA1+EPCAM1<sup>+</sup> late reprogramming intermediates (days 10 and 12 post-OSKM induction) upon *Tia1* overexpression. See Figure S4F for gating strategy. (F) Number of alkaline phosphatase (AP) positive colonies at day 14 post-OSKM induction upon *Tia1* overexpression. Images of representative well are shown below. (D,E,F) Average of biological replicates and SD values (*n*=4). Statistical significance was calculated by *t*-test comparing to the Empty control (\*, \*\*, \*\*\* = *p* value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Holm-Sidak method). (G) Inclusion of *Lef1* exon 6 upon overexpression of T7-*Tia1*, quantified by semi-quantitative RT-PCR and capillary electrophoresis. Values represent average and SD. Statistical significance calculated by *t*-test on average  $\pm$  SD of the area under the curve in each condition yielded a *p* value of 0.079.

overexpression of *Tia1* delayed the gradual skipping of *Lef1* exon 6, a conserved AS event between B cell and MEF reprogramming (**Figure 5G**), which might partially explain the observed reduction in reprogramming efficiency.

Taken together, the observed reduced expression of TIA1 during B cell reprogramming and its impairment of MEF reprogramming when overexpressed suggest that it functions as a general repressor of pluripotency induction.

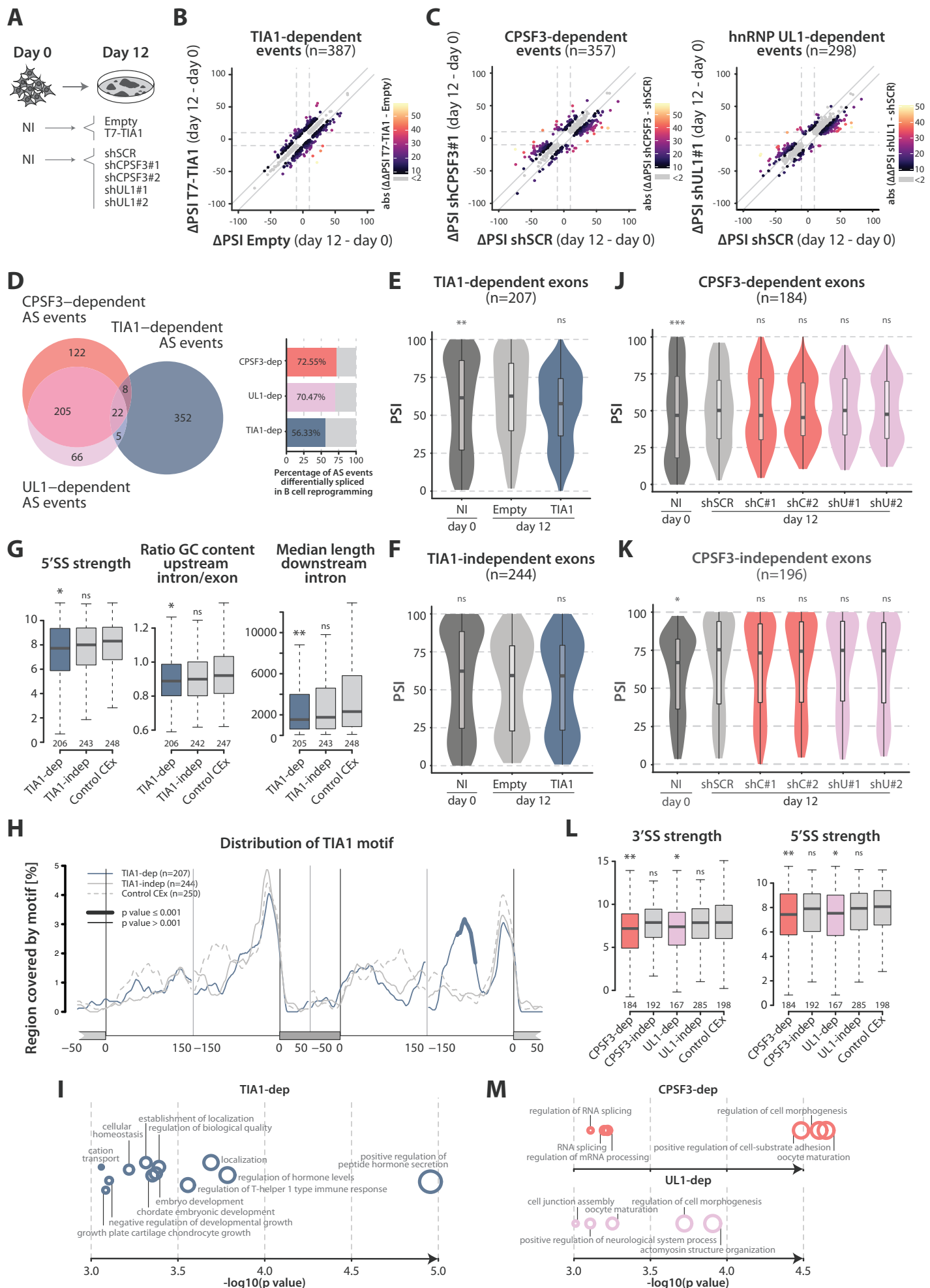
### CPSF3, hnRNP UL1 and TIA1 regulate alternative splicing during reprogramming

To assess the effects of CPSF3, hnRNP UL1 and TIA1 manipulation on AS during reprogramming, RNA-seq analyses were carried out with RNAs isolated from MEFs at 0 and at 12 days post-OSKM induction, comparing the effects of *Cpsf3/Hnnpul1* knockdown (two different shRNAs for each factor) or *Tia1* overexpression with those of the corresponding shSCR/empty vector controls (**Figure 6A**). Quantification of gene expression of *Tia1*, *Cpsf3* and *Hnnpul1* at the two timepoints is shown in **Figures S6A-B**.

To determine the impact of these perturbations on AS we calculated differentially spliced events between day 0 and day 12 in each condition as described before ( $|\Delta PSI(\text{day12-day0})| \geq 10$ , range  $\geq 5$ ).

TIA1-dependent events were defined as those changing during reprogramming only in the control or the overexpression condition, with  $|\Delta\Delta PSI(\text{TIA1-Empty})| \geq 10$  (colored dots in **Figure 6B**). Similarly, CPSF3-dependent and UL1-dependent events were defined as those with  $|\Delta\Delta PSI(\text{average shRNAs-shSCR})| \geq 10$  (only events selected by both specific shRNAs were considered, 54% and 60% overlap respectively, colored dots in **Figures 6C** and **S6C**). For each RBP, control events, changing during reprogramming but not affected by TIA1, CPSF3 or hnRNP UL1 manipulation, were classified as those differentially spliced between days 0 and 12, either in control or knockdown/overexpression conditions ( $|\Delta PSI(\text{day12-day0})| \geq 10$ , range  $\geq 5$ ), that display minimal differences between the  $\Delta PSI$  values in the two conditions (e.g.  $|\Delta\Delta PSI(\text{TIA1-Empty})| < 2$ ) (grey dots in **Figures 6B-C** and **S6C**). All sets are summarized in Table S4.

We thus identified 387 TIA1-dependent events and 558 TIA1-independent events. Similarly, we identified 357 CPSF3-dependent and 298 UL1-dependent events (as well as 429 CPSF3-independent and 662 UL1-independent events). Interestingly, CPSF3-dependent and UL1-dependent events showed a significant overlap (e.g. 70% of UL1-dependent events were also CPSF3-dependent events, of which 98% changed in the same direction) whereas little overlap was observed with TIA1-dependent events (**Figure 6D**, left panel). As we



**Figure 6. Knockdowns of CPSF3 or hnRNP UL1 and overexpression of TIA1 regulate AS during reprogramming.** (legend on next page)

**Figure 6. Knockdowns of CPSF3 or hnRNP UL1 and overexpression of TIA1 regulate AS during reprogramming.** (A) Schematic representation of the experiments performed to study the effect of TIA1 overexpression / CPSF3 or hnRNP UL1 knockdown by RNA-seq. (B) TIA1-dependent events detected during reprogramming. The x axis represents the  $\Delta$ PSI value between Empty day 12 and day 0. The y axis represents the  $\Delta$ PSI value between T7-TIA1 day 12 and day 0 control. TIA1-dependent events ( $|\Delta\Delta\text{PSI}(\text{T7-TIA1} - \text{Empty})| \geq 10$ ) are represented by colored dots (palette representing the  $|\Delta\Delta\text{PSI}(\text{T7-TIA1} - \text{Empty})|$  value,  $n = 387$ ). TIA1-independent events ( $|\Delta\Delta\text{PSI}(\text{T7-TIA1} - \text{Empty})| < 2$ ) are represented by grey dots ( $n=558$ ). (C) CPSF3- and UL1-dependent events detected during reprogramming (left and right, respectively). The x axis represents the  $\Delta$ PSI value between shSCR day 12 and day 0 control. The y axis represents the  $\Delta$ PSI value between shCPSF3#1 or shUL1#1 day 12 and day 0 control. CPSF3/UL1-dependent events ( $\Delta\Delta\text{PSI}(\text{average\_shRNAs} - \text{shSCR}) \geq 10$ ) are represented by non-grey-colored dots (palette representing the  $\Delta\Delta\text{PSI}(\text{average\_shRNAs} - \text{shSCR})$  value). CPSF3/UL1-independent events ( $\Delta\Delta\text{PSI}(\text{average\_shRNAs} - \text{shSCR}) < 2$ ) are represented by grey dots. See Figure S6C for  $\Delta$ PSI values of the same events in shRNA#2 conditions. (D) Venn Diagram representing the overlap between CPSF3-, UL1- and TIA1-dependent events (left, the number of events in each category is shown). Barplot representing the percentage of CPSF3-, UL1- and TIA1-dependent events which are also differentially spliced in B cell reprogramming (right, the percentage of overlap in each category is indicated). (E) Violin plots representing the distribution of PSI values of TIA1-dependent events in non-infected cells (NI, day 0) and day 12 cells infected with Empty or T7-Tia1 vectors. (F) Violin plots representing the distribution of PSI values of TIA1-independent events as in panel E. (E,F) Statistical significance was calculated by Fisher's exact test comparing number of events with intermediate ( $25 < \text{PSI} < 75$ ) or extreme PSI values ( $\text{PSI} \geq 75$  or  $\leq 25$ ) in each condition against Empty control (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively). See also Figure S6D. (G) Boxplots representing the distribution of the indicated sequence features of TIA1-dependent exons, compared to TIA1-independent events and a random set of exons with intermediate PSI values not changing throughout reprogramming (Control CEx). Statistical significance was calculated using *Matt*, by paired Mann-Whitney U test comparing each condition to the Control CEx set (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively). (H) RNA map representing the distribution of TIA1 binding motif in TIA1-dependent exons and flanking introns, compared to TIA1-independent and Control CEx. Thicker segments indicate regions in which enrichment of TIA1 motif is significantly different compared to Control CEx. (I) Gene Ontology (GO) terms enriched in genes containing TIA1-dependent events, compared to a background of all genes containing mapped AS events in the dataset. GO enrichment was calculated using *GOrrilla* and GO terms were summarized for visualization using *REViGO*. The x axis and the size of each bubble represent the  $-\log_{10}(\text{p value})$  of each GO term. (J) Violin plots representing the distribution of PSI values of CPSF3-dependent events in non-infected cells (NI, day 0) and day 12 cells infected with shSCR or two shRNAs specific for CPSF3 (shC#1 and shC#2) or UL1 (shU#1 and shU#2). (K) Violin plots representing the distribution of PSI values of CPSF3-independent events as in panel J. (J,K) Statistical significance was calculated by Fisher's exact test comparing number of events with intermediate ( $25 < \text{PSI} < 75$ ) or extreme PSI values ( $\text{PSI} \geq 75$  or  $\leq 25$ ) in each condition against shSCR control (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively). See also Figure S6E. (L) Boxplots representing the distribution of sequence features of CPSF3- and UL1-dependent exons, compared to the corresponding CPSF3- and UL1-independent events and Control CEx. Statistical significance was calculated using *Matt*, by paired Mann-Whitney U test comparing each condition to the Control CEx set (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively). (M) GO terms enriched in genes containing CPSF3- and UL1-dependent events, compared to a background of all genes containing mapped AS events in the dataset, performed as in panel I.

found no strong evidence for cross-regulation between CPSF3 and hnRNP UL1 factors (either in AS or in gene expression, **Figure S6B**), these data suggest that CPSF3 and hnRNP UL1 contribute to a common program of AS changes relevant for cell reprogramming. Moreover, TIA1-, CPSF3- and UL1-dependent events detected in MEFs showed high overlap with events that were also differentially spliced during B cell reprogramming, suggesting that these factors regulate AS events that generally contribute to the induction of pluripotency (**Figure 6D**, right panel).

Focusing on cassette exon events, TIA1-dependent exons (but not TIA1-independent exons) displayed significantly higher PSI values at day 12 compared to day 0 of reprogramming, as well as a relative increase in exons displaying intermediate PSI values (**Figures**

**6E-F** and **S6E**). These effects are altered by TIA1 overexpression (**Figure 6E**), suggesting that in this system, the typical role of TIA1 is to repress exon inclusion.

Analysis of sequence features associated with TIA1-regulated exons performed using *Matt* (Gohr and Irimia, 2018) revealed weaker 5' splice sites, shorter median length of the flanking downstream introns and a larger difference in GC content between the alternative exons and their flanking upstream introns (**Figure 6G**), suggestive of a strong dependence of these exons on the process of exon definition (Amit *et al.* 2012). Notably, an enrichment in putative TIA1 binding motifs was detected about 100 nucleotides upstream of the distal 3' splice site (**Figure 6H**), suggesting that binding of TIA1 to this region might repress inclusion of the alternative exon (or en-

hance pairing between the distal splice sites), leading to exon skipping.

Gene Ontology (GO) enrichment analysis (*GOrrilla*; Eden *et al.* 2009) on the set of genes containing TIA1-dependent AS events compared to a background list of all genes containing mapped AS events (n=11132) showed an enrichment in functions related to peptide/hormone secretion, T-helper cell functions and embryonic development (**Figure 6I**).

CPSF3- and UL1-dependent exons showed similar inclusion patterns: an increased proportion of intermediate PSIs was observed at day 12 in CPSF3- and UL1-dependent but not in -independent events (**Figures 6J-K** and **S6F**). CPSF3- and UL1-dependent exons also showed higher PSI values at day 12 compared to day 0, and an enrichment in exons displaying intermediate PSI values at the end of the reprogramming process (**Figures 6J-K** and **S6F**). These effects were attenuated, specifically for CPSF3- and UL1-dependent exons upon knockdown of either of these factors (**Figures 6J-K** and **S6F**), supporting the notion that their regulatory programs overlap. Both CPSF3- and UL1-dependent exons displayed weaker 3' and 5' splice sites (**Figure 6L**). GO analyses revealed that CPSF3- and UL1-dependent events belong to genes enriched in functional categories related to the regulation of cell morphogenesis, cell-substrate adhesion and cytoskeleton organization (**Figure 6M**).

Taken together, these results suggest that the AS regulators TIA1, CPSF3 and hnRNP UL1 function in cell reprogramming through their activities on genes relevant for cell fate decisions. Moreover, CPSF3 and hnRNP UL1 act through a highly overlapping program of splicing changes, while TIA1 affects a different set of AS events.

## Discussion

Somatic cell reprogramming holds great promise for fundamental research and medicine. In this study, we have characterized the dynamics of AS changes during the deterministic reprogramming of B cells into iPS cells and have identified three splicing regulators with a role in somatic cell reprogramming, as validated in perturbation experiments performed during MEF reprogramming.

The B cell reprogramming system is a two-step process, in which a pulse of the myeloid-specific transcription factor C/EBP $\alpha$  poises the cells for deterministic reprogramming by the subsequent induction of the OSKM factors, inducing early chromatin opening and epigenetic modifications that silence of the B cell-specific program and induce pluripotency (Di Stefano *et al.* 2014, 2016; Sardina *et al.* 2018; Stadhouders *et al.* 2018). While C/EBP $\alpha$  might not directly impact AS, it remains possible that the pulse of C/EBP $\alpha$  could induce changes in the expression of splicing regulatory factors (e.g. induction of CPSF3 or repression of hnRNP UL1 or TIA1) that trigger a program of post-transcriptional changes that contributes to cell reprogramming, or that C/EBP $\alpha$  could have direct effects on the activity of splicing factors.

We observed widespread changes in AS, whose frequency increased as reprogramming progressed. Clustering analysis of cassette exons revealed groups of exons varying in inclusion at very early stages of B cell reprogramming (including following the C/EBP $\alpha$  pulse), but also at middle and late phases after OSKM induction. Of note, early clusters contain a larger proportion of exons predicted to disrupt open reading frames (ORFs) of isoforms predominantly expressed in the B cells, whereas middle clusters contain more ORF-preserving exons. Together with the observation that the majority of regulated introns tend to be retained (leading to ORF disruption) at early



stages and are progressively spliced during reprogramming, these findings suggest that, in the transition from early to late stages, AS acts as a switch towards a gain in coding capacity, while at intermediate stages of reprogramming AS acts more commonly as a switch between isoforms.

The AS switch observed between day 4 and day 6 of B cell reprogramming coincides with a transcriptional switch, although the proportion of genes displaying both gene expression and AS changes is lower than expected (typically 1/3 of alternatively spliced transcripts are controlled at the level of gene expression by nonsense mediated decay, Lewis, Green and Brenner, 2003; Pan et al. 2004, 2008). The transcriptomic changes observed during this transition might be mediated at least in part by the replacement of the culture medium from serum + LIF cell to N2B27-based 2i medium at day 6. Exons changing their inclusion levels during this intermediate period displayed the highest overlap with exons differentially spliced in the reprogramming of MEFs, arguing for the general relevance of these changes in the induction of pluripotency.

Our analysis of changes in expression of RBPs during B cell reprogramming suggested potential regulators of AS, including factors already known to play a role in cell reprogramming such as the pluripotency repressors MBNL1/2 and CELF2 (Han et al. 2013; Solana et al. 2016), which are downregulated during B cell reprogramming, and positive pluripotency regulators such as U2AF1 and hnRNP H1 (Ohta et al. 2013; Yamazaki et al. 2018), which were upregulated during the process, consistent with results described for MEF reprogramming (Cieply et al. 2016). We focused on three candidate regulators of early AS changes, CPSF3, hnRNP UL1 and TIA1 because they could potentially contribute to the reprogramming advantage acquired by B cells poised by C/EBP $\alpha$  and/or to early events occurring during repro-

gramming. We found that perturbations of each of the three factors impaired the reprogramming of MEFs. CPSF3 plays an important role in mRNA polyadenylation, recognizing the polyadenylation signal (PAS) together with the rest of the CPSF complex. In addition to the well-known intense crosstalk between last intron splicing and 3' end processing (Kyburz et al. 2006), CPSF2 and the CPSF complex have been shown to influence splicing of internal exons independently of cleavage and polyadenylation (Misra et al. 2014, 2015). Furthermore, mutations of the CPSF3 yeast homolog Brr5/Ysh1 have been shown to strongly affect splicing, suggesting that CPSF3 could also play similar direct regulatory functions (Larson et al. 2016). hnRNP UL1, instead, is a largely uncharacterized RBP known to participate in DNA damage response (Polo et al. 2012b) and identified as a surface marker of human ES cells, although its functions in this context remain unknown (Choi et al. 2011). Recent work in Zebrafish suggest a possible role for Hnrnpul1 in the regulation of AS (Blackwell et al. 2020).

Knockdown of Cpsf3 or Hnrnpul1 (with two independent shRNAs for each gene) caused a general repression or delay in MEF reprogramming, revealing their functions as promoters of somatic cell reprogramming. Sequence analysis of regulated AS events suggest that CPSF3 and hnRNP UL1 favor the definition of exons harboring relatively weak splice sites. Remarkably, the AS targets of CPSF3 and hnRNP UL1 identified by our RNA-seq analyses show a high overlap (without evidence of cross-regulation between the two proteins), suggesting a concerted mechanism of splice site selection, perhaps as components of the same complex. It is unlikely that the overlap is an indirect consequence of reduced/delayed reprogramming, because TIA1 overexpression also inhibits reprogramming and yet it is accompanied by largely non-overlapping changes in AS. These observations sug-



gest that multiple, separable AS programs contribute to the regulation of cell reprogramming.

TIA1 is a regulator of RNA metabolism influencing mRNA decay and AS implicated in cell proliferation, cell cycle progression, cellular stress, autophagy and programmed cell death (Förch et al. 2000; Izquierdo et al. 2005; Kedersha et al. 1999; Sánchez-Jiménez & Izquierdo 2013) whose depletion induces mouse embryonic lethality (Piecyk et al. 2000).

We found that TIA1 overexpression at early stages of reprogramming represses the induction of pluripotency, as reported for MBNL proteins (Han et al. 2013). Binding of TIA1 to U-rich sequences downstream of weak 5'SS helps to recruit U1 snRNP and facilitates exon definition (Förch et al. 2000, 2002; Izquierdo et al. 2005; Wang et al. 2014), while it can also inhibit the inclusion of alternative exons located at a distance from its binding site (Ule et al. 2010). As expected, TIA1-dependent exons are enriched in weak 5' splice sites and other features that support a role in exon definition. Interestingly, TIA1-dependent exons in our system feature an enrichment of predicted U-rich TIA1 binding sites approximately 100 nucleotides upstream of the 3' splice site located in the downstream intron, suggesting that this configuration serves to repress the inclusion of exons that contribute to cell reprogramming. Since GO terms most enriched in genes bearing TIA1-dependent exons include 'positive regulation of peptide hormone secretion', we hypothesize that TIA1-regulated transcripts could modulate functions related to the senescence-associated secretory phenotype (SASP), which was shown to promote reprogramming of somatic cells and dedifferentiation in cancer (Brady et al. 2013; Mosteiro et al. 2016; Ritschka et al. 2017). Indeed, 7 out of 23 (30%) of the genes belonging to these GO terms were also found in a list of senescence- or SASP-associated genes (GeneCards database; Stelzer et al. 2016).

The finding that overexpression of TIA1 caused a delay in the conserved gradual skipping of Lef1 exon 6 observed during MEF reprogramming is consistent with a conserved role of the factor in reprogramming, as gradual skipping of this exon was also observed in middle stages of B cell reprogramming, along with a decrease in expression of the gene. LEF1 is a transcription factor with important functions in embryonic and T cell development and activation (Oosterwegel et al. 1993; Travis et al. 1991; Van Genderen et al. 1994; Waterman et al. 1991; Zhou et al. 1995). As Lef1 exon 6 inclusion is promoted by CELF2 during T cell development and activation (Ajith et al. 2016; Mallory et al. 2011, 2015) and Celf2 expression rapidly decreases in both B cell and MEF reprogramming, CELF2 may participate – along with TIA1 – in Lef1 exon 6 regulation during reprogramming. We observed that overexpression of Lef1 at the start of MEF reprogramming improved reprogramming efficiency (observed as an increase of pluripotency markers and percentage of reprogramming intermediates), with overexpression of the iPS-associated Lef1 exon 6-skipping isoform displaying stronger effects than the exon 6-including isoform, suggesting a functional role for this switch in cell reprogramming (Figure S5). We speculate that the gradual skipping of exon 6 affects the interaction of LEF1 with protein partners that either act as coactivators – e.g. ALY (Bruhn et al. 1997) – or corepressors – e.g. Groucho/TLE (Levanon et al. 1998) –, differentially affecting the transcription of target genes. For example, the LEF1-inclusion isoform could promote cell proliferation and maintenance of B cell identity – as observed in the C/EBP $\alpha$ -dependent transdifferentiation system and in pancreatic cancer cell lines (Jesse et al. 2010; Rodriguez-Ubreva et al. 2014) –, while gradual skipping of exon 6 might alter its effect on its target genes with the consequence of inducing pluripotency, most probably in a  $\beta$ -catenin independent way.

In summary, our results provide a comparison between AS changes occurring in two very different reprogramming systems and a wealth of information relevant for cell fate decisions and transitions to pluripotency. Furthermore, they demonstrate the functional involvement of the splicing regulators CPSF3, hnRNP UL1 and TIA1 in efficient cell reprogramming. It thus significantly extends previous work on splicing during reprogramming and suggests that AS is required for the entire transition of somatic into pluripotent stem cells. It would be interesting to determine whether the AS changes and regulators identified in our work play similar roles during the specification of pluripotent stem cells in early embryo development.

## Materials and Methods

### RNA sequencing

For RNA sequencing of B cell reprogramming and MEF reprogramming upon TIA1 overexpression and knockdown of CPSF3 and hnRNP UL1, stranded libraries were prepared from samples and sequenced on an Illumina HiSeq 2500 v4 using a 2x125nt paired-end protocol. Duplicates were sequenced for each condition, with samples pooled in separate lanes. RNA-seq data (triplicates of 2x100nt paired-end sequencing) of MEF reprogramming was downloaded from GEO Database (Cieply *et al.* 2016).

### Data and code accessibility

All scripts and RNA-seq data of MEF reprogramming upon TIA1 overexpression and knockdown of CPSF3 and hnRNP UL1 will be released to public databases when the article is accepted for publication.

### Gene expression and alternative splicing analyses

Reads mapping was performed with STAR (Dobin *et al.* 2013) and gene expression analysis was performed using the *edgeR* package v3.16.5 (McCarthy *et al.* 2012; Robinson *et al.* 2009). A minimum of 5 counts per million (cpm) was required in 33% of both datasets (5 samples for B cell reprogramming, 6 samples for MEF reprogramming) and a minimum coefficient of variation of 0.2 was required.

Alternative splicing (AS) analysis was performed using vast-tools software v2.2.2 (Irimia *et al.* 2014; Tapial *et al.* 2017). Strand-specific mapping was performed and only events with a minimum of 10 actual reads per sample were considered (VLOW quality score). PSI values for single replicates were quantified for all types of alternative events, including simple and complex cassette exons (S, C1, C2, C3), microexons (MIC), alternative 5' and 3' splice sites (Alt5, Alt3) and retained introns (IR-S, IR-C). For cassette exon events (referred to as CEx), PSI values of all annotated exons were also quantified with vast-tools (Annotation module ANN). For both B cell and MEF reprogramming, all possible pairwise comparisons between samples were performed, selecting differentially spliced events with a  $|\Delta\text{PSI}| \geq 10$  and a minimum range of 5 between samples.

For TIA1 overexpression experiments, TIA1-dependent events were defined as events with  $|\Delta\Delta\text{PSI}| \geq 10$ , where  $\Delta\Delta\text{PSI} = (\Delta\text{PSI}_{\text{TIA1\_day12}} - \text{NI\_day0}) - (\Delta\text{PSI}_{\text{Empty\_day12}} - \text{NI\_day0})$ . TIA1-independent events were instead defined as events changing in any of the two conditions ( $|\Delta\text{PSI}|_{\text{TIA1\_day12}} - \text{NI\_day0} \geq 10$ ) or ( $|\Delta\text{PSI}|_{\text{Empty\_day12}} - \text{NI\_day0} \geq 10$ ) and having minimal difference between the two conditions ( $|\Delta\Delta\text{PSI}| < 2$ ).

Similarly, CPSF3- or UL1-dependent events were events for which in both shRNAs  $|\Delta\Delta\text{PSI}| \geq 10$ , where  $\Delta\Delta\text{PSI} = (\Delta\text{PSI}_{\text{shC/U\#1/2\_day12}} - \text{NI\_day0}) - (\Delta\text{PSI}_{\text{shSCR\_day12}} - \text{NI\_day0})$ .

CPSF3- or UL1-independent events were instead defined as events changing in any of the knockdown or control conditions ( $|\Delta\text{PSI}|_{\text{shC/U\#1/2\_day12 - NI\_day0}} \geq 10$ ) or ( $|\Delta\text{PSI}|_{\text{shSCR\_day12 - NI\_day0}} \geq 10$ ) and having minimal difference between the two conditions ( $|\Delta\Delta\text{PSI}| < 2$ ).

### Clustering analysis of alternatively spliced exons and correlation of gene expression profiles

PSI values of exons differentially spliced in at least one pair of B cell reprogramming samples ( $n=4556$ ) were scaled and centered (referred to as 'scaled PSI' values). Fuzzy *c*-means clustering analysis was carried out using the R package *Mfuzz* (Futschik & Carlisle 2005; Kumar & Futschik 2007) on the scaled PSI values. The optimal number of clusters (12) was selected on the basis of the minimum distance to the cluster centroid. A *membership* value was assigned to correlate gene expression profiles to each AS clusters, either of genes containing the AS exons of each cluster or of genes encoding RBPs. To each scaled cpm values vector, we attributed a *membership* value to the centroid of each cluster or to its negative.

### Correlation of B cell reprogramming stages

Pearson correlation coefficient for B cell reprogramming stages was calculated on the most variable expressed genes ( $\text{cpm} \geq 5$  in at least 5 samples and coefficient of variation  $\geq 0.73864$ , corresponding to the 3<sup>rd</sup> quartile,  $n=2961$ ) or the most variable exons (differentially spliced in at least one pairwise comparison and coefficient of variation  $\geq 0.4107$ , corresponding to the 3<sup>rd</sup> quartile,  $n=1139$ ).

### Prediction of the protein impact of alternative exons

Alternative exons detected by vast-tools were classified as described in vastDB v2.2.2

(Tapial et al. 2017). Briefly, following division according to their location in non-coding RNAs, untranslated regions (5' or 3'UTRs) or open reading frame (ORF), exons were predicted to disrupt the ORF if their inclusion or skipping would induce a frameshift in their ORF or if they would induce a premature stop codon (PTC) predicted to be targeted by non-sense-mediated decay (NMD) or truncating the protein by more than 300 amino acids. The rest of the ORF-mapping events are predicted to preserve the transcript coding-potential.

### Principal Component Analysis (PCA) and comparison between B cell and MEF reprogramming

Gene expression values were filtered for minimum variation (coefficient of variation  $\geq 0.66$ , corresponding to the 3<sup>rd</sup> quartile,  $n=2679$ ), scaled and centered (referred to as 'scaled cpm' values). Principal Component Analysis (PCA) was performed on scaled cpm values of the most variable genes expressed in both B cell and MEF reprogramming. The groups shown were obtained by *k*-means clustering. For heatmaps representing scaled PSI and scaled cpm values in B cell and MEF reprogramming, hierarchical clustering was performed on values of both datasets using a Ward's method and Euclidean distance as the distance metric.

### Correlation of RNA-binding proteins expression to AS clusters

The lists of *Mus musculus* RNA-binding proteins (RBPs) and splicing-related proteins were downloaded from the *Uniprot* database. For RBPs, Uniprot keywords had to match 'RNA-binding [KW-0694]', 'mRNA splicing [KW-0508]', 'mRNA processing [KW-0507]' or 'Spliceosome [KW-0747]'. Splicing-related RBPs were defined if keywords were matching either 'mRNA splicing [KW-0508]' or 'Spliceosome [KW-0747]'. Filtered gene expression values were scaled as described above and a *membership* value was

attributed to each RBP profiles for each AS cluster centroid and to its negative (RBPs with membership  $\geq 0.3$  are shown). Word clouds were created with *WordArt.com* online tool, with the size of each RBP name proportional to the corresponding *membership* value and colors highlighting splicing-associated RBPs.

## Sequence feature analysis and RNA maps

Sequence features of exons and flanking introns were analyzed using *Matt* software v1.3.0 (Gohr & Irimia 2019). Features of TIA1-, CPSF3- or UL1-dependent exons were compared with the corresponding independent exons and with a set of control Cassette exons, representing alternative exons not changing during reprogramming. Specifically, the union of alternative non-changing exons (AS\_NC sets,  $10 < \text{average PSI} < 90$  and  $\Delta\text{PSI} \leq 2$ ) in all conditions of each dataset was generated, TIA1- or CPSF3/UL1-dependent events were excluded and a random set was selected, with a size similar to TIA1- or CPSF3-/UL1-dependent and -independent exon sets. Maximum entropy score is calculated as an approximation of splice site strength (Yeo & Burge 2004). RNA maps for TIA1 M075 motif (Ray *et al.* 2013) were generated for the first and last 50 nt of exons and the first and last 150nt of introns (sliding window = 31, p value  $\leq 0.05$  with 1000 permutations).

## Statistical analyses and plots

Statistical tests were performed as indicated in figures legends with R (v3.6.1) or GraphPad Prism (version 8). Heatmaps were plotted with *ggplot2* package or *heatmap.2* function. Venn diagrams were generated with Biovenn (Hulsen *et al.* 2008).

## Gene Ontology enrichment analysis

Genes bearing the differentially spliced exons belonging to each set were analyzed for

GO terms enrichment in contrast to all genes containing any mapped AS event. Analysis was carried out with the 'two unranked lists of genes' module of *GOzilla* (Gene Ontology enrichment analysis and visualization) tool (Eden *et al.* 2009) and summarized using *REVIGO* (Supek *et al.* 2011). Statistical significance was defined with p value  $< 1e-03$ . Only enriched terms for GO Process are shown.

## Cell culture

Platinum E cells (PlatE), 293T/17 cells, stromal S17 cells and primary MEFs serving as feeder cells for B cell reprogramming were kindly provided by the group of T. Graf and cultured in Glutamax Dulbecco's modified Eagle's (DMEM, Life Technologies) medium supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin antibiotics (Pen-Strep, 50 U/ml penicillin; 50 µg/ml streptomycin). All cells were maintained at 37°C under 5% CO<sub>2</sub> and routinely tested for mycoplasma.

## Lentivirus production

293T/17 cells were seeded on gelatin-coated 10 cm plates at a density of  $3 \times 10^6$  cells/plate and incubated overnight. The following day, medium was replaced approximately 2 hours before transfection. For each 10 cm plate, 10 µg of the plasmid of interest, 2.5 µg of VSV-G and 7.5 µg of pΔ8.9 were mixed with 61 µl of CaCl<sub>2</sub> 2.5 M (Sigma) and endotoxin-free water up to 500 µl. While bubbling, 500 µl of 2x HBS pH7.2 (281 mM NaCl, 100 mM HEPES, 1.5 mM Na<sub>2</sub>HPO<sub>4</sub>) were added dropwise, and the mix was incubated for 10 min at room temperature (RT). The solution was added dropwise to the 293T/17 cells, followed by incubation for 24 hours at 37°C. The following day, medium was substituted with 6 ml of fresh medium (reprogramming medium for MEF reprogramming). Viral medium was collected 48 and 72 hours after transfection, filtered (0.22 µm pore size) and supplemented as needed.



## Retrovirus production

4.5x10<sup>6</sup> PlatE cells were seeded in gelatin-coated 10cm plates (0.1% gelatin, Millipore) the day before transfection. To improve efficiency, chloroquine was added approximately one hour before transfection to a final concentration of 30  $\mu$ M. For each plate, 20  $\mu$ g of plasmid were mixed with 60  $\mu$ l of CaCl<sub>2</sub> 2.5 M (Sigma) and endotoxin-free water up to 500  $\mu$ l. While bubbling, 500  $\mu$ l of 2x HBS pH 7.2 (281 mM NaCl, 100 mM HEPES, 1.5 mM Na<sub>2</sub>HPO<sub>4</sub>) were added dropwise, and the mix was incubated for 10 min at room temperature (RT). The solution was added dropwise to the PlatE cells and transfected cells were incubated for 8-10 hours. Medium was substituted with 6 ml of fresh medium (reprogramming medium for MEF reprogramming). Viral medium was collected the day after transfection and the following one, filtered (0.22  $\mu$ m filter pore size) and supplemented as needed.

## C/EBP $\alpha$ -dependent B cell reprogramming

Primary B cell reprogramming was performed as described in (Di Stefano *et al.* 2014, 2016; Di Stefano & Graf 2016; Stadhouders *et al.* 2018).

## MEF reprogramming

Primary MEFs (P0), obtained from male embryos of a *Collagen-OKSM, M2rtta*<sup>+</sup> mouse line (Stadtfeld *et al.* 2010) were cultured on gelatin-coated plates in MEF medium (DMEM, high glucose + Glutamax, FBS 10%, PenStrep 1x, Sodium Pyruvate 1 mM, HEPES 30 mM, NEAA 1x,  $\beta$ -mercaptoethanol 0.1 mM). Primary MEFs were expanded for a maximum of 2 passages before inducing reprogramming. Early passage MEFs were plated in MEF medium on gelatin-coated 6-well plates at a density of 70.000 cells/well. The following day, infection was performed by substituting the medium with the filtered retroviral or lentiviral super-

natant (prepared in reprogramming medium), supplemented with LIF and doxycycline to induce reprogramming. Two subsequent infections were performed, 12 hours apart, after which fresh reprogramming medium was added. The following day (day 2) cells were trypsinized and FACS-sorted for GFP expression by flow cytometry. Sorted MEFs were seeded (10.000 cells/well of 12-well plate) on irradiated feeders on gelatin-coated plates (plated the previous day 100.000 cells/well of 12-well plate). Medium was substituted every 2 days starting from day 4. Harvesting was performed every two days with trypsin 0.25% to ensure the complete dissociation of the feeder layer. Doxycycline was withdrawn from the culture at day 12 post-OSKM induction and AP staining was performed at day 14 with the Alkaline Phosphatase Kit II (StemGent), according to the manufacturer's instructions. The number of AP<sup>+</sup> colonies was quantified with ImageJ software (colony size between 20-2000 pixels).

Reprogramming medium is composed by DMEM, high glucose + Glutamax, ES-qualified FBS 15%, PenStrep 0.5x, Sodium Pyruvate 1 mM, HEPES 30 mM, NEAA 1x,  $\beta$ -mercaptoethanol 0.1 mM. LIF (1000 U/ml) and doxycycline (1 $\mu$ g/ml) were added freshly.

## Flow cytometry analysis

Cells were trypsinized and stained with a mix of antibodies SSEA1-eFluor 660 (MC-480, eBioscience) and EpCAM-PE (G8.8, eBioscience), both at a 1:50 dilution in 100  $\mu$ l per sample. Staining was carried out by incubating on ice for 20 min, followed by washing and staining with DAPI (Sigma). Flow cytometry analysis was performed with a BD LSR II analyser. Gates and compensation between FITC and PE were adjusted on non-stained controls. 10.000 events (alive cells) per sample were acquired.



## Cloning of LEF1 isoforms and TIA1 in retroviral vectors

*Lef1* isoforms were amplified from bulk cDNAs of B cells and iPS (RNA-seq samples) and cloned using Gibson technology (Gibson *et al.* 2009) into a MIG-pMSCV retroviral vector including an IRES-GFP element. Kozak consensus sequences and an N-terminal T7 tag were inserted before the LEF1 ORF to ensure efficient translation and discrimination from endogenous *Lef1*.

cDNA of mouse *Tia1* was purchased from GenScript in pcDNA3.1 vectors (clone ID: OMu08423D) and cloned using a similar Gibson strategy into the MIG-pMSCV retroviral vector bearing the N-terminal T7 and IRES-GFP elements. Gibson reaction master mixes were provided by the CRG Protein Technologies Core Facility.

## Short hairpin RNAs

Five MISSION shRNAs specific for each splicing regulator were purchased from Sigma in pLKO lentiviral vectors and their effects were compared with those of SHC002 mammalian non-targeting MISSION shRNA (Sigma). The effect of each shRNA was tested on MEFs and E14 embryonic stem cells, and knockdown efficiency was assessed by RT-qPCR at 72/96h post-infection using primers specific for each target and normalised by the expression of two housekeeping genes (*Gapdh* and *Rplp0*). The two shRNAs displaying the largest knockdown effects in both cell lines were selected (sequences shown in Table S6).

## RNA extraction and retrotranscription, semi-quantitative RT-PCR and real-time qPCR (RT-qPCR)

RNA extraction and DNase treatment was performed using Maxwell simplyRNA kit (Promega) or RNeasy Mini kit (Qiagen), following the manufacturers' instructions. 200 ng of total RNA were reverse-transcribed

with Superscript III (Invitrogen) following the manufacturer's recommendations.

PCR reactions were carried out using GoTaq enzyme (Promega) with 1 µl of cDNA diluted 1:5. To quantify inclusion of alternatively spliced exons, capillary electrophoresis was performed using a Labchip GX Caliper workstation (Caliper, Perkin Elmer) at the CRG Protein Technologies Core Facility. The nanomolar content of each band was extracted with LabChip GX software and PSI values were calculated as the ratio between the inclusion amplicon and the sum of inclusion and skipping amplicons.

Real Time quantitative PCR (RT-qPCR) was performed on a ViiA7 Real Time PCR System (Applied Biosystems). Reactions in a total volume of 10µl contained 2x SYBR Green Master Mix (Applied Biosystems), primers 400 nM and 1µl of previously synthesized cDNA, diluted 1:5-20. The output Ct values were normalized by the expression of the housekeeping gene GAPDH (unless differently stated) and analysed with  $\Delta Ct/\Delta\Delta Ct$  method. All primers sequences are listed in Table S6.

## Acknowledgements

We thank the CRG Genomics Unit for sequencing; the CRG Tissue Engineering Unit for the inactivated MEF feeders; the CRG/UPF FACS Unit for FACS sorting; Daniel Soronellas for help with the *Mfuzz* package and Konrad Hochedlinger for the *Col-OKSM* reprogrammable mouse strain. We thank members of Manuel Irimia's, T.G.'s and J.V.'s groups for discussions; Manuel Irimia and members of J.V.'s group for critical reading of the manuscript. C.V. was recipient of an FPI-Severo Ochoa Fellowship from the Spanish Ministry of Economy and Competitiveness Work in J.V. laboratory is supported by the European Research Council (ERC AdvG 670146), AGAUR, Spanish Ministry of Economy and Competitiveness (BFU 2017

89308-P) and the Centre of Excellence Severo Ochoa. Work in T.G.'s laboratory was supported by the European Research Council FP7/2007-2013 (ERC Synergy Grant 4D-Genome) the Ministerio de Educación y Ciencia (SAF.2012-37167) and AGAUR.

## Authors contributions

C.V., T.G., and J.V. conceived the study. C.V., P.P., J.L.S. and J.V. designed experiments and data analyses. C.V. and J.V. wrote the manuscript with input from all authors. C.V. performed the majority of the experimental work and data analyses (RNA-seq analyses, MEF reprogramming and molecular biology experiments). R.S., B.D.S. and C.B. performed the B cell reprogramming experiments for RNA-seq. S.G., A.M. and B.P. provided reprogrammable MEFs. J.V. and T.G. acquired funding.

## Declaration of Interests

The authors declare no competing interests.

## References

Ajith, S. *et al.* (2016) 'Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells', *RNA Biology*. Taylor & Francis, 13(6), pp. 569–581. doi: 10.1080/15476286.2016.1176663.

Amit, M. *et al.* (2012) 'Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition', *Cell Reports*. Cell Press, 1(5), pp. 543–556. doi: 10.1016/j.celrep.2012.03.013.

Baralle, F. E. and Giudice, J. (2017) 'Alternative splicing as a regulator of development and tissue identity', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(7), pp. 437–451. doi: 10.1038/nrm.2017.27.

Blackwell, D. *et al.* (2020) 'Hnrnpul1 loss of function affects skeletal and limb development', *bioRxiv*. doi: 10.1101/2020.02.04.934257.

Brady, J. J. *et al.* (2013) 'Early role for IL-6 signalling during generation of induced pluripotent stem cells revealed by heterokaryon RNA-Seq', *Nature Cell Biology*, 15(10), pp. 1244–1252. doi: 10.1038/ncb2835.

Bruhn, L., Munnerlyn, A. and Grosschedl, R. (1997) 'ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCR $\alpha$  enhancer function', *Genes and Development*, 11(5), pp. 640–653. doi: 10.1101/gad.11.5.640.

Choi, H. S. *et al.* (2011) 'Identification and characterization of adenovirus early region 1B-associated protein 5 as a surface marker on undifferentiated human embryonic stem cells', *Stem Cells and Development*, 20(4), pp. 609–620. doi: 10.1089/scd.2010.0265.

Cieply, B. *et al.* (2016) 'Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins', *Cell Reports*, 15(2), pp. 247–255. doi: 10.1016/j.celrep.2016.03.025.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Eden, E. *et al.* (2009) 'GORilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics*, 10. doi: 10.1186/1471-2105-10-48.

Förch, P. *et al.* (2000) 'The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing', *Molecular Cell*, 6(5), pp. 1089–1098. doi: 10.1016/S1097-2765(00)00107-6.

Förch, P. *et al.* (2002) 'The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites', *EMBO Journal*, 21(24), pp. 6882–6892. doi: 10.1093/emboj/cdf668.

Futschik, M. E. and Carlisle, B. (2005) 'Noise-robust soft clustering of gene expression time-course data', *Journal of Bioinformatics and Computational Biology*, 3(4), pp. 965–988. doi: 10.1142/S0219720005001375.

Gabut, M. *et al.* (2011) 'An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming', *Cell*. Elsevier Inc., 147(1), pp. 132–146. doi: 10.1016/j.cell.2011.08.023.

Van Genderen, C. *et al.* (1994) 'Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice', *Genes and Development*, 8(22), pp. 2691–2703. doi: 10.1101/gad.8.22.2691.

Gibson, D. G. *et al.* (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods*. Nature Publishing Group, 6(5), pp. 343–345. doi: 10.1038/nmeth.1318.

Gohr, A. and Irimia, M. (2019) 'Matt: Unix tools for alternative splicing analysis', *Bioinformatics*, 35(1), pp. 130–132. doi: 10.1093/bioinformatics/bty606.

Gopalakrishna-Pillai, S. and Iverson, L. E. (2011) 'A DNMT3B alternatively spliced exon and encoded peptide are novel biomarkers of human pluripotent stem cells', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0020663.

Gopalakrishnan, S. *et al.* (2009) 'A novel DNMT3B splice variant expressed in tumor and pluripotent cells modulates genomic DNA methylation patterns and displays altered DNA binding', *Molecular Cancer Research*,

7(10), pp. 1622–1634. doi: 10.1158/1541-7786.MCR-09-0018.

Han, H. *et al.* (2013) 'MBNL proteins repress ES-cell-specific alternative splicing and reprogramming', *Nature*, 498(7453), pp. 241–245. doi: 10.1038/nature12270.

Hulsen, T., de Vlieg, J. and Alkema, W. (2008) 'BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams', *BMC Genomics*, 9. doi: 10.1186/1471-2164-9-488.

Irimia, M. *et al.* (2014) 'A highly conserved program of neuronal microexons is misregulated in autistic brains', *Cell*. Elsevier Ltd, 159(7), pp. 1511–1523. doi: 10.1016/j.cell.2014.11.035.

Izquierdo, J. M. *et al.* (2005) 'Regulation of fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition', *Molecular Cell*, 19(4), pp. 475–484. doi: 10.1016/j.molcel.2005.06.015.

Jesse, S. *et al.* (2010) 'Lef-1 isoforms regulate different target genes and reduce cellular adhesion', *International Journal of Cancer*, 126(5), pp. 1109–1120. doi: 10.1002/ijc.24802.

Kanitz, A. *et al.* (2019) 'Conserved regulation of RNA processing in somatic cell reprogramming', *BMC Genomics*, 20(1), pp. 1–19. doi: 10.1186/s12864-019-5438-2.

Kedersha, N. L. *et al.* (1999) 'RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2α to the assembly of mammalian stress granules', *Journal of Cell Biology*, 147(7), pp. 1431–1441. doi: 10.1083/jcb.147.7.1431.

Kumar, L. and Futschik, M. E. (2007) 'Mfuzz: A software package for soft clustering of microarray data', *Bioinformatics*, 21(1), pp. 5–7. doi: 10.6026/97320630002005.

Kyburz, A. *et al.* (2006) 'Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3' End Processing and Splicing', *Molecular Cell*, 23(2), pp. 195–205. doi: 10.1016/j.molcel.2006.05.037.

Larson, A., Fair, B. J. and Pleiss, J. A. (2016) 'Interconnections between RNA-processing pathways revealed by a sequencing-based genetic screen for pre-mRNA splicing mutants in fission yeast', *G3: Genes, Genomes, Genetics*, 6(6), pp. 1513–1523. doi: 10.1534/g3.116.027508.

Levanon, D. *et al.* (1998) 'Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors', *Proceedings of the National Academy of Sciences of the United States of America*, 95(20), pp. 11590–11595. doi: 10.1073/pnas.95.20.11590.

Lewis, B. P., Green, R. E. and Brenner, S. E. (2003) 'Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans', *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), pp. 189–192. doi: 10.1073/pnas.0136770100.

Lu, Y. *et al.* (2014) 'Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells', *Cell*

*Stem Cell*, 15(1), pp. 92–101. doi: 10.1016/j.stem.2014.04.002.

Mallory, M. J. *et al.* (2011) 'Signal- and Development-Dependent Alternative Splicing of LEF1 in T Cells Is Controlled by CELF2', *Molecular and Cellular Biology*, 31(11), pp. 2184–2195. doi: 10.1128/mcb.05170-11.

Mallory, M. J. *et al.* (2015) 'Induced transcription and stability of CELF2 mRNA drives widespread alternative splicing during T-cell signaling', *Proceedings of the National Academy of Sciences of the United States of America*, 112(17), pp. E2139–E2148. doi: 10.1073/pnas.1423695112.

McCarthy, D. J., Chen, Y. and Smyth, G. K. (2012) 'Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation', *Nucleic Acids Research*, 40(10), pp. 4288–4297. doi: 10.1093/nar/gks042.

Misra, A. *et al.* (2014) 'Global Promotion of Alternative Internal Exon Usage by mRNA 3' End Formation Factors', *Molecular Cell*. Elsevier Inc., 58(5), pp. 819–831. doi: 10.1016/j.molcel.2015.03.016.

Misra, A. *et al.* (2015) 'Global analysis of CPSF2-mediated alternative splicing: Integration of global iCLIP and transcriptome profiling data', *Genomics Data*. The Authors, 6, pp. 217–221. doi: 10.1016/j.gdata.2015.09.022.

Mosteiro, L. *et al.* (2016) 'Tissue damage and senescence provide critical signals for cellular reprogramming in vivo', *Science*, 354(6315). doi: 10.1126/science.aaf4445.

Ohta, S. *et al.* (2013) 'Global Splicing Pattern Reversion during Somatic Cell Reprogramming', *Cell Reports*, 5(2), pp. 357–366. doi: 10.1016/j.celrep.2013.09.016.

Oosterwegel, M. *et al.* (1993) 'Differential expression of the HMG box factors TCF-1 and LEF-1 during murine embryogenesis', *Development*, 118(2), pp. 439–448. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8223271>.

Pan, Q. *et al.* (2004) 'Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform', *Molecular Cell*, 16(6), pp. 929–941. doi: 10.1016/j.molcel.2004.12.004.

Pan, Q. *et al.* (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nature Genetics*, 40(12), pp. 1413–1415. doi: 10.1038/ng.259.

Piczyk, M. *et al.* (2000) 'TIA-1 is a translational silencer that selectively regulates the expression of TNF-α', *EMBO Journal*, 19(15), pp. 4154–4163. doi: 10.1093/emboj/19.15.4154.

Polo, J. M. *et al.* (2012) 'A molecular roadmap of reprogramming somatic cells into iPS cells', *Cell*. Elsevier Inc., 151(7), pp. 1617–1632. doi: 10.1016/j.cell.2012.11.039.

Polo, S. E. *et al.* (2012) 'Regulation of DNA-End Resection by hnRNPU-like Proteins Promotes DNA Double-Strand Break Signaling and Repair', *Molecular Cell*. Elsevier Inc., 45(4), pp. 505–516. doi: 10.1016/j.molcel.2011.12.035.

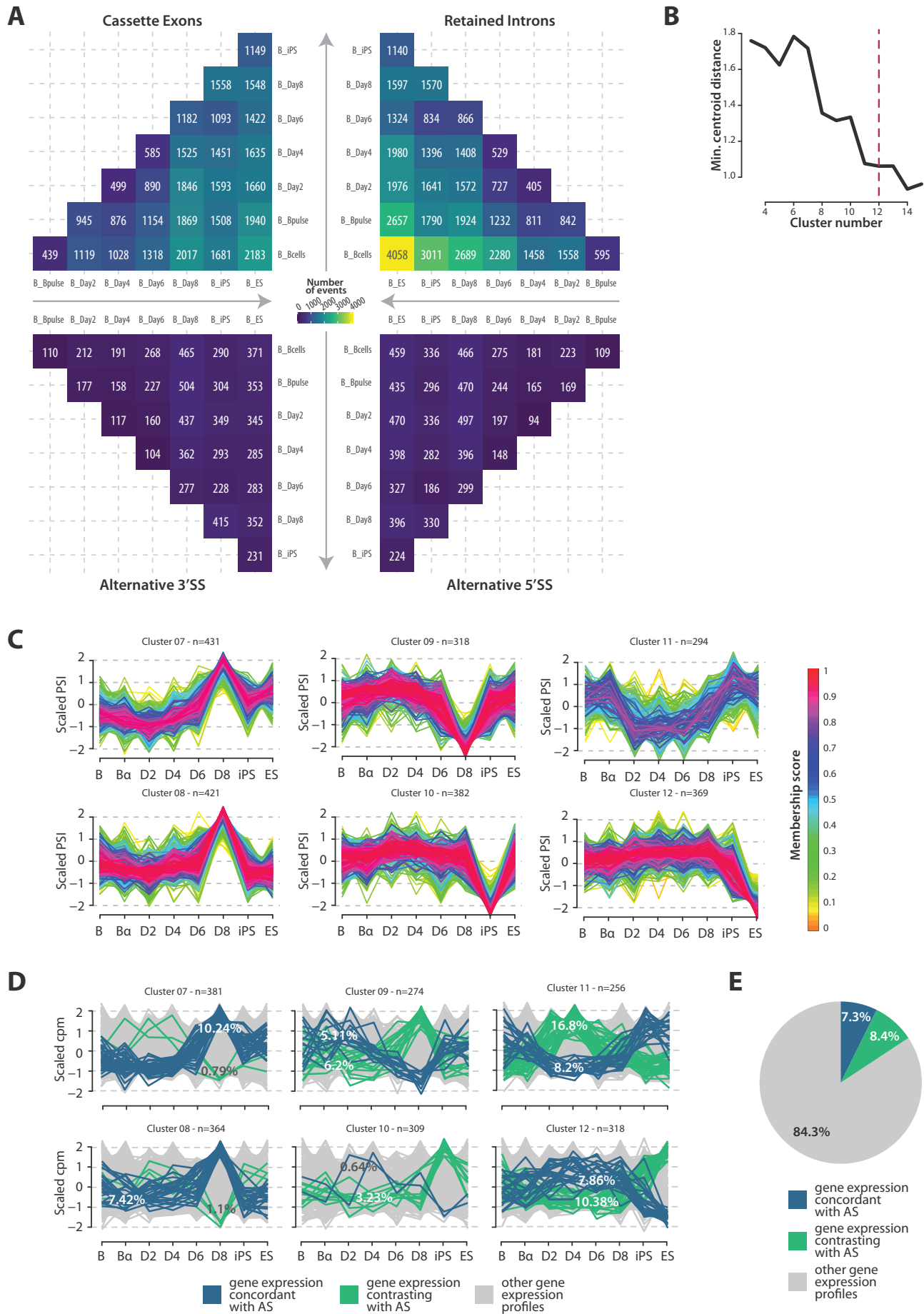
- Ray, D. *et al.* (2013) 'A compendium of RNA-binding motifs for decoding gene regulation', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 499(7457), pp. 172–177. doi: 10.1038/nature12311.
- Ritschka, B. *et al.* (2017) 'The senescence-associated secretory phenotype induces cellular plasticity and tissue regeneration', *Genes and Development*, 31(2), pp. 172–183. doi: 10.1101/gad.290635.116.
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2009) 'edgeR: A Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140. doi: 10.1093/bioinformatics/btp616.
- Rodriguez-Ubrea, J. *et al.* (2014) 'C/EBPα-Mediated Activation of MicroRNAs 34a and 223 Inhibits Lef1 Expression To Achieve Efficient Reprogramming into Macrophages', *Molecular and Cellular Biology*, 34(6), pp. 1145–1157. doi: 10.1128/mcb.01487-13.
- Salomonis, N. *et al.* (2010) 'Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation', *Proceedings of the National Academy of Sciences of the United States of America*, 107(23), pp. 10514–10519. doi: 10.1073/pnas.0912260107.
- Sánchez-Jiménez, C. and Izquierdo, J. M. (2013) 'T-cell Intracellular Antigen (TIA)-Proteins Deficiency in Murine Embryonic Fibroblasts Alters Cell Cycle Progression and Induces Autophagy', *PLoS ONE*, 8(9). doi: 10.1371/journal.pone.0075127.
- Sardina, J. L. *et al.* (2018) 'Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate', *Cell Stem Cell*, 23(5), pp. 727–741.e9. doi: 10.1016/j.stem.2018.08.016.
- Solana, J. *et al.* (2016) 'Conserved functional antagonism of CELF and MBNL proteins controls stem cell-specific alternative splicing in planarians', *eLife*, 5(AUGUST), pp. 1–29. doi: 10.7554/eLife.16797.001.
- Stadhouders, R. *et al.* (2018) 'Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming', *Nature Genetics*. Springer US, 50(2), pp. 238–249. doi: 10.1038/s41588-017-0030-7.
- Stadtfield, M. *et al.* (2010) 'A reprogrammable mouse strain from gene-targeted embryonic stem cells', *Nature Methods*, 7(1), pp. 53–55. doi: 10.1038/nmeth.1409.
- Di Stefano, B. *et al.* (2014) 'C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells', *Nature*. Nature Publishing Group, 506(7487), pp. 235–239. doi: 10.1038/nature12885.
- Di Stefano, B. *et al.* (2016) 'C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4', *Nature Cell Biology*, 18(4), pp. 371–381. doi: 10.1038/ncb3326.
- Di Stefano, B. and Graf, T. (2016) Very rapid and efficient generation of induced pluripotent stem cells from mouse pre-B Cells, *Methods in Molecular Biology*. doi: 10.1007/7651\_2014\_133.
- Stelzer, G. *et al.* (2016) 'The GeneCards suite: From gene data mining to disease genome sequence analyses', *Current Protocols in Bioinformatics*, 2016, pp. 1.30.1-1.30.33. doi: 10.1002/cpbi.5.
- Supek, F. *et al.* (2011) 'Revigo summarizes and visualizes long lists of gene ontology terms', *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0021800.
- Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.
- Tapial, J. *et al.* (2017) 'An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms', *Genome Research*, 27(10), pp. 1759–1768. doi: 10.1101/gr.220962.117.
- Toh, C. X. D. *et al.* (2016) 'RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming', *Cell Reports*, 15(12), pp. 2597–2607. doi: 10.1016/j.celrep.2016.05.049.
- Travis, A. *et al.* (1991) 'LEF-1, a gene encoding a lymphoid-specific protein, an HMG domain, regulates T-cell receptor α enhancer function', *Genes and Development*, 5(5), pp. 880–894. doi: 10.1101/gad.5.5.880.
- Ule, J. *et al.* (2010) 'iCLIP predicts the dual splicing effects of TIA-RNA interactions', *PLoS Biology*, 8(10). doi: 10.1371/journal.pbio.1000530.
- Venables, J. P. *et al.* (2013) 'MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation', *Nature Communications*. Nature Publishing Group, 4(May), p. 2480. doi: 10.1038/ncomms3480.
- Wang, E. T. *et al.* (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456(7221), pp. 470–476. doi: 10.1038/nature07509.
- Wang, I. *et al.* (2014) 'Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1', *Nucleic Acids Research*, 42(9), pp. 5949–5966. doi: 10.1093/nar/uku193.
- Waterman, M. L., Fischer, W. H. and Jones, K. A. (1991) 'A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer', *Genes and Development*, 5(4), pp. 656–669. doi: 10.1101/gad.5.4.656.
- Yamazaki, T. *et al.* (2018) 'TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency', *Genes and Development*, 32(17–18), pp. 1161–1174. doi: 10.1101/gad.316984.118.
- Yeo, G. and Burge, C. B. (2004) 'Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals', in *Journal of Computational Biology*, pp. 377–394. doi: 10.1089/1066527041410418.
- Yeo, G. W. *et al.* (2009) 'Doi:10.1038/Nsmb.1545', *Nature Structural Molecular Biology*, 16(2), pp. 130–137. doi: 10.1038/nsmb.1545.
- Zavolan, M. and Kanitz, A. (2018) 'RNA splicing and its connection with other regulatory layers in somatic cell

reprogramming', *Current Opinion in Cell Biology*. Elsevier Ltd, 52, pp. 8–13. doi: 10.1016/j.ceb.2017.12.002.

Zhou, P. *et al.* (1995) 'Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate', *Genes and Development*, 9(6), pp. 700–713. doi: 10.1101/gad.9.6.700.

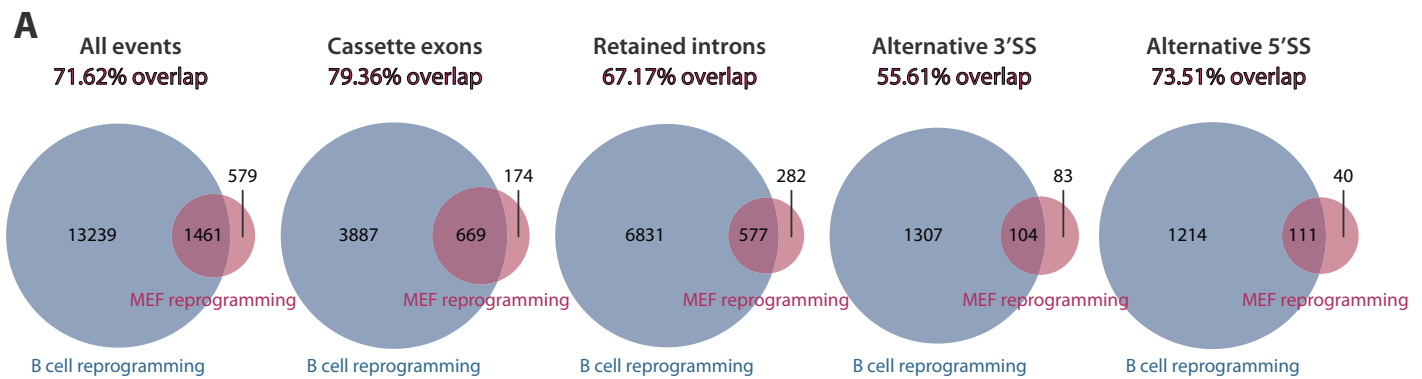


## **Supplementary Figures**

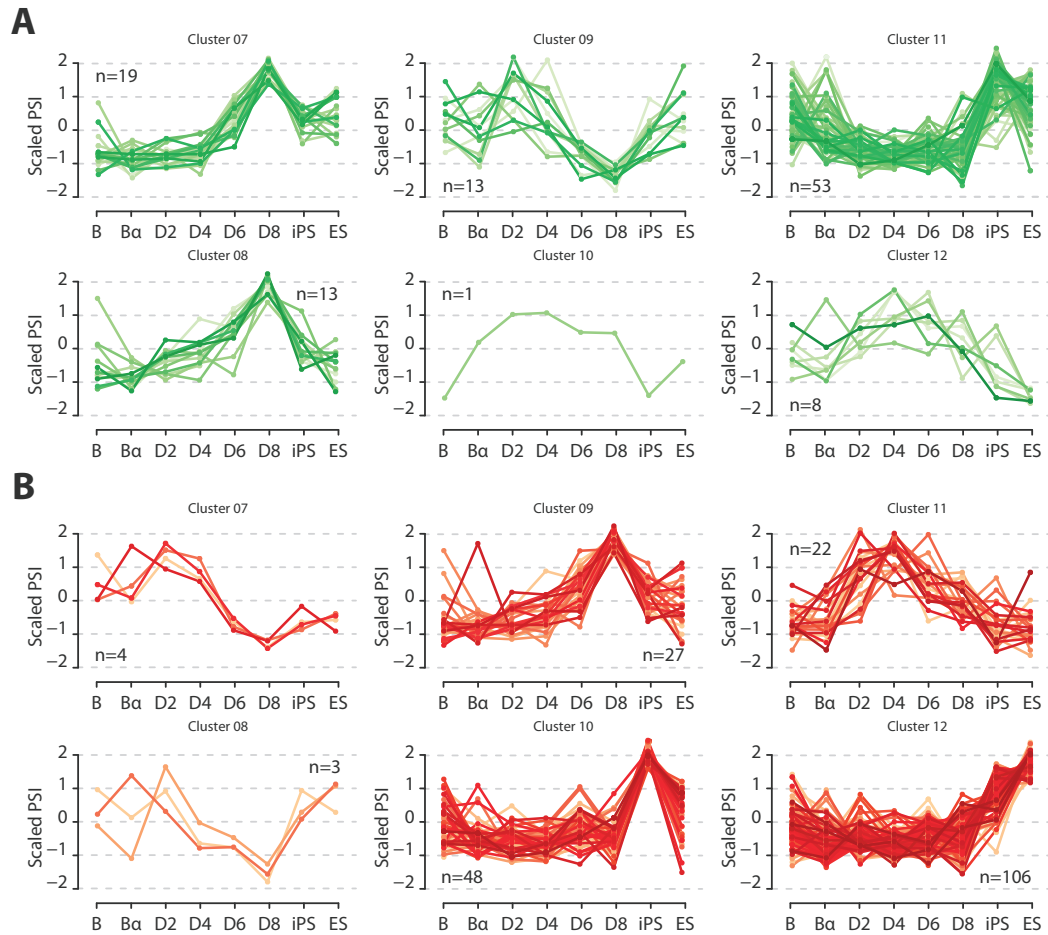


**Supplementary Figure S1.** (legend on next page)

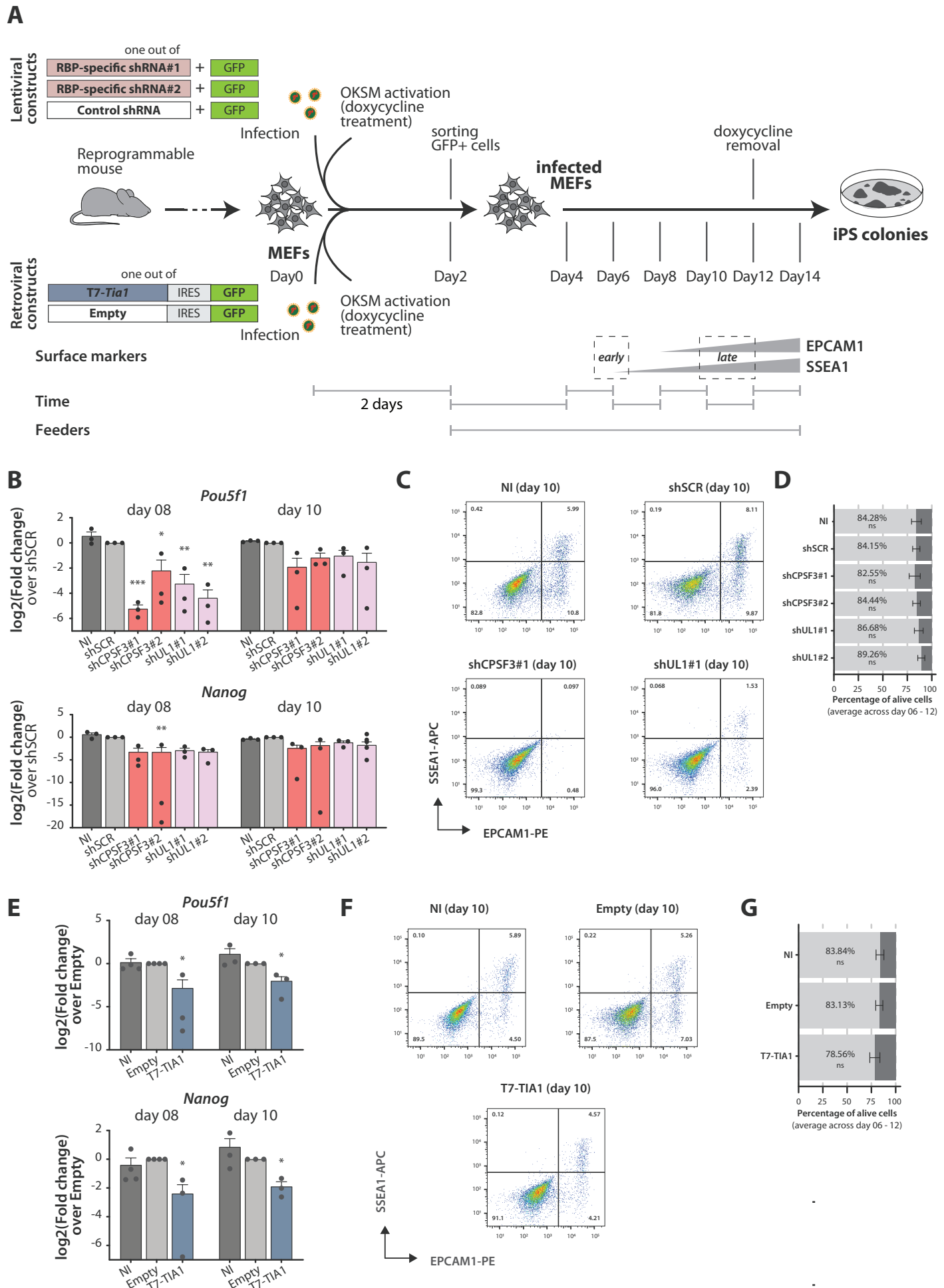
**Supplementary Figure S1.** (A) Heatmaps representing the number of AS events for each AS category (Cassette exons, retained introns, alternative 3' and 5' splice sites) which are differentially spliced between each pair of samples across the C/EBP $\alpha$ -mediated reprogramming dataset. The direction of reprogramming time is indicated on the axes by arrows. Related to Figure 1B. (B) Variation of the minimum centroid distance score with the number of clusters produced, according to which 12 clusters were selected (magenta dashed line). (C) AS clusters not shown in Figure 1C. The size of each cluster is indicated (n). (D) Gene expression patterns of genes containing the exons belonging to each of the AS clusters in panel C. Genes with expression correlating with the cluster centroid or its negative (membership > 0.3) are highlighted in blue or green, respectively. Percentage of concordant/contrasting patterns are displayed for each cluster. (E) Pie chart indicating the fraction of genes whose gene expression changes are concordant / contrasting with the profiles of AS changes corresponding to their respective cassette exons (average of all the clusters). Colors as in panel G.



**Supplementary Figure S2.** (A) Venn diagrams representing the overlap of events differentially spliced between B cell reprogramming (grey) and MEF reprogramming (magenta) from (Cieply et al. 2016).



**Supplementary Figure S3. (A)** Gene expression profiles of RNA-binding proteins correlating with the centroid of each of the remaining AS clusters (shown in Figure S1C and not in 1C) (positive regulators). Average scaled cpm values are represented by each line and the number of regulators of each cluster is indicated (n). **(B)** Gene expression profiles of RNA-binding proteins correlating with the negative of the centroid of each of the remaining AS clusters (shown in Figure S1C and not in 1C) (negative regulators). Average scaled cpm values are represented by each line and the number of regulators in each cluster is indicated (n).

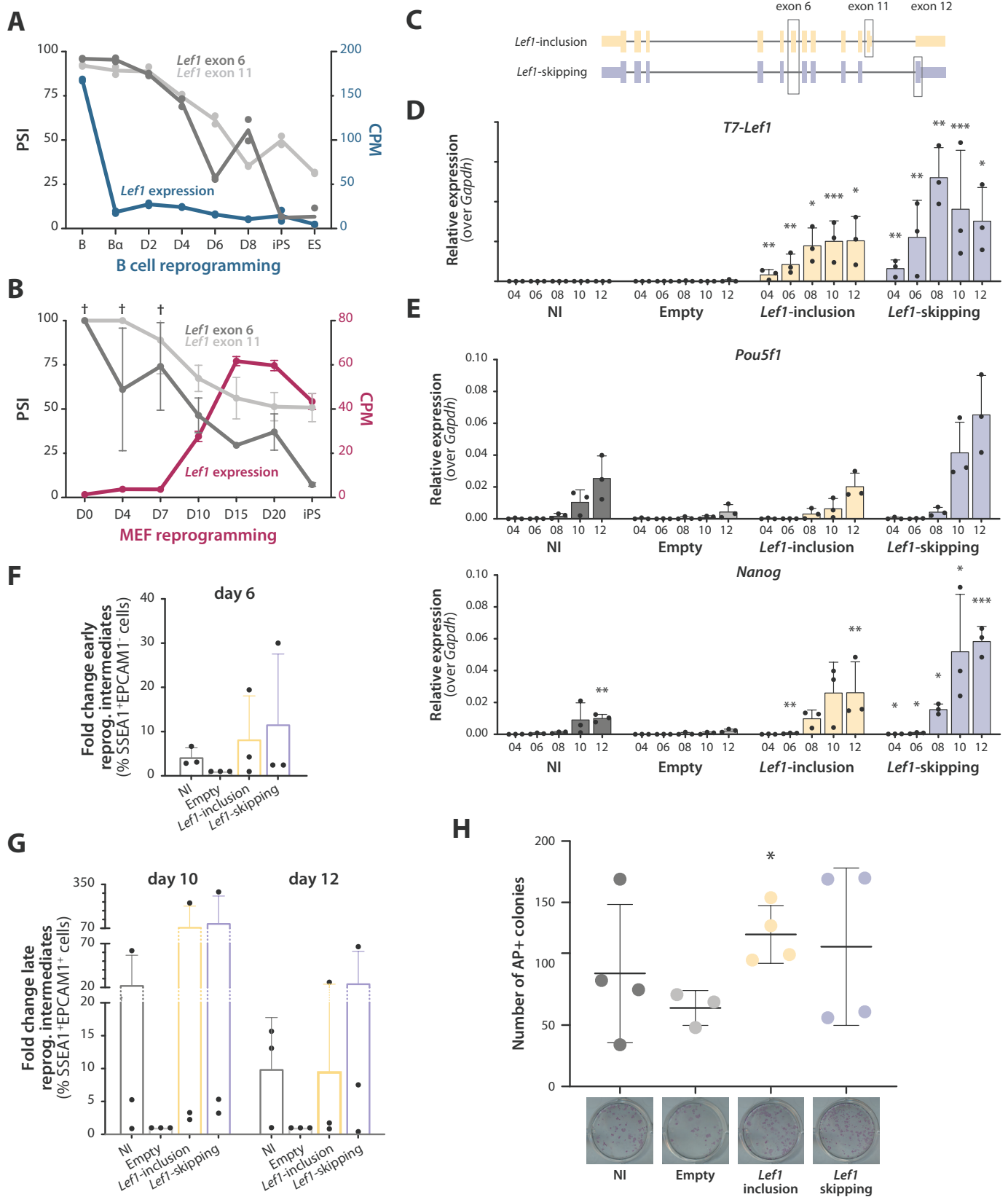


**Supplementary Figure S4.** (legend on next page)



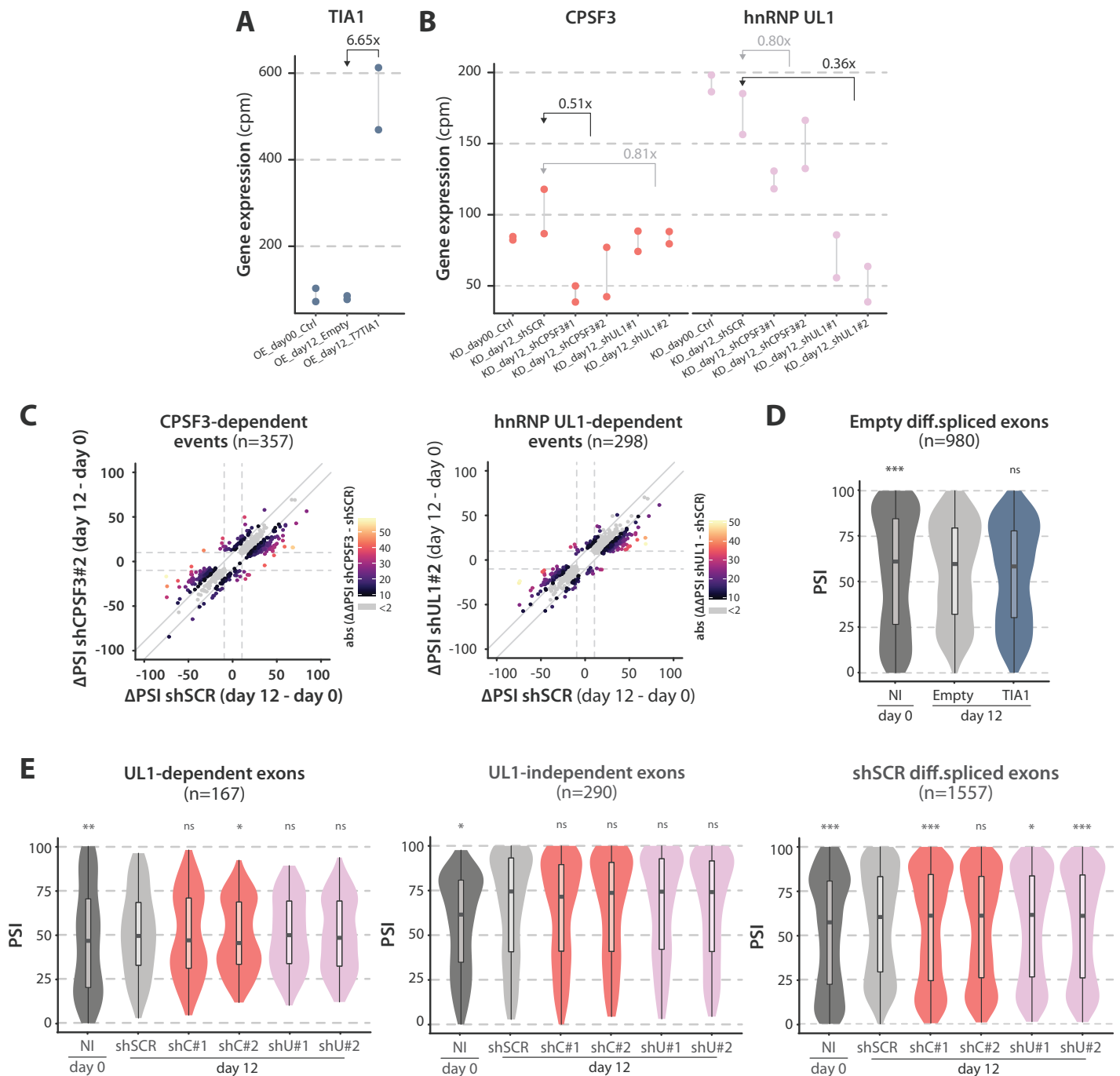
**Supplementary Figure S4.** (A) Schematic representation of the protocol used to knockdown Cpsf3 and Hnrnpul1 or overexpress T7-tagged Tia1 at early stages of MEF reprogramming. (B) Gene expression levels of endogenous Pou5f1 and Nanog mRNAs during MEF reprogramming quantified by RT-qPCR and represented as log2(Fold change) for every condition compared to the shSCR control. Average of biological triplicates and SEM are shown. Statistical significance was calculated by two-way ANOVA on  $\Delta\Delta Ct$  values comparing to shSCR (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Sidak method). (C) Examples of gating set ups for cell sorting of SSEA1 and EPCAM1 markers at day 10 of reprogramming. Numbers within the rectangles indicate percentages of cells in each gate. (D) Barplot representing the percentage of alive (DAPI-, light gray) and dead (DAPI+, dark gray) cells. Average and SD of days 6, 8, 10 and 12 are shown for each condition. Statistical significance was calculated by Fisher's exact test against shSCR condition (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively). (E) Expression levels of endogenous Pou5f1 and Nanog mRNAs during MEF reprogramming quantified by RT-qPCR and represented as log2(Fold change) of every condition compared to the Empty control. Average of biological replicates and SEM are shown. Statistical significance was calculated by two-way ANOVA on  $\Delta\Delta Ct$  values comparing to Empty vector (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Sidak method). (F) Examples of gating set ups for cell sorting of SSEA1 and EPCAM1 markers at day 10 of reprogramming. Numbers within the rectangles indicate percentages of cells in each gate. (G) Barplot representing the percentage of alive (DAPI-, light gray) and dead (DAPI+, dark gray) cells. Average and SD of days 6, 8, 10 and 12 are shown for each condition. Statistical significance was calculated by Fisher's exact test against Empty condition (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively).

34



Supplementary Figure S5. (legend on next page)

**Supplementary Figure S5.** (A) Gene expression of Lef1 gene (cpm values, blue line, right y axis) and PSI values of Lef1 exon 6 and 11 (grey lines, left y axis) in B cell reprogramming. (B) Gene expression of Lef1 gene (cpm values, magenta line, right y axis) and PSI values of Lef1 exon 6 and 11 (grey lines, left y axis) in MEF reprogramming. Asterisks point out PSI values calculated with low coverage (less than 10 actual reads). (C) Lef1 pre-mRNA isoforms including exons 6 and 11 (Lef1-inclusion) or skipping both (Lef1-skipping) are represented. Thin lines represent introns, while wider regions represent exons. Lighter and thinner parts of exons depict 5' and 3'UTR regions. Alternative exons are highlighted. (D) Relative expression levels of T7-Lef1 mRNA quantified by RT-qPCR using specific primers that only amplify exogenous T7-tagged Lef1. The y axis represents relative expression ( $2^{(-\Delta Ct)}$  value) of T7-Lef1 after normalization over Gapdh. (E) Relative expression levels of Pou5f1 and Nanog quantified by RT-qPCR as in panel D. (D,E) Average of biological replicates and SD values are shown (n=4). Statistical significance was calculated by t-test on  $\Delta Ct$  values comparing to the Empty control (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Holm-Sidak method). (F) Increase in early reprogramming intermediates at day 6 post-OSKM induction upon overexpression of Lef1 isoforms. Fold change was calculated from the percentage of SSEA1+EPCAM1- cells (of the total of alive cells) in every condition compared to Empty control using flow cytometry analysis. (G) Increase in late reprogramming intermediates at days 10 and 12 post-OSKM induction upon overexpression of Lef1 isoforms. Fold change was calculated from the percentage of SSEA1+EPCAM1+ cells (of the total of alive cells) in every condition compared to the Empty control using flow cytometry analysis. (H) Number of colonies stained with alkaline phosphatase (AP) at day 14 post-OSKM induction upon Tia1 overexpression. The image of a representative well is shown for every condition. (F,G,H) Average of biological replicates and SD values are shown (n=4). Statistical significance was calculated by t-test comparing to the Empty control (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively, corrected for multiple testing with Holm-Sidak method).



**Supplementary Figure S6.** (A) Expression levels (cpm values) of Tia1 in the corresponding RNA-seq dataset. Fold change between cells overexpressing T7-Tia1 and control cells transduced with an Empty vector is shown (day12, average between replicates). (B) Expression levels (cpm values) of Cpsf3 and Hnrnpul1 in the corresponding RNA-seq dataset. Fold change values between cells infected with specific shRNAs and control shSCR are shown (day12, average between replicates and shRNAs). (C) CPSF3- and UL1-dependent events detected during reprogramming (left and right, respectively), representing the  $\Delta$ PSI with the second shRNA (as in Figure 6C). The x axis represents the  $\Delta$ PSI value between shSCR day 12 and day 0 control. The y axis represents the  $\Delta$ PSI value between shCPSF3#2 or shUL1#2 day 12 and day 0 control. Only the events differentially spliced with both CPSF3- or hnRNPUL1-specific shRNAs are shown. Events coloured with the palette indicated are defined as TIA1-dependent events (average  $|\Delta\Delta\text{PSI}(\text{shRNA} - \text{shSCR})| \geq 10$ ), while the others are represented by grey dots. (D) Relative to Figures 6E-F. Violin plots representing the distribution of PSI values of exons differentially spliced during reprogramming in Empty conditions ( $|\Delta\text{PSI}(\text{Empty day12} - \text{NI day0})| \geq 10$  and range  $\geq 5$ ). (E) Relative to Figures 6J-K. Violin plots representing the distribution of PSI values of UL1-dependent/-independent exons (top and central panel) and exons differentially spliced during reprogramming in shSCR conditions ( $|\Delta\text{PSI}(\text{shSCR day12} - \text{NI day0})| \geq 10$  and range  $\geq 5$ , bottom panel). (D,E) Statistical significance was calculated by Fisher's exact test comparing number of events with intermediate ( $25 < \text{PSI} < 75$ ) or extreme PSI values ( $\text{PSI} \geq 75$  or  $\leq 25$ ) in each condition against shSCR control (\*, \*\*, \*\*\* = p value < 0.05, 0.01, 0.001 respectively).

## Supplementary Tables

**Table S1. Cassette exon events differentially spliced in B cell reprogramming.** This table includes the information about the AS cluster each exon belongs to and the predicted effect on its transcript. Data for other types of events is available upon request.

**Table S2. Cassette exon events differentially spliced in MEF reprogramming.** This table includes the information about the overlap with B cell reprogramming. Data for other types of events is available upon request.

**Table S3. Gene expression levels of RBPs correlating with AS clusters** (positive and negative predicted regulators).

**Table S4. CPSF3-, UL1- and TIA1-dependent and -independent events.** This table includes the sets of control cassette exons (CEX) used for KD (CPSF3- and UL1-dependent CEX) or OE (TIA1-dependent CEX).

**Table S5. Gene Ontology terms enriched in CPSF3-, UL1- and TIA1-dependent events.** This table includes outputs from GOrilla and REVIGO analyses.

**Table S6. Sequences of primers and shRNAs used in this study.**