Temporal epigenomic profiling identifies AHR and GLIS1 as super-enhancer controlled regulators of mesenchymal multipotency

Deborah Gérard[1], Florian Schmidt[2,3], Aurélien Ginolhac[1], Martine Schmitz[1], Rashi Halder[4], Peter Ebert[3], Marcel H. Schulz[2,3], Thomas Sauter[1] and Lasse Sinkkonen[1*]

[1]Life Sciences Research Unit, University of Luxembourg, L-4367 Belvaux, Luxembourg

[2]Excellence Cluster for Multimodal Computing and Interaction, Saarland Informatics Campus, Germany

[3]Computational Biology & Applied Algorithmics, Max Planck Institute for Informatics, Saarland Informatics Campus, Germany

[4]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, L-4362, Luxembourg

Deborah.Gerard@uni.lu; fschmidt@mmci.uni-saarland.de; Aurelien.Ginolhac@uni.lu; Martine.Schmitz@uni.lu; Rashi.Halder@uni.lu; pebert@mpi-inf.mpg.de; mschulz@mmci.uni-saarland.de; Thomas.Sauter@uni.lu; Lasse.Sinkkonen@uni.lu

[*]**Corresponding author:** Dr. Lasse Sinkkonen
Life Sciences Research Unit, University of Luxembourg
6, Avenue du Swing, L-4367 Belvaux, Luxembourg
Tel.: +352-4666446839
E-mail: lasse.sinkkonen@uni.lu

ABSTRACT

Temporal data on gene expression and context-specific open chromatin states can improve identification of key transcription factors (TFs) and the gene regulatory networks (GRNs) controlling cellular differentiation. However, their integration remains challenging. Here, we delineate a general approach for data-driven and unbiased identification of key TFs and dynamic GRNs, called EPIC-DREM. We generated time-series transcriptomic and epigenomic profiles during differentiation of mouse multipotent bone marrow stromal cells (MSCs) towards adipocytes and osteoblasts. Using our novel approach we constructed time-resolved GRNs for both lineages. To prioritize the identified shared regulators, we mapped dynamic super-enhancers in both lineages and associated them to target genes with correlated expression profiles. We identified aryl hydrocarbon receptor (AHR) and Glis family zinc finger 1 (GLIS1) as mesenchymal key TFs controlled by dynamic MSC-specific super-enhancers that become repressed in both lineages. AHR and GLIS1 control differentiation-induced genes and we propose they function as guardians of mesenchymal multipotency.

Keywords: gene regulatory network / epigenomics / super-enhancer / mesenchymal differentiation / AHR / adipocyte / osteoblast / Notch signaling /

INTRODUCTION

Understanding the gene regulatory interactions underlying cell differentiation and identity has become increasingly important, especially in regenerative medicine. Efficient and specific reprogramming of cells towards desired differentiated cell types relies on understanding of the cell type-specific regulators and their targets (Rackham et al., 2016). Similarly, knowledge of the regulatory wiring in the intermediate stages might allow controlled partial dedifferentiation, and thereby endogenous regeneration, also in mammals (Aguirre et al., 2014).

Great progress has been made in reconstruction of GRNs for various cell types in recent years. While successful, many of the approaches derive their regulatory interactions from existing literature and databases, which may be limiting as the majority of enhancers harboring transcription factor (TF) binding sites are cell type-specific (Forrest et al., 2014). Thus, the regulatory interactions derived from existing databases and literature might be misleading and are likely to miss important interactions that have not been observed in other cell types. Therefore, context-specific expression data have been used to overcome such biases and allow a data-driven network reconstruction (Janky et al., 2014). In addition, other approaches taking advantage of time-series data, such as Dynamic Regulatory Events Miner (DREM) (Schulz et al., 2012), have been developed to allow hierarchical identification of the regulatory interactions. However, while time-series epigenomic data has been used in different studies to derive time point-specific GRNs (Ramirez et al., 2017) (Goode et al., 2016), systematic approaches that integrate the different types of data in an intuitive and automated way are missing.

The central key genes of biological networks under multi-way regulation by many TFs and signaling pathways were recently shown to be enriched for disease genes and

are often controlled through so called super-enhancers (SEs), large regulatory regions characterized by broad signals for enhancer marks like H3 lysine 27 acetylation (H3K27ac) (Galhardo et al., 2015) (Hnisz et al., 2013) (Parker et al., 2013) (Siersbaek et al., 2014). Hundreds of SEs can be identified per cell type, many of which are cell type- or lineage-specific and usually control genes that are important for the identity of the given cell type or condition. Thus, SE mapping and SE target identification can facilitate unbiased identification of novel key genes.

An example of lineage specification events with biomedical relevance is the differentiation of multipotent bone marrow stromal progenitor cells (MSCs) towards two mesenchymal cell types: osteoblasts and bone marrow adipocytes. Due to their shared progenitor cells, there is a reciprocal balance in the relationship between osteoblasts and bone marrow adipocytes. Proper osteoblast differentiation and maturation towards osteocytes is important in bone fracture healing and osteoporosis and osteoblast secreted hormones like osteocalcin can influence insulin resistance (N. K. Lee et al., 2007) (Silva & Kousteni, 2012). At the same time bone marrow adipocytes, that occupy as much as 70% of the human bone marrow (Fazeli et al., 2013), are a major source of hormones promoting metabolic health, including insulin sensitivity (Cawthorn et al., 2014). Moreover, increased commitment of the MSCs towards the adipogenic lineage upon obesity and aging was recently shown to inhibit both bone healing and the hematopoietic niche (Ambrosi et al., 2017).

Extensive temporal epigenomic analysis of osteoblastogenesis has been recently reported (H. Wu et al., 2017). However, a parallel investigation of two lineages originating from the same progenitor cells can help to understand both the lineage-specific and the shared regulators important for their (de)differentiation. To identify shared regulators of adipocyte and osteoblast commitment and to delineate a general

approach for systematic unbiased identification of key regulators, we performed time-series epigenomic and transcriptomic profiling at 6 different time points over 15 day differentiation of MSCs towards both adipocytes and osteoblasts. We combine segmentation-based TF binding predictions from time point-specific active enhancer data (Schmidt et al., 2017) with probabilistic modeling of temporal gene expression data (Schulz et al., 2012) to derive dynamic GRNs for both lineages. By merging overlapping SEs identified using H3K27ac signal from different time points we obtained dynamic profiles of SE activity across the two differentiations and use these dynamic SEs to prioritize the key regulators identified through the network reconstruction. With this approach, we identified aryl hydrocarbon receptor (AHR) and Glis family zinc finger 1 (GLIS1) as central regulators of multipotent MSCs under dynamic control from SEs that become repressed in the differentiated cells. The repression of these TFs, and in particular of AHR, allows upregulation of many adipocyte- and osteoblast-specific genes, including *Notch* genes, a family of conserved developmental regulators.

RESULTS

A subset of differentially expressed genes are shared between adipocyte and osteoblast differentiation

To identify shared regulators of MSC differentiation towards adipocytes and osteoblasts, and to delineate a general approach for a systematic unbiased identification of key regulators, we performed time-series ChIP-seq and RNA-seq profiling at 6 different time points over 15 days of differentiation of mouse ST2 MSCs (Figure 1A). Using ChIP-seq, genome-wide profiles of three different histone modifications, indicating active transcription start sites (TSS) (H3K4me3), active

enhancers (H3K27ac), and on-going transcription (H3K36me3) were generated. These data were complemented by corresponding time-series RNA-seq analysis. The successful differentiations were confirmed by induced expression of known lineage-specific marker genes and microscopic inspection of cellular morphology and stainings (Supplementary Figure S1). Interestingly, profiles of the adipogenic marker genes resembled those reported for the yellow adipose tissue (YAT) found in the bone marrow, rather than classic white adipose tissue (WAT) (Scheller et al., 2016), consistent with ST2 cells originating from bone marrow stroma (Figure S1A). Moreover, the expression profiles of *Sp7* and *Runx2* were consistent to those previously reported for mouse osteoblasts (Yoshida et al., 2012).

Principal component analysis of the obtained transcriptome profiles confirmed the specification of the cells towards two different lineages with differential temporal dynamics (Figure 1B). Osteoblastogenesis was accompanied by gradual and consistent progression towards a more differentiated cell type while adipogenesis showed more complex dynamics with a big transcriptome shift after one day of differentiation, followed by a more gradual progression during the following days. This is in keeping with the change in the composition of the differentiation medium from day 2 onwards (see Methods for details). In total the adipocyte differentiation was characterized by a total of 5156 significantly differentially expressed genes ($\log_2 FC \geq 1$, FDR<0.05) across the time series compared to the MSCs (Figure 1C; Supplementary Table S1). During osteoblast differentiation 2072 genes were differentially expressed at least at one time point. 1401 of these genes were affected in both lineages. However, as illustrated by the top 100 genes with highest variance across the time points, which are depicted in the heatmap in Figure 1D, most genes exhibit either lineage-specific or opposing behavior between the lineages. Only a

subset of genes showed similar changes in both lineages (Figure 1D). Thus, narrowing the list of genes that could serve as shared regulators of both differentiation or dedifferentiation processes.

Unbiased data-driven derivation of context-specific dynamic regulatory networks of adipocyte and osteoblast differentiation using EPIC-DREM

In order to take an unbiased and data-driven approach that can benefit from the time series profiles, we have developed a new method to predict condition-specific TF binding using footprint calling in H3K27ac data and TF motif annotation (Gusmao et al., 2016) (Roider et al., 2007). We previously found that using footprints works well for the prediction of gene expression using TEPIC (Schmidt et al., 2017). Our approach uses a randomization strategy, which accurately accounts for differences in footprint lengths and GC-content bias, to assess the significance of TF binding affinity values for each condition or time point (Figure 2A; see Methods for more details). These time point-specific predictions can be combined with the DREM approach (Schulz et al., 2012) to construct lineage-specific networks that are supported by epigenetics data (called EPIC-DREM, Figure 2A).

At first we have used 78 TF ChIP-seq datasets from three ENCODE cell lines (GM12878, HepG2, and K562) to test the ability of our approach to prioritize condition-specific TF binding sites using this pipeline. Using a p-value cutoff < 0.05 we obtained accurate cell-type specific TF binding predictions with a median precision of ~70% without a major decrease in recall (Supplementary Figure S2) and, thus, used the same cut-off for our time-series differentiation data.

We used EPIC-DREM to analyse all timepoints of the two differentiation time series of adipogenesis and osteoblastogenesis. Depending on the time point and lineage we predicted 0.6 to 1.4 million footprints per time point, consistent with previous reports on the presence of approximately 1.1 million DNase-seq footprints per cell type (Neph et al., 2012). These footprints were annotated with TFs that are expressed during the differentiation and were associated to target genes within 50 kb of their most 5'TSSs to obtain the TF scores per gene and per time point (see Methods and Figure 2). The full matrix of the time point-specific TF-target gene interactions per lineage can be downloaded at doi:10.5061/dryad.r32t3.

The derived matrix of the predicted time point-specific TF-target gene interactions was combined with the time series gene expression data to serve as input for DREM to identify bifurcation events, where genes split into paths of co-expressed genes (Figures 2 and 3). Knowing the time point-specific TF-target gene interactions allows an accurate association of split points and paths with the key TFs regulating them. To directly test the accuracy and the biological relevance of the EPIC-DREM predictions compared to alternative methods, we performed the same analysis of the time-series expression data also with three alternative approaches for the prediction of TF-gene interactions: 1) based on DREM 2.0, using existing, time-point unspecific ChIP-seq datasets of TF binding sites; 2) DREM-TRAP, where the TF binding site predictions are computed in 2 kb windows centered at the gene TSSs without the time- and context-specific epigenomic profiles, and; 3) Random shuffling of the EPIC-DREM TF-gene interaction matrix (Figures 2B and 2C). To address the biological relevance of these predictions, we considered the top 15 TFs with best split scores from each split at day 0 and combined them into lists of potential master regulators identified for both lineages and each prediction method (Figure 2B, Supplementary Table S2). Next

we performed a literature search for each of the predicted TFs on these lists to see whether they have already been associated with adipogenesis or osteoblastogenesis (Supplementary Table S2). From the randomly assigned TFs only 20-30% had been previously found to be associated with the two cell types, while for the TFs predicted by the time-point unspecific methods DREM 2.0 or DREM-TRAP this fraction increased to between 53% and 59%, depending on the lineage (Figure 2B). However, applying EPIC-DREM for the prediction further improved the fraction of biologically relevant known TFs to as high as 92% for adipocyte and 74% for osteoblast differentiation, respectively. Consistently, the DREM enrichment scores obtained by EPIC-DREM were overall higher than those obtained by the other methods (Figure 2C), indicating that the TFs identified by EPIC-DREM can much better explain the observed expression dynamics.

DREM clusters co-expressed genes along the time-points and identifies split points with transcriptional regulators assigned to them. Figure 3 shows the split points and the paths of co-expressed genes identified for adipocyte (Figure 3A) and osteoblast (Figure 3B) differentiation. The entire predicted gene regulatory networks consist of up to tens of thousands of nodes and edges, preventing their illustration in an intuitive fashion. However, for the selected DREM paths the TF-TF-networks of the top 25 TFs (based on their split score) are illustrated in Figure 3, with the size of the TF nodes corresponding to the total number of time point-specific predicted target genes across the genome. The full list of TFs controlling the individual paths are provided in Supplementary Table S3 and the full matrix of the gene regulatory networks can be downloaded at doi:10.5061/dryad.r32t3.

Inspection of the TF-TF networks of the identified top TFs confirmed many known positive (e.g. KLF5 on day 0, CEBPA on D1, and TGIF2 on D3) and negative (e.g.

HES1 and NR4A3 on day 0, FOXC1 on D1) regulators of adipocyte differentiation (Chao et al., 2008; Farmer, 2006; Horie et al., 2008; Omatsu et al., 2014) (Figure 3A), consistently with the above-mentioned findings that most of the top TFs have been associated with adipogenesis in the literature (Supplementary Table S2). Focusing on the size of the individual nodes in the TF-TF networks directly reveals that TFs controlling the first time point directly after differentiation initiation have the highest number of predicted target genes, with some TFs such as HES1 having more that 10 000 predicted target genes (Figure 3A). Similarly, the results of the regulatory network of osteoblastogenesis confirms several known positive regulators such as the aforementioned HES1, TEAD2, and BHLHE40 (Hakelien et al., 2014; Hassan et al., 2004; Iwata et al., 2006), while revealing many other factors that have not been previously associated to osteoblast differentiation (Figure 3B).

Curiously, EPIC-DREM analysis did not highlight TFs such as PPARG among the top 25 TFs at any bifurcation point, despite PPARG being a well-established master regulator of adipocyte differentiation (Tontonoz & Spiegelman, 2008). To better understand this result and to further explore capabilities of EPIC-DREM we had a detailed look at the predictions related to PPARG and the PPARG:RXRA heterodimer. Indeed, both the PPARG monomer and the heterodimer with RXRA appear among the predicted regulators at most split points (see Supplementary Table S3), in particular for the paths of genes upregulated in adipogenesis. The highest ranking of PPARG is obtained for the genes induced directly after differentiation initiation on day 0, where it is ranked as TF number 34 based on the split score. To better illustrate PPARG regulation in yellow adipocyte differentation, we generated a PPARG-centric GRN on day 3 of differentiation (Supplementary Figure S3), the time point where the highest induction of *Pparg* mRNA takes place (Supplementary Figure

S1). Supplementary Figure S3 shows all the TFs predicted to control *Pparg* expression at this time point and the top 200 targets out of the total of 3405 genes predicted to be regulated by PPARG:RXRA (Supplementary Table S4). To see whether these predicted targets are consistent with known direct target genes, we used results from Lefterova *et al*. who experimentally identified PPARG targets as genes with at least 3-fold change in their expression in differentiated adipocytes and with binding of PPARG within 50 kb from the TSS based on ChIP-seq data (Lefterova et al., 2008). Importantly, our predicted PPARG:RXRA targets were significantly enriched for these experimentally validated targets (hypergeometric test p=2.78e-31).

The fact that PPARG is missing from the lists of the very top TFs explaining the gene expression dynamics during adipogenesis could be due to its modest induction during MSC differentiation (Supplementary Figure S1). Indeed, *Pparg* was noted already in earlier work to be weaker induced in bone marrow derived MSCs than in the 3T3-L1 cells that are more commonly used to study white adipocyte differentiation (Cawthorn et al., 2012). In keeping with these observations, analysis of the histone modifications associated with active transcription at the *Pparg* locus in published data from 3T3-L1 cells confirmed the appearance of strong enhancers (based on H3K27ac signal), increased transcription (H3K36me3), and activation of both alternative transcription start sites (H3K4me3) after differentiation initiation (Supplementary Figure S4; (Mikkelsen et al., 2010)). In contrast, in MSCs *Pparg* transcription was abundant already prior to differentiation (H3K36me3), with only one TSS marked by H3K4me3 (transcript variant 1) and strong enhancer regions active across the locus (H3K27ac) (Supplementary Figure S4A). Moreover, while also SE regions could be identified at the *Pparg* locus, consistently with previous work (Hnisz et al., 2013), these did not change significantly during MSC differentiation. Thus, our data confirms PPARG as

an important SE-controlled regulator with thousands of putative targets in MSCs and the bone marrow adipocytes, albeit with only modest changes during differentiation.

One of our main aims in analyzing two parallel lineages was the identification of shared regulators that could control the differentiation or dedifferentiation of both cell types. With this aim in mind, we combined the predictions from all bifurcation points per time point to generate separate lists of top 20 TFs with most targets for each time point in both lineages (Figure 3C, see Methods). Next the top 20 TFs at corresponding time points in both lineages were overlapped to uncover the extent and identity of the shared top TFs. As might be expected, the highest level of overlap between the regulators was observed at early differentiation with 9 out of 20 top TFs being same for both lineages (Figure 3C). These are likely to include TFs important for maintaining the multipotent state of MSCs and thus negative regulators of both differentiation processes. Indeed, TFs like the AHR:ARNT heterodimer, E2F4, GLIS1, HIF1A, and TEAD2 have already been associated with gene regulation in different stem cell types (Forristal et al., 2010; Gialitakis et al., 2017; S. Y. Lee et al., 2017; Lv et al., 2017; Maekawa et al., 2011; Singh et al., 2011; Tamm et al., 2011). Interestingly, at day 1 of differentiation the number of shared TFs decreases to four and at later time points the only shared factor is SETDB1, better known as a co-regulator (Frietze et al., 2010). The only TF to appear as shared regulator at two different time points is FOXN1, which is highly connected in GRNs of both lineages, strongly downregulated in adipogenesis but highly expressed in osteoblasts, suggesting FOXN1 might play opposite roles in the two lineages (Figure 3 and Supplementary Table S1).

Taken together, the EPIC-DREM approach can identify many known key regulators of adipocyte and osteoblast differentiation and predicts additional novel regulators

and the bifurcation events they control in an entirely unbiased manner relying only on the available time series data.

Identification of dynamic SEs in adipocyte and osteoblast differentiation

While EPIC-DREM can efficiently identify many of the main regulators of the differentiation time courses, it still yields tens of TFs predicted to control thousands of target genes, most of which are likely to contribute to the cellular phenotype. However, testing tens of different TFs is often not practical. In order to further prioritize the identified main regulators, we hypothesized that the key TFs of the differentiation processes would be under high regulatory load and controlled by SEs with dynamic profiles. To obtain such profiles of SEs across time points, we first identified all SEs with a width of at least 10 kb separately in each of the 10 H3K27ac ChIP-seq data sets from the two time courses. Next, to allow a quantification of the SE signals across time, while also accounting for the changes in the width of the SEs, we combined all SEs from all time points with at least 1 nucleotide overlap into one broader genomic region, called a merged SE. Figure 4A illustrates how the 49 SEs found at the *Cxcl12* locus, producing a chemokine essential for maintenance of the hematopoietic stem cells (HSCs) (Ambrosi et al., 2017), can be combined into one exceptionally large merged SE region that enables the quantification of the SE signal across time. The normalized read counts in the merged SE region were collected and normalized relative to D0 (Figure 4B). To confirm that the obtained profile is a reasonable estimate of the SE activity and to see whether it could indeed be regulating the *Cxcl12* gene expression, we compared the SE signal profiles to the *Cxcl12* mRNA profiles. As shown in Figure 4B, in both lineages the SE signal closely followed the mRNA expression profile with Pearson correlation coefficients of 0.83 and 0.89,

respectively. Moreover, similar correlations could be seen for most other identified merged SEs, further confirming the applicability of our approach.

In total, we identified 1052 merged SEs across the two lineages (Supplementary Table S5). 120 and 79 merged SEs showed a dynamic profile ($\log_2 FC \geq 1$ in at least one time point) in adipocyte and osteoblast differentiation, respectively. Of these, 25 dynamic SEs showed concordant changes in both differentiation processes (Figure 4C-4E; Supplementary Table S6). Consistent with the different dynamics in transcriptomic changes (Figure 1), the adipocyte SEs could be divided into four separate main profiles based on their dynamics (Figure 4C; Supplementary Figure S5) while most osteoblast SEs could be assigned into one of two simple profiles that either increase or decrease in signal over time (Figure 4D).

In order to identify the genes regulated by these dynamic SEs we calculated the Pearson correlation for all of the dynamic SEs and all of their putative target genes with their TSS located within +/- 500 kb, and associated the SEs to the genes with the highest correlation coefficient (Supplementary Table S6). Out of the 151 genes we found to be associated with the dynamic SEs within the defined distance, 12 were known regulatory genes, classified either as TFs (Heinaniemi et al., 2013) or microRNA genes (Figure 4E). From the 12 regulatory genes, 7 were associated with SEs dynamic only in adipogenesis, namely, *Creb5*, *Hoxc13*, *Hsf2*, *Trp63*, *Trps1*, *Zhx3*, and *Foxn1*. From these FOXN1 was already highlighted among the top TFs by EPIC-DREM analysis, while HOXC13 and HSF2 were also predicted to regulate genes induced during adipogenesis (Supplementary Table S3). The regulatory genes associated with SEs dynamic only in osteoblastogenesis were *Nfe2l1* and *miR-322/503*. From these NFE2L1 was induced and, together with its dimerization partner MAFG, predicted by EPIC-DREM as positive regulator of both osteoblast and

adipocyte differentiation (Figure 3, Supplementary Table S3), while miR-322 and miR-503 have been shown to enhance osteogenesis (Gamez et al., 2013; Sun et al., 2017).

Finally, three of the regulatory genes, *aryl hydrocarbon receptor* (*Ahr*), *GLIS family zinc finger 1* (*Glis1*), and *Homeobox A10* (*Hoxa10*), were associated with SEs dynamic in both lineages (Figure 4E). Interestingly, all three TFs are predicted as regulators of MSC differentiation by EPIC-DREM (Supplementary Table S3), and both AHR and GLIS1 belong to the top TFs found to be shared between the two lineages (Figure 3C). AHR, the AHR:ARNT heterodimer, and GLIS1 belong to the TF-TF network controlling genes in early adipogenesis and depicted in Figure 3A. In addition GLIS1 is also included in the TF-TF network controlling genes induced on day 3 of adipogenesis. Similarly, AHR and AHR:ARNT are found in the TF-TF network controlling genes after day 1 of osteoblastogenesis (Figure 3B). Moreover, *Ahr* is associated with four separate merged SEs, more than any other TF in our analysis, and predicted to function upstream of GLIS1 in MSCs (Figure 3A and 4E). Therefore, we next focused on *Ahr* and *Glis1* regulation.

***Ahr* and *Glis1* are controlled by SEs in MSCs and repressed with lineage-specific dynamics**

The four adjacent merged SEs (SE_{283}, SE_{284}, SE_{285}, and SE_{286}) at the *Ahr* locus together cover a continuous region over 300 kb of active enhancer signal downstream of the *Ahr* gene in the MSCs (Figure 5A and 5B). SE_{831} covers a region of 29 kb in the 3'end of the *Glis1* gene in the same cells (Figure 6A and 6B). All four SE regions at the *Ahr* locus follow similar dynamics across the time points (Figure 5A-5D)

although SE_{283} was identified as being dynamic only in adipogenesis due the FC cut-off in the initial analysis in Figure 4. Moreover, all four SEs showed a very high correlation ($r \geq 0.95$) with *Ahr* mRNA levels as measured by RNA-seq (Figure 5C-5D, upper panels) and validated by RT-qPCR (Figure 5C-5D, lower panels). In adipogenesis *Ahr* expression is repressed already by D1 and remains repressed throughout the differentiation, while the signal from all the SE regions is also becoming reduced already after 1 day. In contrast in osteoblasts, the SE regions first show an increase in signal on D1, followed by gradual reduction from D3 onwards, consistent with the *Ahr* mRNA levels.

Similarly, the *Glis1* expression changes show high correlation ($r \geq 0.95$) with the signal of SE_{831} in both lineages and with both RNA-seq and RT-qPCR (Figure 6C-6D). While in osteoblast differentiation *Glis1* is gradually decreased, in adipocyte differentiation a more dynamic pattern of downregulation is observed with a temporary induction on day 5.

Finally, for both genes and in both lineages the repression is accompanied by a decreased signal of H3K36me3 in the gene bodies of the two genes, confirming their repression at the transcriptional level (Figure 5A-B and 6A-B).

While the newly identified SE clusters at the *Ahr* locus are also flanked by another active gene, *Snx13*, it is unlikely that the SEs contribute to its regulation; during the two differentiations *Snx13* shows very few changes in its expression or H3K36me3 signal (Figure 5A-5B and data not shown). Moreover, inspection of Hi-C data surrounding the *Ahr* locus in different mouse cell types indicates that *Ahr* and the four SEs are located in their own topological domain (TAD) separate from the *Snx13* gene (Supplementary Figure S6) (Tremethick, 2007). Similarly, SE_{831} could be confirmed

to be located in the same TAD with *Glis1* and the neighboring *Dmrtb1* gene, which is silenced in MSCs, suggesting that *Glis1* is the main target of the SE.

Interestingly, the SEs controlling *Ahr* and *Glis1* appear to be specific for the multipotent MSCs as only weak or no H3K27ac signal could be detected in the corresponding genomic region in mouse 3T3-L1 pre-adipocytes that are more committed towards the white adipocyte lineage (Figure 5E and 6E, (Mikkelsen et al., 2010)). Importantly, the large SE domains downstream of the *Ahr* gene and the enhancer signals at the 3'end of *Glis1* could be identified also in human MSCs, but not in other inspected human cell types, suggesting that the complex regulation of *Ahr* and *Glis1* expression in these multipotent cells could be conserved and relevant also in human development (Supplementary Figure S7).

AHR and GLIS1 regulate mesenchymal multipotency through repression of lineage-specific genes

Based on the above results, we reasoned that *Ahr* and *Glis1* could play important roles in maintaining MSCs in a multipotent state. Indeed, previous work has separately shown that both adipogenesis and osteoblastogenesis can be inhibited by toxic compounds like dioxin that are xenobiotic ligands of AHR (Alexander et al., 1998), (Naruse et al., 2002), (Korkalainen et al., 2009), while GLIS1 has been shown to promote reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs) (Maekawa et al., 2011).

To get the first idea whether the two TFs could influence the multipotent state of the MSCs, we performed a knock-down (KD) of AHR and GLIS1 in the MSCs and tested the expression of lineage-specific marker genes used earlier to confirm MSC

differentiation (Figure 7A, Supplementary Figure S8A-C and S1). For AHR a KD of approximately 50% could be confirmed at both mRNA and protein level, while Glis1 reduction was around 30% and could not be confirmed at the protein level due to lack of a specific antibody. Still, GLIS1-KD led to a modest but significant induction of *Cebpa*, *Lpl* and *Bglap* expression, while AHR-KD only affected one of the marker genes (*Lpl*). Thus, both AHR and GLIS1 might contribute to maintenance of the MSC state.

While AHR is best known for its role as a xenobiotic receptor, it has been recently suggested that it could also play a role in stem cell maintenance in HSCs under the control of endogenous ligands (Gasiewicz et al., 2014). Moreover, since GLIS1-KD was not more robust and could not yet be confirmed on the protein level, we next focused on the AHR-KD. To directly test whether the endogenous activity of AHR is important for the maintenance of the appropriate transcriptional program of the MSCs, we knocked down AHR in undifferentiated cells, and those differentiated for one day towards either lineage (Figure 7B, Supplementary Figure S8B). Two days following the KD total RNA was extracted and subjected to RNA-seq analysis.

The number of genes affected by the AHR-KD (FDR$\leq$0.1) compared to the respective control siRNA transfections was greatly dependent on the cellular condition and ranged from 614 genes in the undifferentiated cells to 722 and 1819 genes in the one day differentiated osteoblasts and adipocytes, respectively (Figure 7C, Supplementary Table S7). The higher extent of genes affected in the differentiated cells is well in keeping with the number of genes normally changing in the early osteoblasto- or adipogenesis, with adipogenesis associated with more changes (Figure 1). This suggests that early changes induced by the knock-down in the undifferentiated cells are amplified in the differentiating cells.

At first, we took advantage of these context-specific KD data to ask whether the EPIC-DREM predicted primary AHR targets at the corresponding time points were indeed affected by depletion of AHR. As shown in Figure 7D, at each condition the predicted AHR targets were significantly more affected by AHR-KD than all genes on average. Especially the top genes with highest affinity scores for AHR regulation on day 1 of adipogenesis were clearly shifted towards more upregulation at each condition (to the right in the cumulative distribution plot), arguing for functional relevance of the EPIC-DREM predictions and for AHR's role as transcriptional repressor. As highlighted earlier in Figure 3A, AHR was predicted by EPIC-DREM to act as a direct regulator of *Glis1* in MSCs and early adipogenesis. To test this specific prediction we performed RT-qPCR for *Glis1* in all AHR-KD samples and confirmed a significant change in *Glis1* levels upon AHR depletion in both differentiating adipocytes and osteoblasts (Figure 7E).

To better understand the function of the putative AHR target genes in the MSCs, we overlapped the differentially expressed genes from the undifferentiated cells with those identified in the two other KD experiments and obtained 266 high-confidence target genes that were affected in all three conditions (Figure 7C). Enrichment analysis for tissue-specific expression profiles of these genes in human and mouse gene atlas databases revealed smooth muscle, adipocytes, and osteoblasts as the most enriched cell types, where the AHR regulated genes are normally expressed (Figure 7F). Consistently, inspection of publicly available ChIP-seq data from various cell types revealed the other regulators of the AHR targets to include CEBPB, PPARG, and NR1H3, all of which are important regulators involved in induction of genes during adipocyte differentiation (Galhardo et al., 2014) (Nielsen et al., 2008) .

To further elucidate the role of AHR in the undifferentiated cells and the potential impact on lineage-specific genes, we asked how the deregulated genes were expressed in normally differentiated day 15 adipocytes and osteoblasts. The scatter plots in Figure 7G indicate the transcriptome changes taking place upon AHR-KD in MSCs compared to changes of the same genes in the differentiated cell types. As indicated by the color coding, approximately half of the genes affected by AHR-KD were also differentially expressed after normal differentiation (286 in adipocytes and 314 in osteoblasts). Moreover, genes that were upregulated in AHR-KD were preferentially induced in the differentiated cell types (Figure7G, upper right quartile). Therefore, AHR might serve as a guardian of the multipotent state in the undifferentiated cells, with its downregulation allowing increased expression of the lineage-specific genes.

Taken together, the above findings support a role for AHR as a regulator of lineage-specific genes that need to remain repressed in the multipotent MSCs. Among such genes induced upon AHR-KD we identified *Notch3*, a known regulator of cellular differentiation (Bray, 2016) (Figure 7G-H). In both differentiation time courses *Notch3* expression showed an anti-correlated profile compared to *Ahr*, with *Notch3* becoming induced while *Ahr* levels decreased (Figure 5 and Supplementary Figure 8D-E). The *Notch3* induction was also accompanied by increased *Notch4* levels in both lineages while the third abundantly expressed receptor, *Notch1*, was concomitantly downregulated (Supplementary Figure S8D-E). Finally, to see whether the impact of AHR on *Notch3* is mediated via GLIS1, we tested by RT-qPCR the effect of GLIS1-KD on three of the *Notch* genes in the corresponding conditions (Figure 7I). Interestingly, no effect could be seen on *Notch3* expression upon GLIS1-KD while both *Notch1* and *Notch2* were modestly affected. Thus, it appears that a

reprogramming of Notch signaling could be involved in the commitment of MSCs and this rewiring is at least partially under AHR- and GLIS1-mediated control.

DISCUSSION

The ability to obtain unbiased GRNs and to identify their key nodes for any given cell state transition in a data-driven manner is becoming increasingly relevant for regenerative and personalized medicine. Understanding such dynamic networks can be improved by obtaining genome-wide time-series data sets such as transcriptomics or epigenomics data. To seamlessly integrate such data sets we have combined time point-specific high accuracy TF binding predictions with probabilistic modeling of temporal gene expression data and applied it to our own time-series data from mesenchymal differentiation (Figure 2 and Figure 3). Similar time series data collections have previously been used to study for example hematopoiesis (Goode et al., 2016) and myeloid differentiation (Ramirez et al., 2017). However, the derived dynamic GRNs have relied on experimentally identified TF binding sites covering only a fraction of all TF-target gene interactions, or on a sub-network of selected TF-TF interactions, respectively.

EPIC-DREM can reveal the key TFs controlling co-expressed gene sets of interest. Still, consistent with the co-operative nature of TF activity (Siersbaek et al., 2014), the number of putative master regulators is often very large. Recent work elucidating the role of SEs in controlling cell type-specific master regulators has provided researchers with a new tool for data-driven identification of such regulators (Hnisz et al., 2013) (Parker et al., 2013). We hypothesized that merged SEs with dynamic behavior during differentiation based on their H3K27ac signal would allow finding

the genes, including the TFs, most relevant for the dynamic process. Indeed, quantification of merged SEs shows high correlation with expression levels of their target genes over time, both validating the approach and allowing for more accurate association of SEs to their target genes (Figure 4). A recent study applied a similar strategy for SE quantification and further showed that SE dynamics, as measured by MED1 occupancy, were predictive of enhancer looping to target genes, and highlighted H3K27ac as the histone modification that best predicted such loop dynamics, further supporting the validity of our approach (Siersbaek et al., 2017).

Combining EPIC-DREM results and dynamic SE profiling points towards several TFs with potentially significant roles in both lineages, including *Foxn1*, *Ahr* and *Glis1* (Figure 4). In addition, dynamic SE profiling supports a role for *Hoxa10*, although in EPIC-DREM analysis it did not make it among the top TFs. Both *Ahr* and *Glis1*, and their SEs, show an overall reduction in signal during both differentiations, although with differential and lineage-specific dynamics (Figure 5 and 6), suggesting that they could play an important role in maintaining MSCs. On the contrary, the SE at the *Foxn1* locus shows a decrease in its signal in adipocyte differentiation (data not shown), while in osteoblast differentiation the signal is further increased. Thus suggesting a role as a positive driver of osteoblast differentiation, while potentially having a negative role in adipocytes. Interestingly, FOXN1 has not been previously shown to be involved in MSC differentiation and our results warrant a further investigation of FOXN1 function in adipocyte and osteoblast differentiation.

Also GLIS1 has not been functionally associated to adipocyte or osteoblast differentiation although recent work has implicated it as differentially expressed in brown adipocyte differentiation (Pradhan et al., 2017). However, consistent with a potential role in the multipotent progenitors, GLIS1 has been shown capable of

promoting reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs) (Maekawa et al., 2011). Interestingly, our analysis for enhancer signals at the Glis1 locus in different human cell types confirmed the presence of active enhancers also in human MSCs but not in embryonic stem cells (ESCs) (Supplementary Figure S7). Thus, the ability of GLIS1 to promote cellular reprogramming of iPSCs might reflect its endogenous functions in multipotent stem cells like MSCs, rather than the pluripotent ESCs. Our results implicate GLIS1 as a regulator of MSCs functioning downstream of AHR with its reduction influencing several lineage-specific genes and some of the *Notch* family genes (Figure 7). In addition, both network analysis and the expression dynamics suggest a role for GLIS1 during adipogenesis (Figure 3A and 6). Further work will be needed to elucidate the role of GLIS1 in MSCs and adipocyte differentiation.

Unlike for GLIS1 and FOXN1, previous work has already linked AHR separately to inhibition of both adipocyte and osteoblast differentiation through studies on biological impact of dioxin, an environmental toxin capable of activating AHR (Alexander et al., 1998), (Naruse et al., 2002), (Korkalainen et al., 2009). In 3T3-L1 adipocytes this inhibition is known to be mediated through overexpression of *Ahr* in a dioxin-independent manner (Shimba et al., 2001), while increased levels of *Ahr* expression in MSCs in rheumatoid arthritis are inhibitory of osteogenesis (Tong et al., 2017). Still, the biological function of AHR in the MSCs has remained unclear while maintenance of HSCs, located in the same niche of bone marrow with MSCs, has been suggested to depend on the normal function of AHR (Gasiewicz et al., 2014). Moreover, as HSC maintenance also depends on the cytokines and chemokines provided by the MSCs, AHR is likely to impact HSCs through its gene regulatory

functions in both MSCs and HSCs (Ambrosi et al., 2017) (Boitano et al., 2010) (Jensen et al., 2003).

Here we show that repression of *Ahr* in MSC differentiation happens with lineage-specific dynamics and is accompanied by similar reduction in signal of an exceptionally large SE cluster downstream of *Ahr*. Together *Ahr* and the SEs form their own TAD in mouse cells and in human cells the SE signal is specific for MSCs. Analyzing the contributions of the different constituents of the *Ahr*-SEs will be important for understanding which pathways converge to regulate *Ahr* in MSCs and whether they function in a synergistic manner as suggested for some other large SEs (Hnisz et al., 2015) (Dukler et al., 2016).

A detailed characterization and validation of the function of AHR, and other candidate TFs, in MSCs will require the generation of a knock-out system where the TF can be fully depleted in the multipotent cells in an inducible manner. Nevertheless, a reduction of AHR expression in the undifferentiated and early stage differentiated cells already confirmed many of the EPIC-DREM predictions and also revealed an enrichment for lineage-specific genes among AHR targets, including *Notch3* (Figure 6). Notch signaling has been implicated in numerous developmental processes with highly diverse outcomes (Bray, 2016) and also in MSCs *Notch3* regulation is accompanied by changes in other *Notch* genes (Supplementary Figure S8). Interestingly, AHR has been recently linked to regulation of Notch signaling in mouse lymphoid cells and testis (J. S. Lee et al., 2011) (Huang et al., 2016). However, the affected Notch receptors varied depending on the cell type in question. *Notch* genes show different expression profiles across tissues and cell types and *Ahr*-mediated regulation of Notch signaling could be context-specific depending on the prevailing GRN or chromatin landscape. Curiously, *Notch3* has a cell type-selective expression

profile, favouring mesenchymal tissues like bone, muscle and adipose tissue (C. Wu et al., 2016).

Our approach for identification of the dynamic GRNs and SEs allows key regulator identification in various time series experiments involving cell state changes. Our current results together with previous data identify AHR and GLIS1 as likely guardians of mesenchymal multipotency, implicate additional TFs as novel regulators of adipocyte and osteoblast differentiation, and provide an extensive resource for further analyses of mesenchymal lineage commitment.

METHODS

Cell culture

The mouse MSC line ST2, established from Whitlock-Witte type long-term bone marrow culture of BC8 mice (Ogawa et al., 1988), was used during all experiments. Cells were grown in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco, Life Technologies, 32404014) supplemented with 10% fetal bovine serum (FBS) (Gibco, Life Technologies, 10270-106, lot #41F8430K) and 1% L-Glutamine (Lonza, BE17-605E) in a constant atmosphere of 37°C and 5 % CO_2. For differentiation into adipocytes and osteoblasts, ST2 cells were seeded 4 days before differentiation (day-4), reached 100% confluency after 48 hours (day-2) and were further maintained for 48 hours post-confluency (day 0). Adipogenic differentiation was subsequently initiated on day 0 (D0) by adding differentiation medium I consisting of growth medium, 0.5 mM isobutylmethylxanthine (IBMX) (Sigma-Aldrich, I5879), 0.25 µM dexamethasone (DEXA) (Sigma-Aldrich, D4902) and 5 µg/mL insulin (Sigma-Aldrich, I9278). From day 2 (D2) on differentiation medium II consisting of growth medium, 500 nM rosiglitazone (RGZ) (Sigma-Aldrich, R2408) and 5 µg/mL insulin

(Sigma-Aldrich, I9278) was added and replaced every 2 days until 15 days of differentiation. Osteoblastic differentiation was induced with growth medium supplemented with 100 ng/mL bone morphogenetic protein-4 (BMP-4) (PeproTech, 315-27). Same media was replaced every 2 days until 15 days of osteoblastogenesis.

Gene silencing

Undifferentiated ST2 cells (day-1) were transfected with Lipofectamine RNAiMAX (Life Technologies, 13778150) according to manufacturer's instructions using 50 nM of gene-specific siRNAs against mouse *Ahr* (si*Ahr*) (Dharmacon, M-044066-01-0005), *Glis1* (si*Glis1*) (Dharmacon, M-065576-01-0005) or 50 nM of a negative control siRNA duplexes (si*Control*) (Dharmacon, D-001206-14-05). Cells were collected 48 h post-transfection. Sequences of the siRNAs are listed in Supplementary Table S8.

Western blotting

After a washing step of the cells with 1x PBS, and addition of 1x Läemmli buffer, the lysates were vortexed and the supernatants were heated at 95°C for 7 minutes. Proteins were subjected to SDS-PAGE (10% gel) and probed with the respective antibodies. The following antibodies were used: anti-AHR (Enzo Life Biosciences, BML-SA210-0025), anti-ACTIN (Merck Millipore, MAB1501). HRP-conjugated secondary antibodies were purchased from Cell Signaling. Signals were detected on a Fusion FX (Vilber Lourmat) imaging platform, using an ECL solution containing 2.5 mM luminol, 100 mM Tris/HCl pH 8.8, 0.2 mM para-coumaric acid, and 2.6 mM hydrogenperoxide.

RNA extraction and cDNA synthesis

Total RNA was extracted from ST2 cells using TRIsure (Bioline, BIO-38033). Medium was aspirated and 1000 µL of TRIsure was added to 6-wells. To separate RNA from DNA and proteins, 200 µL of chloroform (Carl Roth, 6340.1) was added. To precipitate RNA from the aqueous phase, 400 µL of 100% isopropanol (Carl Roth, 6752.4) was added and RNA was incubated at -20°C overnight. cDNA synthesis was done using 1 µg of total RNA, 0.5 mM dNTPs (ThermoFisher Scientific, R0181), 2.5 µM oligo dT-primer (Eurofins MWG GmbH, Germany), 1 U/µL Ribolock RNase inhibitor (ThermoFisher Scientific, EO0381) and 1 U/µL M-MulV Reverse transcriptase (ThermoFisher Scientific, EP0352) for 1h at 37°C or 5 U/ µL RevertAid Reverse transcriptase for 1 h at 42°C. The PCR reaction was stopped by incubating samples at 70°C for 10 minutes.

Quantitative PCR

Real-time quantitative PCR (qPCR) was performed in an Applied Biosystems 7500 Fast Real-Time PCR System and using Thermo Scientific Absolute Blue qPCR SYBR Green Low ROX Mix (ThermoFisher Scientific, AB4322B). In each reaction 5 µL of cDNA, 5 µL of primer pairs (2 µM) and 10 µL of the Absolute Blue qPCR mix were used. The PCR reactions were carried out at the following conditions: 95°C for 15 minutes followed by 40 cycles of 95°C for 15 seconds, 55°C for 15 seconds and 72°C for 30 seconds. To calculate the gene expression level the $2^{-(\Delta\Delta Ct)}$ method were used where ΔΔCt is equal to $(\Delta Ct_{(\text{target gene})} - \Delta Ct_{(\text{housekeeping gene})})_{\text{tested condition}} - (\Delta Ct_{(\text{target gene})} - \Delta Ct_{(\text{housekeeping gene})})_{\text{control condition}}$. *Rpl13a* was used as a stable housekeeping gene

and D0 or si*Control* were used as control condition. Sequences of the primer pairs are listed in Supplementary Table S8.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation of histone modifications was performed on indicated time points of adipocyte and osteoblast differentiation. Cells were grown on 10 cm^2 dishes. First, chromatin was cross-linked with formaldehyde (Sigma-Aldrich, F8775-25ML) at a final concentration of 1% in the culture media for 8 minutes at room temperature. Then, the cross-linked reaction was quenched with glycine (Carl Roth, 3908.3) at a final concentration of 125 mM for 5 minutes at room temperature. The formaldehyde-glycine solution was removed and cells were washed twice with ice-cold phosphate-buffered saline (PBS) (Lonza, BE17-516F) containing cOmplete™ mini Protease Inhibitor (PI) Cocktail (Roche, 11846145001). Then, cells were lysed in 1.7 mL of ice-cold lysis buffer [5 mM 1,4-Piperazinediethanesulfonic acid (PIPES) pH 8.0 (Carl Roth, 9156.3); 85 mM potassium chloride (KCl) (PanReac AppliChem, A2939); 0.5 % 4-Nonylphenyl-polyethylene glycol (NP-40) (Fluka Biochemika, 74385)] containing PI and incubated for 30 minutes on ice. The cell lysates were then centrifuged at 660 xg for 10 min at 7°C and the pellet was resuspended in 400 µL of ice-cold shearing buffer [50 mM Tris Base pH 8.1 (Carl Roth, 4855.2); 10 mM Ethylenediamine tetraacetic acid (EDTA) (Carl Roth, CN06.3); 0.1 % Sodium Dodecylsulfate (SDS) (PanReac Applichem, A7249); 0.5 % Sodium deoxycholate (Fluka Biochemika, 30970)] containing PI. Chromatin was sheared with a sonicator (Bioruptor®Standard Diagenode, UCD-200TM-EX) during 20 cycles at high intensity (30 s off and 30 s on

) for the ST2 cells differentiated into adipocytes and osteoblasts and 25 cycles at high intensity (30 s off and 30 s on) for the ST2 differentiated into osteoblasts for 9 days on. The sheared cell lysate was then centrifuged at 20817 xg for 10 minutes at 7°C and the supernatant containing the sheared chromatin was transferred to a new tube. For each immunoprecipitation 10 µg (for H3K4me3) or 15 µg (for H3K27ac and H3K36me3) of sheared chromatin and 4 µg as input were used. The sheared chromatin was diluted 1:10 with modified RIPA buffer [140 mM NaCl (Carl Roth, 3957.2); 10 mM Tris pH 7.5 (Carl Roth, 4855.2); 1 mM EDTA (Carl Roth, CN06.3); 0.5 mM ethylene glycol-bis(β-amino-ethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) (Carl Roth, 3054.3); 1 % Triton X-100 (Carl Roth, 3051.2); 0.01 % SDS (PanReac Applichem, A7249); 0.1 % sodium deoxycholate (Fluka Biochemika, 30970)] containing PI. The diluted sheared chromatin was incubated overnight with the recommended amount provided by the manufacturer of an antibody against H3K4me3 (Millipore, 17-614), 5 µg of an antibody against H3K27ac (Abcam, ab4729) or 5 µg of an antibody against H3K36me3 (Abcam, ab9050). The next day, the antibodies were captured using 25 µL of PureProteome™ Protein A Magnetic (PAM) Bead System (Millipore, LSKMAGA10) for 2 hours at 4°C on a rotating wheel. After, the PAM beads were captured using a DynaMag™-2 magnetic stand (Life Technologies, 12321D). The supernatant was discarded and the PAM beads were washed twice with 800 µL of Immunoprecipitation wash buffer 1 (IPWB1) [20 mM Tris, pH 8.1 (Carl Roth, 4855.2); 50 mM NaCl (Carl Roth, 3957.2); 2 mM EDTA (Carl Roth, CN06.3); 1 % Triton X-100 (Carl Roth, 3051.2); 0.1 % SDS (PanReac Applichem, A7249)], once with 800 µL of Immunoprecipitation wash buffer 2 (IPWB2) [10 mM Tris, pH 8.1 (Carl Roth, 4855.2); 150 mM NaCl (Carl Roth, 3957.2); 1 mM EDTA (Carl Roth, CN06.3), 1 % NP-40 (Fluka Biochemika, 74385), 1 % sodium deoxycholate (Fluka

Biochemika, 30970), 250 mM of lithium chloride (LiCl) (Carl Roth, 3739.1)], and twice with 800 µL of Tris-EDTA (TE) buffer [10 mM Tris, pH 8.1 (Carl Roth, 4855.2); 1 mM EDTA (Carl Roth, CN06.3), pH 8.0]. Finally, the PAM beads and the inputs were incubated with 100 µL of ChIP elution buffer [0.1 M sodium bicarbonate ($NaHCO_3$) (Sigma-Aldrich, S5761); 1 % SDS (PanReac Applichem, A7249)]. The cross-linking was reversed by adding 10 µg of RNase A (ThermoFisher, EN0531) and 20 µg of proteinase K (ThermoFisher, EO0491) at 65°C overnight. Then, the eluted chromatin was purified using a MinElute Reaction Cleanup Kit (Qiagen, 28206) according to the manufacturer's instructions. The DNA concentration was measured using the Qubit® dsDNA HS Assay Kit (ThermoFisher, Q32851) and the Qubit 1.0 fluorometer (Invitrogen, Q32857) according to the manufacturer's instructions.

ChIP-Seq

The sequencing of the ChIP samples was done at the Genomics Core Facility in EMBL Heidelberg, Germany. For sequencing, single-end and unstranded reads were used and the samples were processed in an Illumina CBot and sequenced in an Illumina HiSeq 2000 machine. In total, 979 572 918 raw reads were obtained. Raw reads quality was assessed by fastqc [v0.11, (Ruvkun & Hobert, 1998)]. This quality control unveiled that some reads were containing part of the adapters. Those spurious sequences were cleaned up from the genuine mouse sequences by AdapterRemoval (Lindgreen, 2012) [v1.5]. The PALEOMIX pipeline (Schubert et al., 2014) [v1.0.1] was used for all steps from FASTQ files to BAM files including trimming, mapping, and duplicate marking. This workflow ensures that all files are complete and valid. Retained reads were required to have a minimum length of 25 bp. Bases with unreliable Phred scores (0-2) were trimmed out. In total 31 909 435 reads were discarded (3.26%). Eventually, 947 663 483 reads were retained (96.74%). Trimmed

reads were further mapped using BWA (Li & Durbin, 2009) [v0.7.10] with the backtrack algorithm dedicated to short sequences. The mouse reference was the mouse genome GRCm38.p3 (mm10, patch 3) downloaded from NCBI. For validating, merging BAM files, and marking duplicates, we used the suite tool Picard [v1.119, (Adams et al., 2000)]. Duplicates were marked but not removed. Only reads with a mapping quality of 30 were retained to ensure a unique location on the genome resulting in 661 364 143 reads (69.79% of the trimmed reads). The samples with a coverage of less than 8 million reads (mapping quality > 30) were excluded from the downstream analysis. Raw FASTQ and BAM files have been deposited in the European Nucleotide Archive with the accession number PRJEB20933.

The ChIP-Seq peaks were called with Model-based analysis of ChIP-Seq (Zhang et al., 2008) (MACS) version 2.1.0 for H3K4me3, with HOMER (Heinz et al., 2010) for H3K27ac, and with SICER (Zang et al., 2009) version 1.1 for H3K36me3, using input from undifferentiated ST2 cells as control for IPs from D0 cells and input from D5 adipocyte- or osteoblast-differentiated cells for the IPs from the respectively differentiated cells.

RNA-Seq

The sequencing of the time course samples was done at the Genomics Core Facility in EMBL Heidelberg, Germany. For sequencing, single-end and unstranded reads were used and the samples were processed in an Illumina CBot and sequenced in an Illumina NextSeq machine.

The sequencing of the AHR knock-down samples was performed at the Luxembourg Center for Systems Biomedicine (LCSB) Sequencing Facility. The TruSeq Stranded mRNA Library Prep kit (Illumina) was used to prepare the library for sequencing with

1 µg of RNA as starting material according to manufacturer's protocol. The library quality was checked using an Agilent 2100 Bioanalyzer and quantified using Qubit dsDNA HS assay Kit. The libraries were then adjusted to 4 nM and sequenced on a NextSeq 500 (Illumina) according to the manufacturer's instructions.

The obtained reads were quality checked using FastQC version 0.11.3 (Ruvkun & Hobert, 1998). Cutadapt version 1.8.1 (Martin, 2011) was used to trim low quality reads (`-q 30` parameter), remove Illumina adapters (`-a` parameter), remove reads shorter than 20 bases (`-m 20` parameter) with an error tolerance of 10% (`-e 0.1` parameter). Then, removal of reads mapping to rRNA species was performed using SortMeRNA (Kopylova et al., 2012) with the parameters `--other`, `--log`, `-a`, `-v`, `--fastx` enabled. Lastly, the reads were quality checked using FastQC version 0.11.3 to control whether bias could have been introduced after the removal of Illumina adapters, low quality reads and rRNA reads. Then, the reads were mapped to the mouse genome mm10 (GRCm38.p3) and using the gene annotation downloaded from Ensembl (release 79) using the Spliced Transcripts Alignment to a Reference (Dobin et al., 2013) (STAR) version 2.5.2b using the previously described parameters (Baruzzo et al., 2017) . The reads were counted using the function *featureCounts* from the R package *Rsubread* (Liao et al., 2014) version 1.4.6-p3 and the statistical analysis was performed using DESeq2 (Love et al., 2014) version 1.14.1 in R 3.3.2 and RStudio (RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA).

EPIC-DREM analysis

To identify TFs that have a regulatory function over time, we designed a new computational workflow that combines the computational TF prediction method

TEPIC (Schmidt et al., 2017) with DREM (Schulz et al., 2012), a tool to analyze the dynamics of transcriptional regulation.

We identified TF footprints in the H3K27ac signal using HINT-BC (Gusmao et al., 2016), which is included in the Regulatory Genomics Toolbox, version 0.9.9. Next, we predicted TF binding in those footprints using TEPIC, version 2.0. We used the provided set of 687 PWMs for *Mus musculus* and mouse genome version mm10 (GRCm38) to predict TF affinities using TRAP (Roider et al. 2007) within TEPIC. As DREM requires a time point-specific prediction of binding of a regulator with its target, we needed to develop an approach to determine a suitable TF-specific affinity cut-off, for each time point. For this, we created a similar set of random regions that mirrors the GC content and length distribution of the original sequences of the footprints. TF affinities a_r calculated in the random regions are used to determine a suitable cut-off for the original affinities a_o using the frequency distribution of the TF affinities. Affinities for TF i are denoted by a_{ri} and a_{oi}. Let $r \in R$ denote a randomly chosen genomic region that is screened for TF binding, and let $|r|$ denote its length. Analogously, let $o \in O$ denote a footprint that is screened for TF binding, and let $|o|$ denote its length. We normalize both a_{ri} and a_{oi} by the length of their corresponding region and obtain the normalized TF affinities a'_{ri} and a'_{oi}:

$$a'_{ri} = \frac{a_{ri}}{|r|},\ a'_{oi} = \frac{a_{oi}}{|r|}.$$

Using the distribution of a'_{ri}, values we derive a TF-specific affinity threshold t_i for a p-value cut-off of *0.05* (See section *ChIP-seq validation of TEPIC affinity cut-off* for how this p-value was chosen). For a TF i, we compute a binary affinity value b_{oi} from the original affinity a'_{oi} according to the cut-off t_i with:

$$b_{oi} = \begin{cases} 1, a'_{oi} > t_i \\ 0, a'_{oi} \leq t_i. \end{cases}$$

The binary affinity values b_{oi} can be used to compute a binary TF – gene association a_{gi} between gene g and TF i:

$a_{gi} = \begin{cases} 1, & \exists\, o \in O_{gw}: a'_{oi} = 1 \\ 0, & otherwise \end{cases}$, where O_{gw} denotes all footprint regions that occur within a window of size w around the TSS of gene g.

Informally, a gene g is associated to TF i if there is a predicted binding site within a window of predefined size w around the gene's TSS. Here, we use $w = 50kb$.

Together with gene expression estimates, the TF– gene associations can be directly used as input to DREM. In this analysis, we used version 2.0.3 of DREM. The entire workflow of EPIC-DREM is shown in Figure 2A. TEPIC is available online at https://github.com/SchulzLab/TEPIC, DREM can be downloaded at http://www.sb.cs.cmu.edu/drem/.

ChIP-seq validation of TEPIC affinity cut-off

To validate that the affinity threshold described above indeed results in an adequate separation between bound and unbound sites, we conducted a comparison to TF-ChIP-seq data. We obtained TF-ChIP-seq data from ENCODE for K562 (18 TFs), HepG2 (36 TFs), and GM12878 (24 TFs). In addition, we downloaded H3K27ac data for the mentioned cell lines from ENCODE. A list of all ENCODE accession numbers is provided in Supplementary Table S9. As described above, we called footprints using HINT-BC and calculated TF affinities in the footprints as well as in the randomly selected regions that map the characteristics of the footprints. To understand the influence of different thresholds, we calculated affinity thresholds for the following p-values: 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, and 0.5. All affinities below the selected affinity value are set to zero, the remaining values are set to one. The quality of the discretization is assessed through the following "peak centric"

validation scheme, as used before in (Cuellar-Partida et al., 2012). The positive set of the gold standard is comprised of all ChIP-seq peaks that contain a motif predicted by FIMO (Grant et al., 2011) , the negative set contains all remaining ChIP-seq peaks. A prediction is counting as a true positive (TP) if it overlaps the positive set, it counts as a false positive (FP) it if overlaps the negative set. The number of false negatives (FN) is the number of all entries in the positive set that are not overlapped by any prediction. For all TFs in all cell lines we calculate Precision (PR) and Recall (REC) according to

$$PR = \frac{TP}{TP+FP}, \quad REC = \frac{TP}{TP+FN}.$$

As one can see, Precision is increasing with a stricter p-value threshold, while Recall is decreasing. We found that using 0.05 seems to be a reasonable compromise between Precision and Recall. The median Precision and Recall values calculated over all cell lines and all TFs are shown in Supplementary Figure S2A. Detailed results on a selection of TFs that are present in all three cell lines are shown in Supplementary Figure S2B.

Generating TF-TF interaction networks for DREM splits

We devised a general strategy to create TF-TF interaction networks for individual DREM splits. For a split of interest, we retrieve the top 25 regulators *T*, ranked according to the DREM split score. For each regulator $t \in T$, we determine the target genes G_t by declaring each gene *g* to be a target of TF *t*, if and only if $a_{g,t} = 1$, where $a_{g,t}$ is the binary value for the TF affinity of TF *t* in the vicinity of gene *g*, as introduced above. Here, we want to show only interactions among the top regulators, thus, in order to include a directed edge from *t* to *g* in the TF-TF network, it must hold

that $g \in T$. For reasons of simplicity, we limit the number of shown interactions per TF *t* to 10, ranked by the numerical affinity values across all $g \in G_t$.
Additionally, we scale the size of each node representing a regulator $t \in T$, according to its total number of target genes $|G_t|$ in six discrete levels (<2000, <4000, <6000, <8000, <10 000, and >10 000). This allows an interpretation of the importance of the regulator outside the scope of the interactions among the top TFs for the shown split. The node colour indicates the expression change of the depicted TF compared to time-point zero. Orange refers to downregulation, whereas blue refers to upregulation. The networks are arranged with *graphviz* and the *neato* layout algorithm (Gansner et al., 2005).

DREM-TRAP and Random TF-gene assignment

To assess the impact of epigenetics data on DREM predictions, we designed a baseline approach, assuming that most transcriptional regulation for a gene is occurring in its promoter region. Hence, for each gene *g*, we consider a 2 kb window around the most 5' TSS of *g*. Within this window, we compute TF affinities as described above for 687 PWMs for *Mus musculus* and mouse genome version mm10 (GRCm38). Also, we applied the same thresholding approach as described for EPIC-DREM. Thereby, DREM-TRAP represents approaches that are purely sequence and annotation based and do not account for changes in the chromatin.
As another sanity check for EPIC-DREM, we permuted the EPIC-DREM input matrix, which is based on time-point specific footprint calls. In detail, we randomly shuffled column 1 (TF), column 2 (target gene), and column 4 (time point). Thus the number of TFs, target genes, and timepoint entries is not affected, only the mapping between them changed.

Enrichment score of DREM splits and its aggregation

DREM uses a hypergeometric distribution to assess the association of a TF to genes in a distinct path, the so called split-score, where a lower value means a stronger association. Because we carry out a considerable number of tests per split (as we test more than 600 TFs), we correct the split-scores for multiple testing using Bonferroni correction.

To compare the enrichment scores across the different versions of DREM used in this manuscript, we –log2 transformed the Bonferroni corrected split-scores and show violin plots of the transformed scores, which we call *DREM split score*. For those scores, the higher the value, the better is the association of the TFs identified at split points to genes on the adjacent paths.

To identify essential TFs across multiple splits at one time point, we perform p-value aggregation using Fisher's method on the split scores, i.e. we compute

$$\boxed{X_{t,n} = -2 \sum_{i=1}^{s} \log\left(p_{n,i,t}\right)}$$

Here, $\boxed{X_{t,n}}$ is the aggregated score for TF *t* at time point *n*, *i* is the current split, *s* indicates the number of splits at Timepoint *n* and $\boxed{p_{n,i,t}}$ is the split score for TF *t* at split *i* at time point *n*. Next, we rank the $\boxed{X_{t,n}}$ by time point n and thereby obtain a list of top regulators per time point.

Identification of dynamic merged SEs

In order to identify temporal SEs across both lineages, BedTools (Quinlan & Hall, 2010) version 2.24.0, Hypergeometric Optimization of Motif EnRichment (Heinz et al., 2010) (HOMER) version 4.7.2 and Short Time-series Expression Miner (Ernst &

Bar-Joseph, 2006) (STEM) version 1.3.8 were used. First, the coverage of individual SEs was summarized using `genomeCoverageBed` command using `-g mm10` and `-bg` parameters. Then, `unionBedGraphs` command was used to combine multiple SEs coverage into a single map such that the SEs' coverage is directly comparable across multiple samples. Finally, `mergeBed` command was used to combine SEs overlapping ≥ 1 bp into a single merged SE which spans all the combined SEs. In order to calculate the normalized read count number of merged SEs, `annotatePeaks.pl` with `-size given` and `-noann` parameters was used. Lastly, STEM was used to cluster and identify SEs temporal profiles and SEs with *Maximum_Unit_Change_in_Model_Profiles_between_Time_Points* 2 and *Minimum Absolute Expression Change* 1.0 were considered as dynamic.

For the dynamic SEs we calculated the Pearson correlation for the putative target genes with their TSS located within +/- 500 kb, and associated the SEs to the genes with the highest correlation coefficient.

Enrichment analysis

EnrichR (14.4.2017) (Chen et al., 2013; Kuleshov et al., 2016) was used to perform gene enrichment analysis.

Availability of data and materials

The datasets generated and analysed during the current study are available in the European Nucleotide Archive with the accession number PRJEB20933. All scripts and generated GRNs can be retrieved from https://github.com/sysbiolux and from https://github.com/SchulzLab/TEPIC.

ACKNOWLEDGEMENTS

We would like to thank Dr. Maria Bouvy-Liivrand for help with establishing the ST2 cell culture and differentiation and EMBL Gene Core at Heidelberg for support with high-throughput sequencing. The experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2014).

COMPETING INTERESTS

The authors declare they have no competing interests.

FUNDING

This work was supported by funding from the University of Luxembourg. DG was supported by fellowship from the Luxembourg National Research Fund (FNR) (AFR 7924045).

AUTHOR CONTRIBUTIONS

DG, TS and LS conceived the project and designed the experiments and analysis. DG performed all the experiments and DG and LS analyzed the results. DG and AG performed the RNA-seq and ChIP-seq analysis. FS and MHS developed the EPIC-DREM approach. DG, FS and MHS performed the EPIC-DREM analysis. MS and DG performed the Western blotting. RH prepared the libraries and performed the sequencing for the AHR-KD experiments. PE developed the randomization method to derive control footprint regions. All authors commented on the manuscript.

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L.,

Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., & Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science, 287*(5461), 2185-2195. doi:10.1126/science.287.5461.2185

Aguirre, A., Montserrat, N., Zacchigna, S., Nivet, E., Hishida, T., Krause, M. N., Kurian, L., Ocampo, A., Vazquez-Ferrer, E., Rodriguez-Esteban, C., Kumar, S., Moresco, J. J., Yates, J. R., 3rd, Campistol, J. M., Sancho-Martinez, I., Giacca, M., & Izpisua Belmonte, J. C. (2014). In vivo activation of a conserved microRNA program induces mammalian heart regeneration. *Cell Stem Cell, 15*(5), 589-604. doi:10.1016/j.stem.2014.10.003

Alexander, D. L., Ganem, L. G., Fernandez-Salguero, P., Gonzalez, F., & Jefcoate, C. R. (1998). Aryl-hydrocarbon receptor is an inhibitory regulator of lipid synthesis and of commitment to adipogenesis. *J Cell Sci, 111*, 3311-3322.

Ambrosi, T. H., Scialdone, A., Graja, A., Gohlke, S., Jank, A. M., Bocian, C., Woelk, L., Fan, H., Logan, D. W., Schurmann, A., Saraiva, L. R., & Schulz, T. J. (2017). Adipocyte Accumulation in the Bone Marrow during Obesity and

1 Aging Impairs Stem Cell-Based Hematopoietic and Bone Regeneration. *Cell*
2 *Stem Cell, 20*(6), 771-784.e776. doi:10.1016/j.stem.2017.02.009
3 Baruzzo, G., Hayer, K. E., Kim, E. J., Di Camillo, B., FitzGerald, G. A., & Grant, G.
4 R. (2017). Simulation-based comprehensive benchmarking of RNA-seq
5 aligners. *Nat Methods, 14*(2), 135-139. doi:10.1038/nmeth.4106
6 Boitano, A. E., Wang, J., Romeo, R., Bouchez, L. C., Parker, A. E., Sutton, S. E.,
7 Walker, J. R., Flaveny, C. A., Perdew, G. H., Denison, M. S., Schultz, P. G.,
8 & Cooke, M. P. (2010). Aryl hydrocarbon receptor antagonists promote the
9 expansion of human hematopoietic stem cells. *Science, 329*(5997), 1345-1348.
10 doi:10.1126/science.1191536
11 Bray, S. J. (2016). Notch signalling in context. *Nat Rev Mol Cell Biol, 17*(11), 722-
12 735. doi:10.1038/nrm.2016.94
13 Cawthorn, W. P., Bree, A. J., Yao, Y., Du, B., Hemati, N., Martinez-Santibanez, G.,
14 & MacDougald, O. A. (2012). Wnt6, Wnt10a and Wnt10b inhibit
15 adipogenesis and stimulate osteoblastogenesis through a beta-catenin-
16 dependent mechanism. *Bone, 50*(2), 477-489. doi:10.1016/j.bone.2011.08.010
17 Cawthorn, W. P., Scheller, E. L., Learman, B. S., Parlee, S. D., Simon, B. R., Mori,
18 H., Ning, X., Bree, A. J., Schell, B., Broome, D. T., Soliman, S. S.,
19 DelProposto, J. L., Lumeng, C. N., Mitra, A., Pandit, S. V., Gallagher, K. A.,
20 Miller, J. D., Krishnan, V., Hui, S. K., Bredella, M. A., Fazeli, P. K.,
21 Klibanski, A., Horowitz, M. C., Rosen, C. J., & MacDougald, O. A. (2014).
22 Bone marrow adipose tissue is an endocrine organ that contributes to
23 increased circulating adiponectin during caloric restriction. *Cell Metab, 20*(2),
24 368-375. doi:10.1016/j.cmet.2014.06.003

Chao, L. C., Bensinger, S. J., Villanueva, C. J., Wroblewski, K., & Tontonoz, P. (2008). Inhibition of adipocyte differentiation by Nur77, Nurr1, and Nor1. *Mol Endocrinol, 22*(12), 2596-2608. doi:10.1210/me.2008-0161

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics, 14*, 128. doi:10.1186/1471-2105-14-128

Cuellar-Partida, G., Buske, F. A., McLeay, R. C., Whitington, T., Noble, W. S., & Bailey, T. L. (2012). Epigenetic priors for identifying active transcription factor binding sites. *Bioinformatics, 28*(1), 56-62. doi:10.1093/bioinformatics/btr614

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dukler, N., Gulko, B., Huang, Y. F., & Siepel, A. (2016). Is a super-enhancer greater than the sum of its parts? *Nat Genet, 49*(1), 2-3. doi:10.1038/ng.3759

Ernst, J., & Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics, 7*, 191. doi:10.1186/1471-2105-7-191

Farmer, S. R. (2006). Transcriptional control of adipocyte formation. *Cell Metab, 4*(4), 263-273. doi:10.1016/j.cmet.2006.07.001

Fazeli, P. K., Horowitz, M. C., MacDougald, O. A., Scheller, E. L., Rodeheffer, M. S., Rosen, C. J., & Klibanski, A. (2013). Marrow fat and bone--new perspectives. *J Clin Endocrinol Metab, 98*(3), 935-945. doi:10.1210/jc.2012-3634

Forrest, A. R., Kawaji, H., Rehli, M., Baillie, J. K., de Hoon, M. J., Haberle, V., Lassmann, T., Kulakovskiy, I. V., Lizio, M., Itoh, M., Andersson, R., Mungall, C. J., Meehan, T. F., Schmeier, S., Bertin, N., Jorgensen, M., Dimont, E., Arner, E., Schmidl, C., Schaefer, U., Medvedeva, Y. A., Plessy, C., Vitezic, M., Severin, J., Semple, C., Ishizu, Y., Young, R. S., Francescatto, M., Alam, I., Albanese, D., Altschuler, G. M., Arakawa, T., Archer, J. A., Arner, P., Babina, M., Rennie, S., Balwierz, P. J., Beckhouse, A. G., Pradhan-Bhatt, S., Blake, J. A., Blumenthal, A., Bodega, B., Bonetti, A., Briggs, J., Brombacher, F., Burroughs, A. M., Califano, A., Cannistraci, C. V., Carbajo, D., Chen, Y., Chierici, M., Ciani, Y., Clevers, H. C., Dalla, E., Davis, C. A., Detmar, M., Diehl, A. D., Dohi, T., Drablos, F., Edge, A. S., Edinger, M., Ekwall, K., Endoh, M., Enomoto, H., Fagiolini, M., Fairbairn, L., Fang, H., Farach-Carson, M. C., Faulkner, G. J., Favorov, A. V., Fisher, M. E., Frith, M. C., Fujita, R., Fukuda, S., Furlanello, C., Furino, M., Furusawa, J., Geijtenbeek, T. B., Gibson, A. P., Gingeras, T., Goldowitz, D., Gough, J., Guhl, S., Guler, R., Gustincich, S., Ha, T. J., Hamaguchi, M., Hara, M., Harbers, M., Harshbarger, J., Hasegawa, A., Hasegawa, Y., Hashimoto, T., Herlyn, M., Hitchens, K. J., Ho Sui, S. J., Hofmann, O. M., Hoof, I., Hori, F., Huminiecki, L., Iida, K., Ikawa, T., Jankovic, B. R., Jia, H., Joshi, A., Jurman, G., Kaczkowski, B., Kai, C., Kaida, K., Kaiho, A., Kajiyama, K., Kanamori-Katayama, M., Kasianov, A. S., Kasukawa, T., Katayama, S., Kato, S., Kawaguchi, S., Kawamoto, H., Kawamura, Y. I., Kawashima, T., Kempfle, J. S., Kenna, T. J., Kere, J., Khachigian, L. M., Kitamura, T., Klinken, S. P., Knox, A. J., Kojima, M., Kojima, S., Kondo, N., Koseki, H., Koyasu, S., Krampitz, S., Kubosaki, A., Kwon, A. T., Laros, J. F., Lee, W., Lennartsson,

A., Li, K., Lilje, B., Lipovich, L., Mackay-Sim, A., Manabe, R., Mar, J. C., Marchand, B., Mathelier, A., Mejhert, N., Meynert, A., Mizuno, Y., de Lima Morais, D. A., Morikawa, H., Morimoto, M., Moro, K., Motakis, E., Motohashi, H., Mummery, C. L., Murata, M., Nagao-Sato, S., Nakachi, Y., Nakahara, F., Nakamura, T., Nakamura, Y., Nakazato, K., van Nimwegen, E., Ninomiya, N., Nishiyori, H., Noma, S., Noma, S., Noazaki, T., Ogishima, S., Ohkura, N., Ohimiya, H., Ohno, H., Ohshima, M., Okada-Hatakeyama, M., Okazaki, Y., Orlando, V., Ovchinnikov, D. A., Pain, A., Passier, R., Patrikakis, M., Persson, H., Piazza, S., Prendergast, J. G., Rackham, O. J., Ramilowski, J. A., Rashid, M., Ravasi, T., Rizzu, P., Roncador, M., Roy, S., Rye, M. B., Saijyo, E., Sajantila, A., Saka, A., Sakaguchi, S., Sakai, M., Sato, H., Savvi, S., Saxena, A., Schneider, C., Schultes, E. A., Schulze-Tanzil, G. G., Schwegmann, A., Sengstag, T., Sheng, G., Shimoji, H., Shimoni, Y., Shin, J. W., Simon, C., Sugiyama, D., Sugiyama, T., Suzuki, M., Suzuki, N., Swoboda, R. K., t Hoen, P. A., Tagami, M., Takahashi, N., Takai, J., Tanaka, H., Tatsukawa, H., Tatum, Z., Thompson, M., Toyodo, H., Toyoda, T., Valen, E., van de Wetering, M., van den Berg, L. M., Verado, R., Vijayan, D., Vorontsov, I. E., Wasserman, W. W., Watanabe, S., Wells, C. A., Winteringham, L. N., Wolvetang, E., Wood, E. J., Yamaguchi, Y., Yamamoto, M., Yoneda, M., Yonekura, Y., Yoshida, S., Zabierowski, S. E., Zhang, P. G., Zhao, X., Zucchelli, S., Summers, K. M., Suzuki, H., Daub, C. O., Kawai, J., Heutink, P., Hide, W., Freeman, T. C., Lenhard, B., Bajic, V. B., Taylor, M. S., Makeev, V. J., Sandelin, A., Hume, D. A., Carninci, P., & Hayashizaki, Y. (2014). A promoter-level mammalian expression atlas. *Nature, 507*(7493), 462-470. doi:10.1038/nature13182

1 Forristal, C. E., Wright, K. L., Hanley, N. A., Oreffo, R. O., & Houghton, F. D.
2 (2010). Hypoxia inducible factors regulate pluripotency and proliferation in
3 human embryonic stem cells cultured at reduced oxygen tensions.
4 *Reproduction, 139*(1), 85-97. doi:10.1530/rep-09-0300
5 Frietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X., & Farnham, P. J. (2010). ZNF274
6 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes.
7 *PLoS One, 5*(12), e15082. doi:10.1371/journal.pone.0015082
8 Galhardo, M., Berninger, P., Nguyen, T. P., Sauter, T., & Sinkkonen, L. (2015). Cell
9 type-selective disease-association of genes under high regulatory load. *Nucleic*
10 *Acids Res, 43*(18), 8839-8855. doi:10.1093/nar/gkv863
11 Galhardo, M., Sinkkonen, L., Berninger, P., Lin, J., Sauter, T., & Heinaniemi, M.
12 (2014). Integrated analysis of transcript-level regulation of metabolism reveals
13 disease-relevant nodes of the human metabolic network. *Nucleic Acids Res,*
14 *42*(3), 1474-1496. doi:10.1093/nar/gkt989
15 Gamez, B., Rodriguez-Carballo, E., Bartrons, R., Rosa, J. L., & Ventura, F. (2013).
16 MicroRNA-322 (miR-322) and its target protein Tob2 modulate Osterix (Osx)
17 mRNA stability. *J Biol Chem, 288*(20), 14264-14275.
18 doi:10.1074/jbc.M112.432104
19 Gansner, E. R., Koren, Y., & North, S. (2005). *Graph Drawing by Stress*
20 *Majorization*, Berlin, Heidelberg, 239-250 Retrieved from
21 https://link.springer.com/chapter/10.1007/978-3-540-31843-9_25
22 doi:10.1007/978-3-540-31843-9_25
23 Gasiewicz, T. A., Singh, K. P., & Bennett, J. A. (2014). The Ah receptor in stem cell
24 cycling, regulation, and quiescence. *Ann N Y Acad Sci, 1310*, 44-50.
25 doi:10.1111/nyas.12361

1 Gialitakis, M., Tolaini, M., Li, Y., Pardo, M., Yu, L., Toribio, A., Choudhary, J. S.,
2 Niakan, K., Papayannopoulos, V., & Stockinger, B. (2017). Activation of the
3 Aryl Hydrocarbon Receptor Interferes with Early Embryonic Development.
4 *Stem Cell Reports, 9*(5), 1377-1386. doi:10.1016/j.stemcr.2017.09.025
5 Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie, A. L. M., Lilly, A. J., Hannah, R.,
6 Lichtinger, M., Batta, K., Florkowska, M., Patel, R., Challinor, M., Wallace,
7 K., Gilmour, J., Assi, S. A., Cauchy, P., Hoogenkamp, M., Westhead, D. R.,
8 Lacaud, G., Kouskoff, V., Gottgens, B., & Bonifer, C. (2016). Dynamic Gene
9 Regulatory Networks Drive Hematopoietic Specification and Differentiation.
10 *Dev Cell, 36*(5), 572-587. doi:10.1016/j.devcel.2016.01.024
11 Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences
12 of a given motif. *Bioinformatics, 27*(7), 1017-1018.
13 doi:10.1093/bioinformatics/btr064
14 Gusmao, E. G., Allhoff, M., Zenke, M., & Costa, I. G. (2016). Analysis of
15 computational footprinting methods for DNase sequencing experiments. *Nat*
16 *Methods, 13*(4), 303-309. doi:10.1038/nmeth.3772
17 Hakelien, A. M., Bryne, J. C., Harstad, K. G., Lorenz, S., Paulsen, J., Sun, J.,
18 Mikkelsen, T. S., Myklebost, O., & Meza-Zepeda, L. A. (2014). The
19 regulatory landscape of osteogenic differentiation. *Stem Cells, 32*(10), 2780-
20 2793. doi:10.1002/stem.1759
21 Hassan, M. Q., Javed, A., Morasso, M. I., Karlin, J., Montecino, M., van Wijnen, A.
22 J., Stein, G. S., Stein, J. L., & Lian, J. B. (2004). Dlx3 transcriptional
23 regulation of osteoblast differentiation: temporal recruitment of Msx2, Dlx3,
24 and Dlx5 homeodomain proteins to chromatin of the osteocalcin gene. *Mol*
25 *Cell Biol, 24*(20), 9248-9261. doi:10.1128/mcb.24.20.9248-9261.2004

1 Heinaniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L.,
2 Zhou, J. X., Kreisberg, R., Kauffman, S. A., Huang, S., & Shmulevich, I.
3 (2013). Gene-pair expression signatures reveal lineage control. *Nat Methods,*
4 *10*(6), 577-583. doi:10.1038/nmeth.2445
5 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X.,
6 Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-
7 determining transcription factors prime cis-regulatory elements required for
8 macrophage and B cell identities. *Mol Cell, 38*(4), 576-589.
9 doi:10.1016/j.molcel.2010.05.004
10 Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke,
11 H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity
12 and disease. *Cell, 155*(4), 934-947. doi:10.1016/j.cell.2013.09.053
13 Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I.,
14 Bradner, J. E., & Young, R. A. (2015). Convergence of developmental and
15 oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell,*
16 *58*(2), 362-370. doi:10.1016/j.molcel.2015.02.014
17 Horie, T., Ono, K., Kinoshita, M., Nishi, H., Nagao, K., Kawamura, T., Abe, Y.,
18 Wada, H., Shimatsu, A., Kita, T., & Hasegawa, K. (2008). TG-interacting
19 factor is required for the differentiation of preadipocytes. *J Lipid Res, 49*(6),
20 1224-1234. doi:10.1194/jlr.M700578-JLR200
21 Huang, B., Butler, R., Miao, Y., Dai, Y., Wu, W., Su, W., Fujii-Kuriyama, Y.,
22 Warner, M., & Gustafsson, J. A. (2016). Dysregulation of Notch and ERalpha
23 signaling in AhR-/- male mice. *Proc Natl Acad Sci U S A, 113*(42), 11883-
24 11888. doi:10.1073/pnas.1613269113

Iwata, K., Li, J., Follet, H., Phipps, R. J., & Burr, D. B. (2006). Bisphosphonates suppress periosteal osteoblast activity independently of resorption in rat femur and tibia. *Bone, 39*(5), 1053-1058. doi:10.1016/j.bone.2006.05.006

Janky, R., Verfaillie, A., Imrichova, H., Van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., Svetlichnyy, D., Kalender Atak, Z., Fiers, M., Marine, J. C., & Aerts, S. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol, 10*(7), e1003731. doi:10.1371/journal.pcbi.1003731

Jensen, B. A., Leeman, R. J., Schlezinger, J. J., & Sherr, D. H. (2003). Aryl hydrocarbon receptor (AhR) agonists suppress interleukin-6 expression by bone marrow stromal cells: an immunotoxicology study. *Environ Health, 2*(1), 16. doi:10.1186/1476-069x-2-16

Kopylova, E., Noe, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics, 28*(24), 3211-3217. doi:10.1093/bioinformatics/bts611

Korkalainen, M., Kallio, E., Olkku, A., Nelo, K., Ilvesaro, J., Tuukkanen, J., Mahonen, A., & Viluksela, M. (2009). Dioxins interfere with differentiation of osteoblasts and osteoclasts. *Bone, 44*(6), 1134-1142. doi:10.1016/j.bone.2009.02.019

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res, 44*(W1), W90-97. doi:10.1093/nar/gkw377

Lee, J. S., Cella, M., McDonald, K. G., Garlanda, C., Kennedy, G. D., Nukaya, M., Mantovani, A., Kopan, R., Bradfield, C. A., Newberry, R. D., & Colonna, M. (2011). AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. *Nat Immunol, 13*(2), 144-151. doi:10.1038/ni.2187

Lee, N. K., Sowa, H., Hinoi, E., Ferron, M., Ahn, J. D., Confavreux, C., Dacquin, R., Mee, P. J., McKee, M. D., Jung, D. Y., Zhang, Z., Kim, J. K., Mauvais-Jarvis, F., Ducy, P., & Karsenty, G. (2007). Endocrine regulation of energy metabolism by the skeleton. *Cell, 130*(3), 456-469. doi:10.1016/j.cell.2007.05.047

Lee, S. Y., Noh, H. B., Kim, H. T., Lee, K. I., & Hwang, D. Y. (2017). Glis family proteins are differentially implicated in the cellular reprogramming of human somatic cells. *Oncotarget, 8*(44), 77041-77049. doi:10.18632/oncotarget.20334

Lefterova, M. I., Zhang, Y., Steger, D. J., Schupp, M., Schug, J., Cristancho, A., Feng, D., Zhuo, D., Stoeckert, C. J., Jr., Liu, X. S., & Lazar, M. A. (2008). PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev, 22*(21), 2941-2952. doi:10.1101/gad.1709008

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923-930. doi:10.1093/bioinformatics/btt656

1 Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing
2 reads. *BMC Res Notes, 5*, 337. doi:10.1186/1756-0500-5-337
3 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change
4 and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550.
5 doi:10.1186/s13059-014-0550-8
6 Lv, B., Li, F., Fang, J., Xu, L., Sun, C., Han, J., Hua, T., Zhang, Z., Feng, Z., & Jiang,
7 X. (2017). Hypoxia inducible factor 1alpha promotes survival of mesenchymal
8 stem cells under hypoxia. *Am J Transl Res, 9*(3), 1521-1529.
9 Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka,
10 T., Kawamura, Y., Mochizuki, H., Goshima, N., & Yamanaka, S. (2011).
11 Direct reprogramming of somatic cells is promoted by maternal transcription
12 factor Glis1. *Nature, 474*(7350), 225-229. doi:10.1038/nature10106
13 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
14 sequencing reads. *EMBnet.journal, 17*(1), 10-12. doi:10.14806/ej.17.1.200
15 Mikkelsen, T. S., Xu, Z., Zhang, X., Wang, L., Gimble, J. M., Lander, E. S., & Rosen,
16 E. D. (2010). Comparative epigenomic analysis of murine and human
17 adipogenesis. *Cell, 143*(1), 156-169. doi:10.1016/j.cell.2010.09.006
18 Naruse, M., Ishihara, Y., Miyagawa-Tomita, S., Koyama, A., & Hagiwara, H. (2002).
19 3-Methylcholanthrene, which binds to the arylhydrocarbon receptor, inhibits
20 proliferation and differentiation of osteoblasts in vitro and ossification in vivo.
21 *Endocrinology, 143*(9), 3575-3581. doi:10.1210/en.2002-220003
22 Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B.,
23 Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., Maurano, M. T.,
24 Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M.,
25 Roach, V., Dunn, D., Neri, J., Schafer, A., Hansen, R. S., Kutyavin, T., Giste,

E., Weaver, M., Canfield, T., Sabo, P., Zhang, M., Balasundaram, G., Byron, R., MacCoss, M. J., Akey, J. M., Bender, M. A., Groudine, M., Kaul, R., & Stamatoyannopoulos, J. A. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature, 489*(7414), 83-90. doi:10.1038/nature11212

Nielsen, R., Pedersen, T. A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., Denissov, S., Borgesen, M., Francoijs, K. J., Mandrup, S., & Stunnenberg, H. G. (2008). Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev, 22*(21), 2953-2967. doi:10.1101/gad.501108

Ogawa, M., Nishikawa, S., Ikuta, K., Yamamura, F., Naito, M., Takahashi, K., & Nishikawa, S. (1988). B cell ontogeny in murine embryo studied by a culture system with the monolayer of a stromal cell clone, ST2: B cell progenitor develops first in the embryonal body rather than in the yolk sac. *Embo j, 7*(5), 1337-1343.

Omatsu, Y., Seike, M., Sugiyama, T., Kume, T., & Nagasawa, T. (2014). Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. *Nature, 508*(7497), 536-540. doi:10.1038/nature13071

Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., Black, B. L., Visel, A., Pennacchio, L. A., & Collins, F. S. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A, 110*(44), 17921-17926. doi:10.1073/pnas.1317023110

Pradhan, R. N., Bues, J. J., Gardeux, V., Schwalie, P. C., Alpern, D., Chen, W., Russeil, J., Raghav, S. K., & Deplancke, B. (2017). Dissecting the brown adipogenic regulatory network using integrative genomics. *Scientific Reports, 7*, 42130. doi:10.1038/srep42130

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Rackham, O. J., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., Petretto, E., Forrest, A. R., Hayashizaki, Y., Polo, J. M., & Gough, J. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nat Genet, 48*(3), 331-335. doi:10.1038/ng.3487

Ramirez, R. N., El-Ali, N. C., Mager, M. A., Wyman, D., Conesa, A., & Mortazavi, A. (2017). Dynamic Gene Regulatory Networks of Human Myeloid Differentiation. *Cell Syst, 4*(4), 416-429.e413. doi:10.1016/j.cels.2017.03.005

Roider, H. G., Kanhere, A., Manke, T., & Vingron, M. (2007). Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics, 23*(2), 134-141. doi:10.1093/bioinformatics/btl565

Ruvkun, G., & Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. *Science, 282*(5396), 2033-2041. doi:10.1126/science.282.5396.2033

Scheller, E. L., Burr, A. A., MacDougald, O. A., & Cawthorn, W. P. (2016). Inside out: Bone marrow adipose tissue as a source of circulating adiponectin. *Adipocyte, 5*(3), 251-269. doi:10.1080/21623945.2016.1149269

Schmidt, F., Gasparoni, N., Gasparoni, G., Gianmoena, K., Cadenas, C., Polansky, J. K., Ebert, P., Nordstrom, K., Barann, M., Sinha, A., Frohler, S., Xiong, J., Dehghani Amirabad, A., Behjati Ardakani, F., Hutter, B., Zipprich, G., Felder, B., Eils, J., Brors, B., Chen, W., Hengstler, J. G., Hamann, A., Lengauer, T., Rosenstiel, P., Walter, J., & Schulz, M. H. (2017). Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. *Nucleic Acids Res, 45*(1), 54-66. doi:10.1093/nar/gkw1061

Schubert, M., Ermini, L., Der Sarkissian, C., Jonsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernandez, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc, 9*(5), 1056-1082. doi:10.1038/nprot.2014.063

Schulz, M. H., Devanny, W. E., Gitter, A., Zhong, S., Ernst, J., & Bar-Joseph, Z. (2012). DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Systems Biology, 6*, 104-104. doi:10.1186/1752-0509-6-104

Shimba, S., Wada, T., & Tezuka, M. (2001). Arylhydrocarbon receptor (AhR) is involved in negative regulation of adipose differentiation in 3T3-L1 cells: AhR inhibits adipose differentiation independently of dioxin. *J Cell Sci, 114*(Pt 15), 2809-2817.

Siersbaek, R., Madsen, J. G. S., Javierre, B. M., Nielsen, R., Bagge, E. K., Cairns, J., Wingett, S. W., Traynor, S., Spivakov, M., Fraser, P., & Mandrup, S. (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte

Differentiation. *Mol Cell, 66*(3), 420-435.e425. doi:10.1016/j.molcel.2017.04.010

Siersbaek, R., Rabiee, A., Nielsen, R., Sidoli, S., Traynor, S., Loft, A., La Cour Poulsen, L., Rogowska-Wrzesinska, A., Jensen, O. N., & Mandrup, S. (2014). Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep, 7*(5), 1443-1455. doi:10.1016/j.celrep.2014.04.042

Silva, B. C., & Kousteni, S. (2012). Role of Osteoblasts in Regulation of Energy Metabolism. *Clinical Reviews in Bone and Mineral Metabolism, 11*(1), 2-10. doi:10.1007/s12018-012-9128-8

Singh, K. P., Garrett, R. W., Casado, F. L., & Gasiewicz, T. A. (2011). Aryl hydrocarbon receptor-null allele mice have hematopoietic stem/progenitor cells with abnormal characteristics and functions. *Stem Cells Dev, 20*(5), 769-784. doi:10.1089/scd.2010.0333

Sun, Y., Xu, J., Xu, L., Zhang, J., Chan, K., Pan, X., & Li, G. (2017). MiR-503 Promotes Bone Formation in Distraction Osteogenesis through Suppressing Smurf1 Expression. *Sci Rep, 7*(1), 409. doi:10.1038/s41598-017-00466-4

Tamm, C., Bower, N., & Anneren, C. (2011). Regulation of mouse embryonic stem cell self-renewal by a Yes-YAP-TEAD2 signaling pathway downstream of LIF. *J Cell Sci, 124*(Pt 7), 1136-1144. doi:10.1242/jcs.075796

Tong, Y., Niu, M., Du, Y., Mei, W., Cao, W., Dou, Y., Yu, H., Du, X., Yuan, H., & Zhao, W. (2017). Aryl hydrocarbon receptor suppresses the osteogenesis of mesenchymal stem cells in collagen-induced arthritic mice through the inhibition of beta-catenin. *Exp Cell Res, 350*(2), 349-357. doi:10.1016/j.yexcr.2016.12.009

Tontonoz, P., & Spiegelman, B. M. (2008). Fat and beyond: the diverse biology of PPARgamma. *Annu Rev Biochem, 77*, 289-312. doi:10.1146/annurev.biochem.77.061307.091829

Tremethick, D. J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell, 128*(4), 651-654. doi:10.1016/j.cell.2007.02.008

Varrette, S., Bouvry, P., Cartiaux, H., & Georgatos, F. (2014, 21-25 July 2014). *Management of an academic HPC cluster: The UL experience.* Paper presented at the 2014 International Conference on High Performance Computing & Simulation (HPCS), 959-967 Retrieved from http://ieeexplore.ieee.org/document/6903792/ doi:10.1109/HPCSim.2014.6903792

Wu, C., Jin, X., Tsueng, G., Afrasiabi, C., & Su, A. I. (2016). BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res, 44*(D1), D313-316. doi:10.1093/nar/gkv1104

Wu, H., Gordon, J. A., Whitfield, T. W., Tai, P. W., van Wijnen, A. J., Stein, J. L., Stein, G. S., & Lian, J. B. (2017). Chromatin dynamics regulate mesenchymal stem cell lineage specification and differentiation to osteogenesis. *Biochim Biophys Acta, 1860*(4), 438-449. doi:10.1016/j.bbagrm.2017.01.003

Yoshida, C. A., Komori, H., Maruyama, Z., Miyazaki, T., Kawasaki, K., Furuichi, T., Fukuyama, R., Mori, M., Yamana, K., Nakamura, K., Liu, W., Toyosawa, S., Moriishi, T., Kawaguchi, H., Takada, K., & Komori, T. (2012). SP7 inhibits osteoblast differentiation at a late stage in mice. *PLoS One, 7*(3), e32364. doi:10.1371/journal.pone.0032364

Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., & Peng, W. (2009). A clustering approach for identification of enriched domains from histone

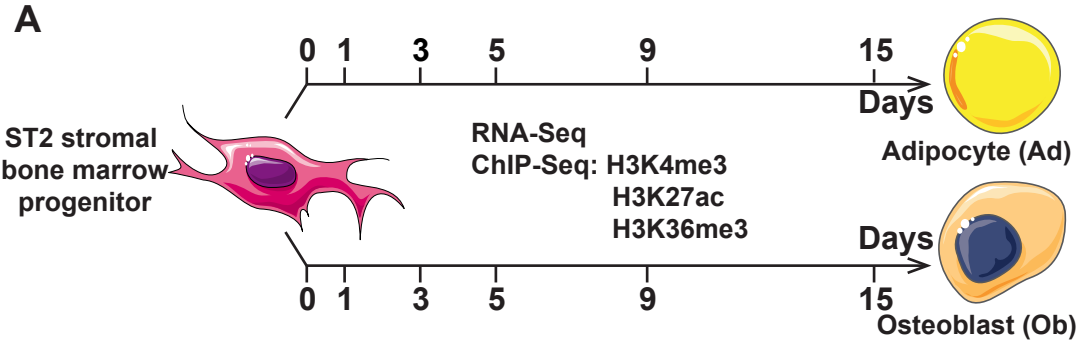
modification ChIP-Seq data. *Bioinformatics, 25*(15), 1952-1958. doi:10.1093/bioinformatics/btp340

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi:10.1186/gb-2008-9-9-r137

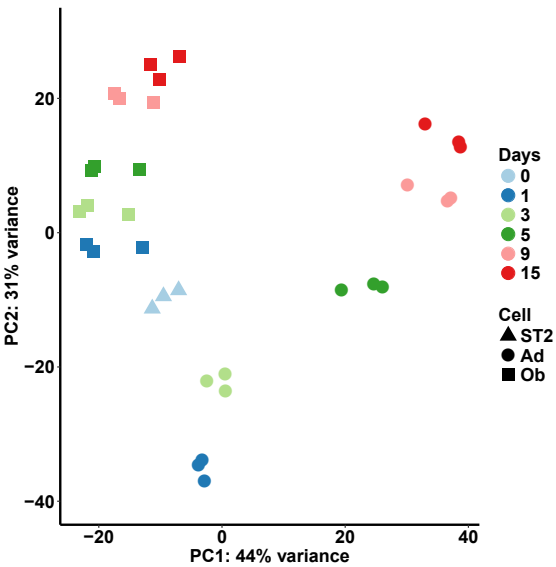
Figure 1. Time-series RNA-seq and ChIP-seq profiling of adipocyte and osteoblast differentiation from shared bone marrow progenitor cells. (A) Schematic representation of the experimental set-up. ST2 MSCs were differentiated towards adipocytes or osteoblasts. Total RNA and chromatin was collected at the indicated time points from both differentiation time courses and subjected to RNA-seq or ChIP-seq analysis using antibodies against the indicated histone modifications. (B) Principle component analysis (PCA) of the RNA-seq data. All replicates are indicated with each time point in a different colour. Triangles marks the undifferentiated ST2 cells, circles the adipocyte (Ad) samples and squares the osteoblast (Ob) samples. (C) Venn diagram comparing all differentially expressed genes at any time point of adipogenesis ($\log_2 FC > 1$, FDR<0.05) to those similarly affected in osteoblastogenesis. (D) Heatmap depicting the top 100 genes with the highest variance across the time points and lineages.

Figure 1

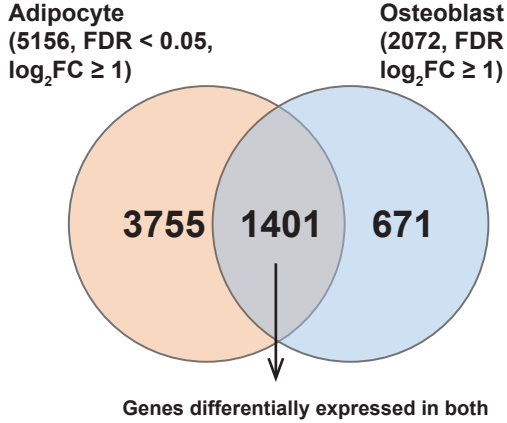
A



B



C



D

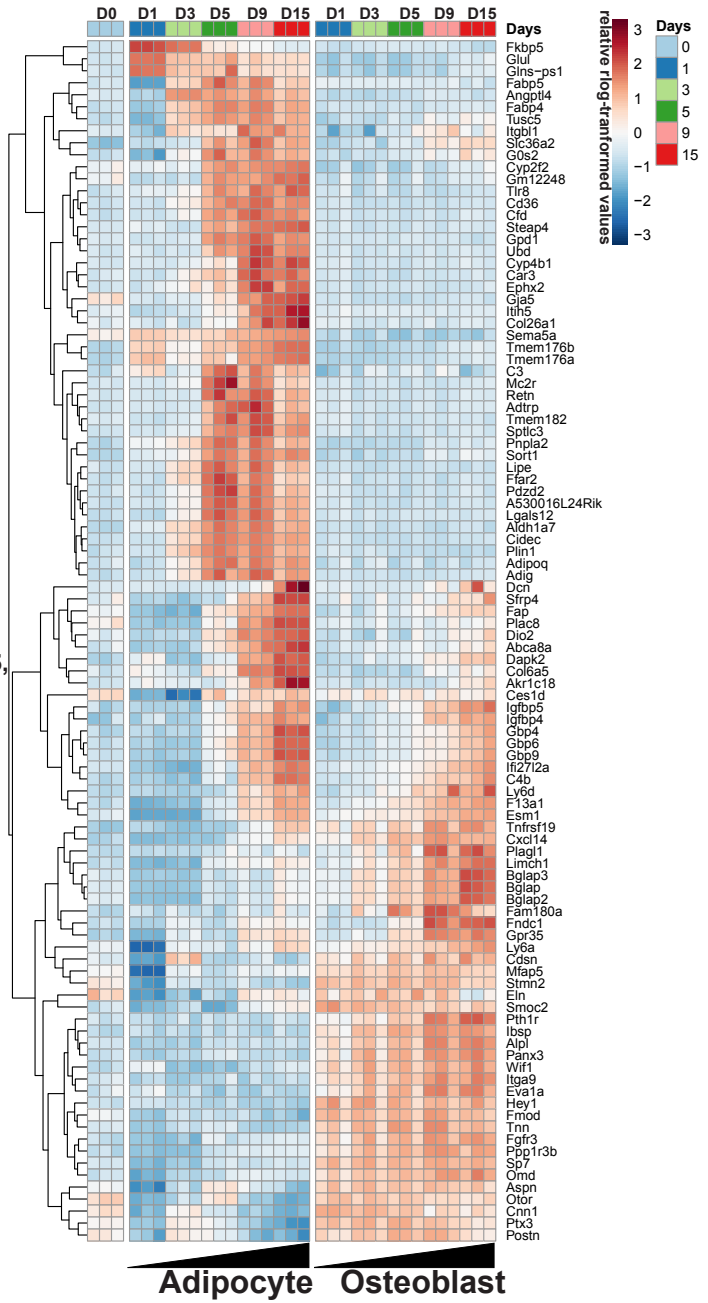


Figure 2: **Workflow of the EPIC-DREM approach and benchmarking against other methods.** (A) (i-iii) From time-series epigenetic experiments, e.g. DNase1-seq or histone ChIP-seq, putative TF binding sites are identified through footprint/peak calling and are annotated with TF affinities using TRAP. Random genomic regions with similar genomic characteristics (GC-content and length) compared to measured footprints/peaks are generated and annotated with TF affinities as well. A TF- and timepoint-specific affinity threshold can be obtained by applying an empirical *p*-value threshold (e.g. 0.05) on the distribution of TF affinities calculated on the randomly selected regions. Applying the threshold on the TF affinities computed in (iii) results in a set of discrete TF affinities per TF and timepoint. (iv) Using TEPIC, the discrete TF affinities are integrated into discrete, timepoint specific TF-gene association scores. (iv-vi) TF-gene associations and time-series gene expression data are used as input to DREM, which predicts regulators that distinguish subsets of genes that show similar gene expression changes over time. Different time points are indicated by different colours. (B) Literature benchmarking of the different methods. The time-series data from MSC differentation to adipocytes and osteoblasts (Figure 1A) was used as input for 4 different approaches (DREM2.0, EPIC-DREM, DREM-TRAP and random TF-target assignment) to identify top TFs (15 TFs per time point) controlling the respective lineages (see Supplementary Table S2 for the detailed list). For each identified TF the literature was searched for existing evidence for that TFs role in the corresponding differentiation. The fraction of the predicted TFs for which literature exists are shown per method and lineage, with the total number of identified unique top TFs indicated. (C) The distributions of the obtained DREM split scores per method and lineage are shown, with EPIC-DREM typically obtaining highest scores at most split points.

Figure 2

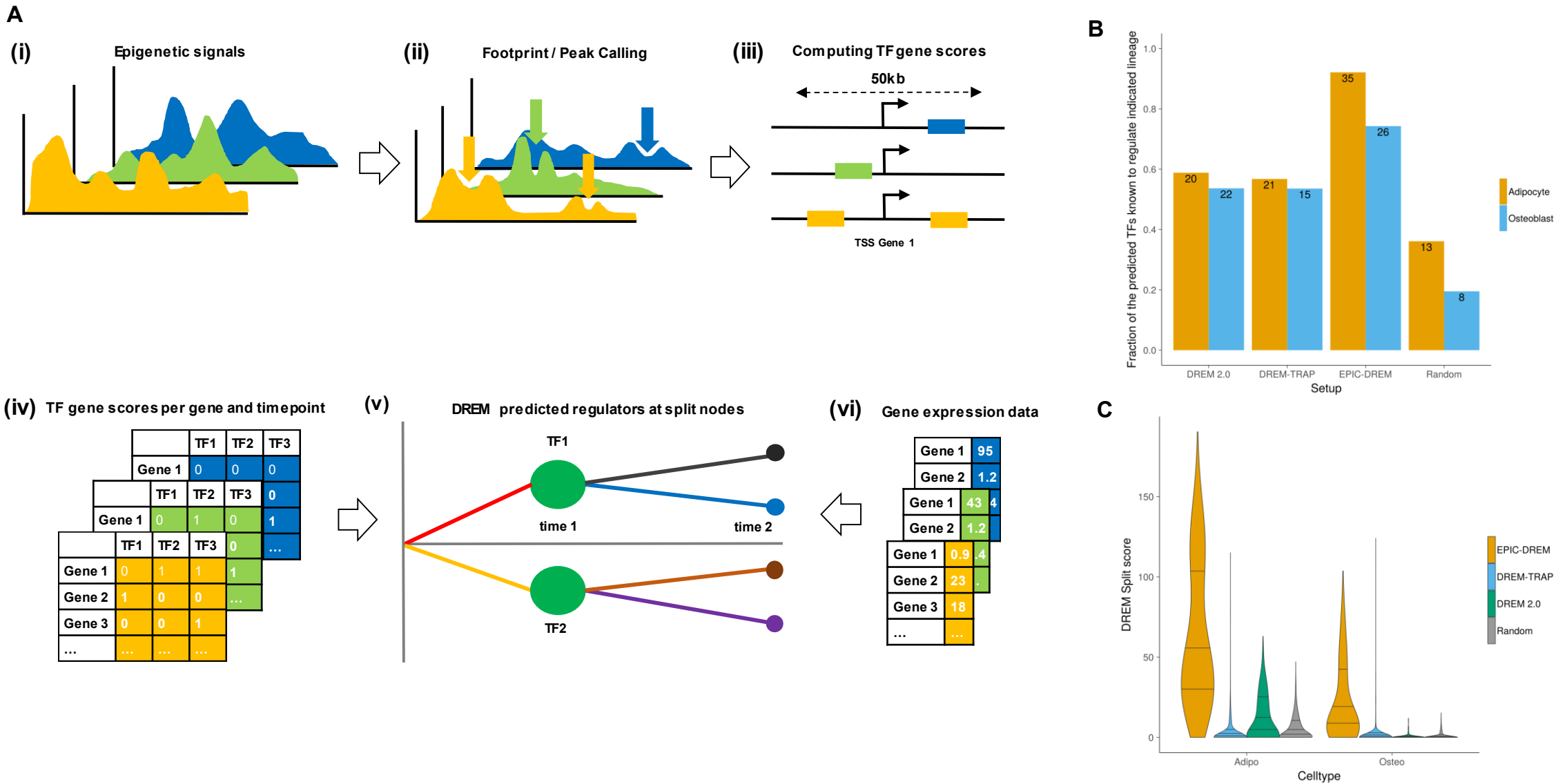

A
(i)
Epigenetic signals
(ii)
Footprint / Peak Calling
(iii)
Computing TF gene scores
50kb
TSS Gene 1
(iv)
TF gene scores per gene and timepoint
TF1
TF2
TF3
Gene 1
Gene 2
Gene 3
(v)
DREM predicted regulators at split nodes
TF1
time 1
time 2
TF2
(vi)
Gene expression data
Gene 1
Gene 2
Gene 3
95
1.2
43
1.2
0.9
23
18
B
Fraction of the predicted TFs known to regulate indicated lineage
DREM 2.0
DREM-TRAP
EPIC-DREM
Random
Setup
Adipocyte
Osteoblast
20
22
21
15
35
26
13
8
C
DREM Split score
Adipo
Osteo
Celltype
EPIC-DREM
DREM-TRAP
DREM 2.0
Random

Figure 3. Derivation of lineage-specific regulatory interactions using EPIC-DREM. EPIC-DREM clusters genes into paths of similar expression over time in (A) adipocyte and (B) osteoblast differentiation. Split points differentiate time point-specific changes in gene expression, that are annoted by DREM with matching TFs based on the predicted TF-target gene interactions. For each such path a GRN of thousands of interactions has been derived. The top TFs based on their time point- and path-specific split scores can be used to generate TF-TF networks underlying the observed gene regulation. The TF-TF networks of top 25 TFs are indicated for (A) 3 selected paths in adipocyte and (B) 2 selected paths in osteoblast differentiation. The top TFs are colored depending whether they are upregulated (blue) or downregulated (orange) compared to the previous time point and the size of the node corresponds to the total number of predicted target at the indicated time point. The full set of TFs predicted to control each path with at least 30% of the genes in the upstream split point with a split score ≤ 0.01 are listed in Supplementary Table S3. (C) Overlap of the time point-specific top TFs between the lineages. For both lineages lists of top 20 TFs per time point were generated by combining the predictions from individual splits. The Venn diagrams indicate the extent of overlap between the lineage-specific top TFs at each time point and the shared top TFs are indicated.

Figure 3

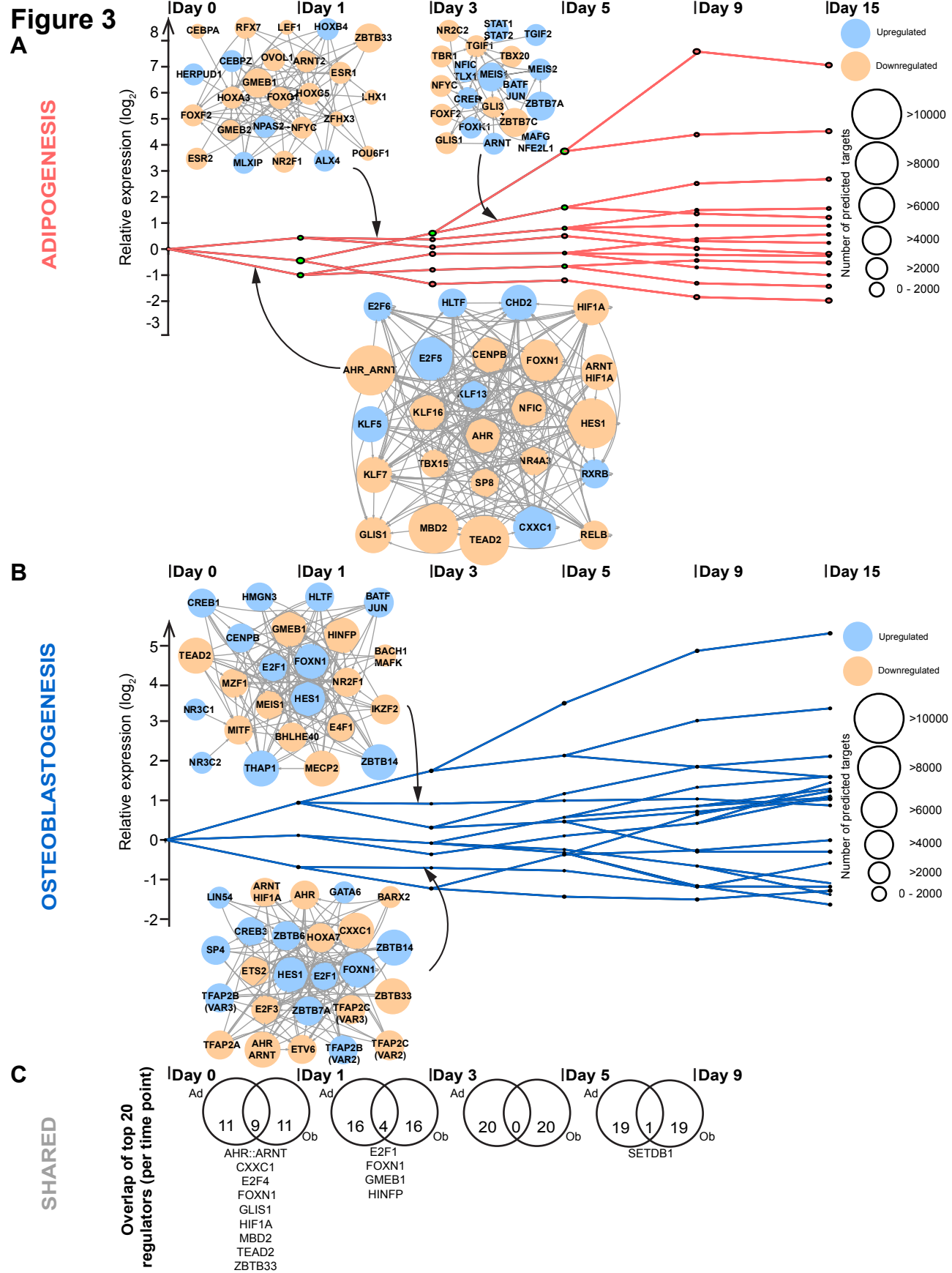

A
ADIPOGENESIS
Day 0
Day 1
Day 3
Day 5
Day 9
Day 15
Relative expression (log2)
Upregulated
Downregulated
Number of predicted targets
>10000
>8000
>6000
>4000
>2000
0 - 2000
CEBPA
RFX7
LEF1
HOXB4
ZBTB33
CEBPZ
OVOL1
ARNT2
HERPUD1
ESR1
GMEB1
HOXA3
FOXG1
HOXG5
LHX1
FOXF2
ZFHX3
GMEB2
NPAS2
NFYC
ESR2
MLXIP
NR2F1
ALX4
POU6F1
NR2C2
STAT1
STAT2
TGIF2
TGIF1
TBR1
TBX20
NFIC
MEIS2
TLX1
MEIS1
NFYC
BATF
JUN
CREB
ZBTB7A
GLI3
FOXF2
ZBTB7C
FOXK1
GLIS1
ARNT
MAFG
NFE2L1
E2F6
HLTF
CHD2
HIF1A
AHR_ARNT
E2F5
CENPB
FOXN1
ARNT
HIF1A
KLF13
KLF16
NFIC
KLF5
HES1
AHR
TBX15
NR4A3
KLF7
RXRB
SP8
GLIS1
MBD2
TEAD2
CXXC1
RELB
B
OSTEOBLASTOGENESIS
Day 0
Day 1
Day 3
Day 5
Day 9
Day 15
Relative expression (log2)
Upregulated
Downregulated
Number of predicted targets
>10000
>8000
>6000
>4000
>2000
0 - 2000
CREB1
HMGN3
HLTF
BATF
JUN
GMEB1
CENPB
HINFP
TEAD2
BACH1
MAFK
E2F1
FOXN1
MZF1
NR2F1
HES1
NR3C1
MEIS1
IKZF2
MITF
E4F1
BHLHE40
NR3C2
THAP1
MECP2
ZBTB14
LIN54
ARNT
HIF1A
AHR
GATA6
BARX2
CREB3
ZBTB6
HOXA7
CXXC1
SP4
ZBTB14
ETS2
HES1
E2F1
FOXN1
TFAP2B
(VAR3)
E2F3
ZBTB7A
TFAP2C
(VAR3)
ZBTB33
TFAP2A
AHR
ARNT
ETV6
TFAP2B
(VAR2)
TFAP2C
(VAR2)
C
SHARED
Overlap of top 20 regulators (per time point)
Day 0
Ad
11
9
11
Ob
AHR::ARNT
CXXC1
E2F4
FOXN1
GLIS1
HIF1A
MBD2
TEAD2
ZBTB33
Day 1
Ad
16
4
16
Ob
E2F1
FOXN1
GMEB1
HINFP
Day 3
Ad
20
0
20
Ob
Day 5
Ad
19
1
19
Ob
SETDB1
Day 9

Figure 4. Identification of dynamic SEs. (A) Overview depicting the enrichment of H3K27ac at the *Cxcl12* locus across the time points of adipogenesis (in magenta) and osteoblastogenesis (in light blue). The magenta and light blue bars indicate the SE regions identified by HOMER in adipogenesis and osteoblastogenesis, respectively, and black bar indicates the merged SE derived through overlapping of the individual SEs across the lineages. (B) The correlation between mRNA levels and SE signal. The *Cxcl12* mRNA levels as measured by RNA-seq and the merged SE signal at the *Cxcl12* locus as measured by the reads detected in the H3K27ac IPs are depicted across the time-series of adipocyte (upper panel) and osteoblast (lower panel) differentiation. Intact line = mRNA level and dashed line = merged SE signal. r = Pearson correlation co-efficient (C-D) STEM clustering of the merged SEs according to their dynamic profiles identifies (C) 4 main profiles in adipogenesis and (D) two main profiles in osteoblastogenesis. The additional profiles are shown in Supplementary Figure S5. y-axis indicates the signal change in $\log_2$-scale. (E) Venn analysis of the dynamic merged SEs from both lineages identifies 95 adipocyte-specific, 54 osteoblast-specific, and 25 shared dynamic SEs with $\log_2 FC \geq 1$. The dynamic SEs were associated to their target genes based on Pearson correlation with gene expression levels across the time series and the regulatory genes (TFs and miRNAs) associated with dynamic SEs are indicated in the corresponding boxes. All dynamic SEs and their putative target genes are listed in Supplementary Table S6. Adipocyte D9 sample for H3K27ac was not included in the above analysis due to low number of mappable high quality reads.

Figure 4

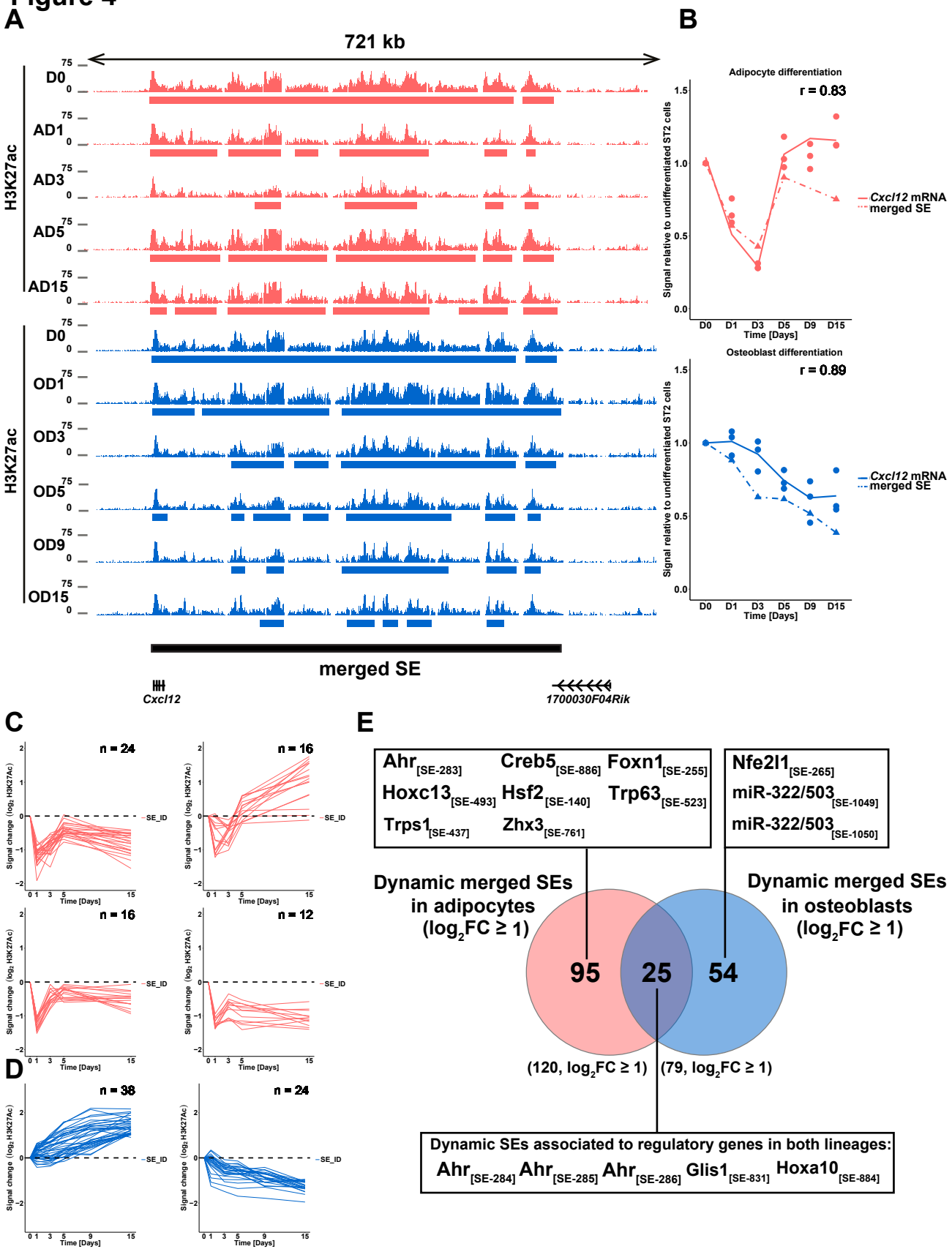

A
721 kb
H3K27ac
D0
AD1
AD3
AD5
AD15
H3K27ac
D0
OD1
OD3
OD5
OD9
OD15
merged SE
Cxcl12
1700030F04Rik
B
Adipocyte differentiation
r = 0.83
Cxcl12 mRNA
merged SE
Signal relative to undifferentiated ST2 cells
Time [Days]
Osteoblast differentiation
r = 0.89
Cxcl12 mRNA
merged SE
C
n = 24
n = 16
n = 16
n = 12
SE_ID
Signal change (log2 H3K27Ac)
Time [Days]
D
n = 38
n = 24
E
Ahr[SE-283]
Creb5[SE-886]
Foxn1[SE-255]
Hoxc13[SE-493]
Hsf2[SE-140]
Trp63[SE-523]
Trps1[SE-437]
Zhx3[SE-761]
Nfe2l1[SE-265]
miR-322/503[SE-1049]
miR-322/503[SE-1050]
Dynamic merged SEs in adipocytes (log2FC ≥ 1)
Dynamic merged SEs in osteoblasts (log2FC ≥ 1)
95
25
54
(120, log2FC ≥ 1)
(79, log2FC ≥ 1)
Dynamic SEs associated to regulatory genes in both lineages:
Ahr[SE-284] Ahr[SE-285] Ahr[SE-286] Glis1[SE-831] Hoxa10[SE-884]

Figure 5. *Ahr* is regulated by multiple SEs with lineage-specific dynamics. (A-B) Overview depicting the enrichments of H3K4me3 (in dark blue and purple), H3K36me3 (in light and dark green), and H3K27ac (magenta and light blue) at the *Ahr* locus across the time points of adipogenesis (A) and osteoblastogenesis (B), respectively. The magenta and light blue bars indicate the merged SE regions identified through the analysis described in Figure 4. See also Supplementary Figure S4. (C-D) *Ahr* downregulation correlates with a decreased signal from all four SEs. The *Ahr* mRNA level was measured across the differentiation by RNA-seq (upper panel) and RT-qPCR (lower panel) in both adipocyte (C) and osteoblast (D) differentiation and is indicated as the intact line. Dashed lines represent the signal from the individual indicated merged SEs. r = Pearson correlation co-efficient. The statistical significance for RT-qPCR measurements compared to the value on D0 was determined by two-tailed Student's t-test. *=$p<0.05$, **=$p<0.01$ and ***=$p<0.001$. Data points represent mean of 3 biological replicates +/- SEM. AD9 sample for H3K27ac and OD15 for H3K4me3 were not included in the above analysis due to lower number of mappable high quality reads. (E) Overview depicting the enrichment of H3K27ac at the *Ahr* locus in the confluent undifferentiated (D0) and differentiated (D7) 3T3-L1 adipocyte cell line. No SE formation could be detected in these more lineage-committed cells. The data were obtained from (Mikkelsen et al., 2010). The H3K27ac enrichments at the corresponding locus in human cell types are indicated in Supplementary Figure S7A.

Figure 5

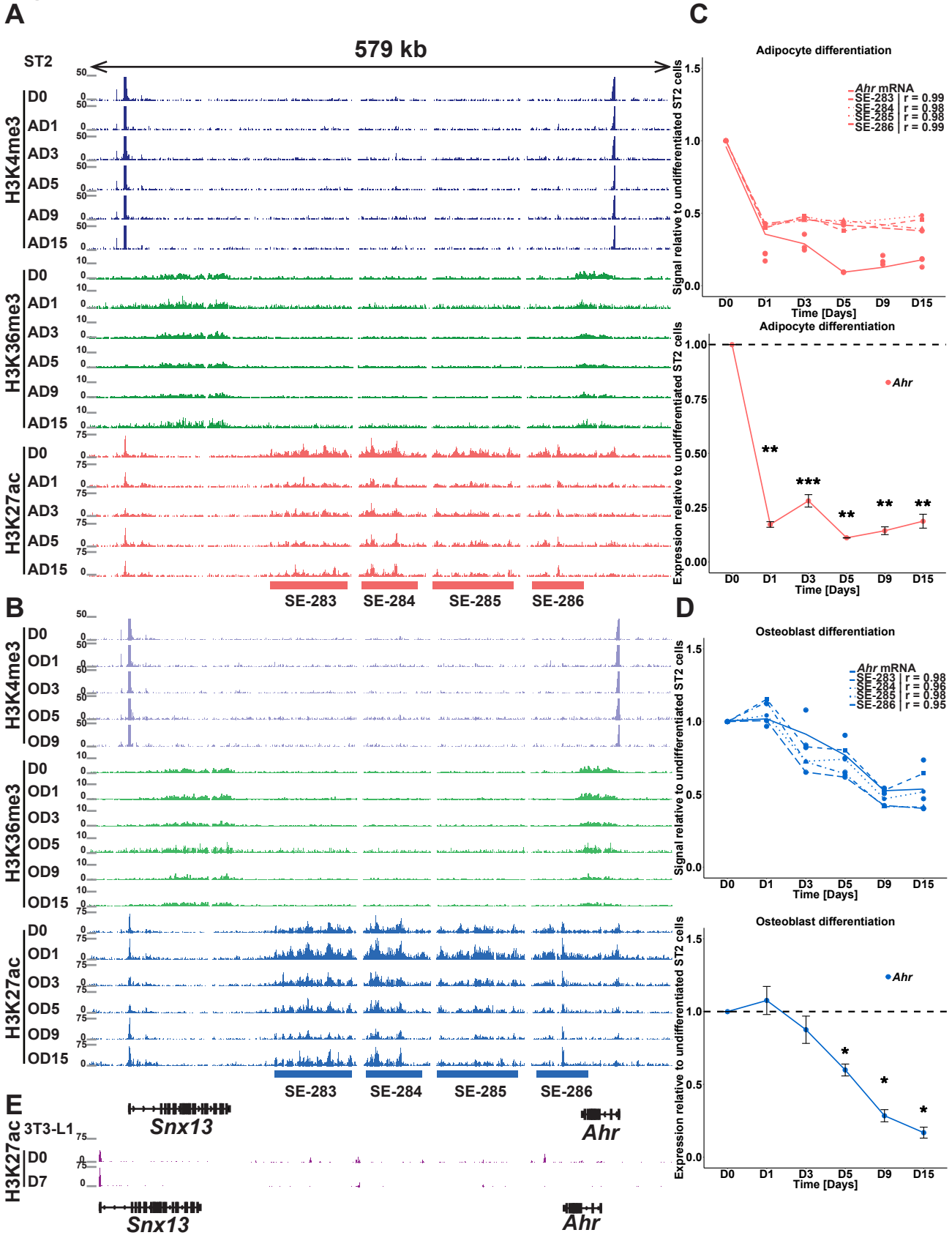

A
ST2
579 kb
H3K4me3
H3K36me3
H3K27ac
D0
AD1
AD3
AD5
AD9
AD15
SE-283
SE-284
SE-285
SE-286
B
OD1
OD3
OD5
OD9
OD15
C
Adipocyte differentiation
Ahr mRNA
SE-283 r = 0.99
SE-284 r = 0.98
SE-285 r = 0.98
SE-286 r = 0.99
Signal relative to undifferentiated ST2 cells
Time [Days]
Ahr
Expression relative to undifferentiated ST2 cells
D
Osteoblast differentiation
SE-283 r = 0.98
SE-284 r = 0.96
SE-285 r = 0.98
SE-286 r = 0.95
E
3T3-L1
D7
Snx13
Ahr

Figure 6. *Glis1* is regulated by a SE with lineage-specific dynamics. (A-B) Overview depicting the enrichments of H3K4me3 (in dark blue and purple), H3K36me3 (in light and dark green), and H3K27ac (magenta and light blue) at the *Glis1* locus across the time points of adipogenesis (A) and osteoblastogenesis (B), respectively. The magenta and light blue bars indicate the merged SE regions identified through the analysis described in Figure 4. See also Supplementary Figure S6. (C-D) *Glis1* downregulation correlates with the signal from SE_{831}. The *Glis1* mRNA level was measured across the differentiation by RNA-seq (upper panel) and RT-qPCR (lower panel) in both adipocyte (C) and osteoblast (D) differentiation and is indicated as the intact line. Dashed line represents the signal from the SE_{831}. r = Pearson correlation co-efficient. The statistical significance for RT-qPCR measurements compared to the value on day 0 was determined by two-tailed Student's t-test. *=$p<0.05$, **=$p<0.01$ and ***=$p<0.001$. Data points represent mean of 3 biological replicates +/- SEM. AD9 sample for H3K27ac and OD15 for H3K4me3 were not included in the above analysis due to lower number of mappable high quality reads. (E) Overview depicting the enrichment of H3K27ac at the *Glis1* locus in the confluent undifferentiated (D0) and differentiated (D7) 3T3-L1 adipocyte cell line. No SE formation could be detected in these more lineage-committed cells. The data were obtained from (Mikkelsen et al., 2010). The H3K27ac enrichments at the corresponding locus in human cell types are indicated in Supplementary Figure S7B.

Figure 6

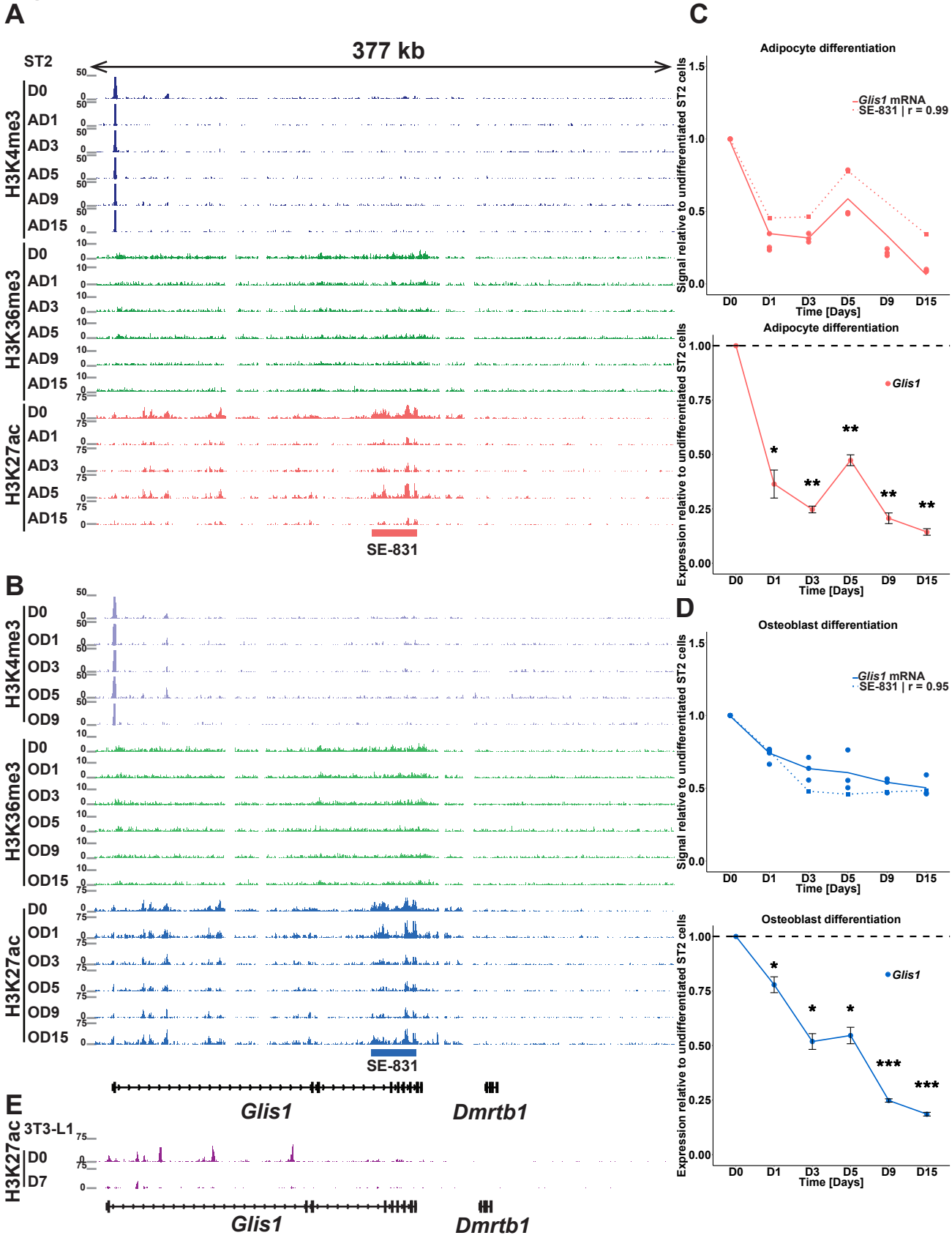
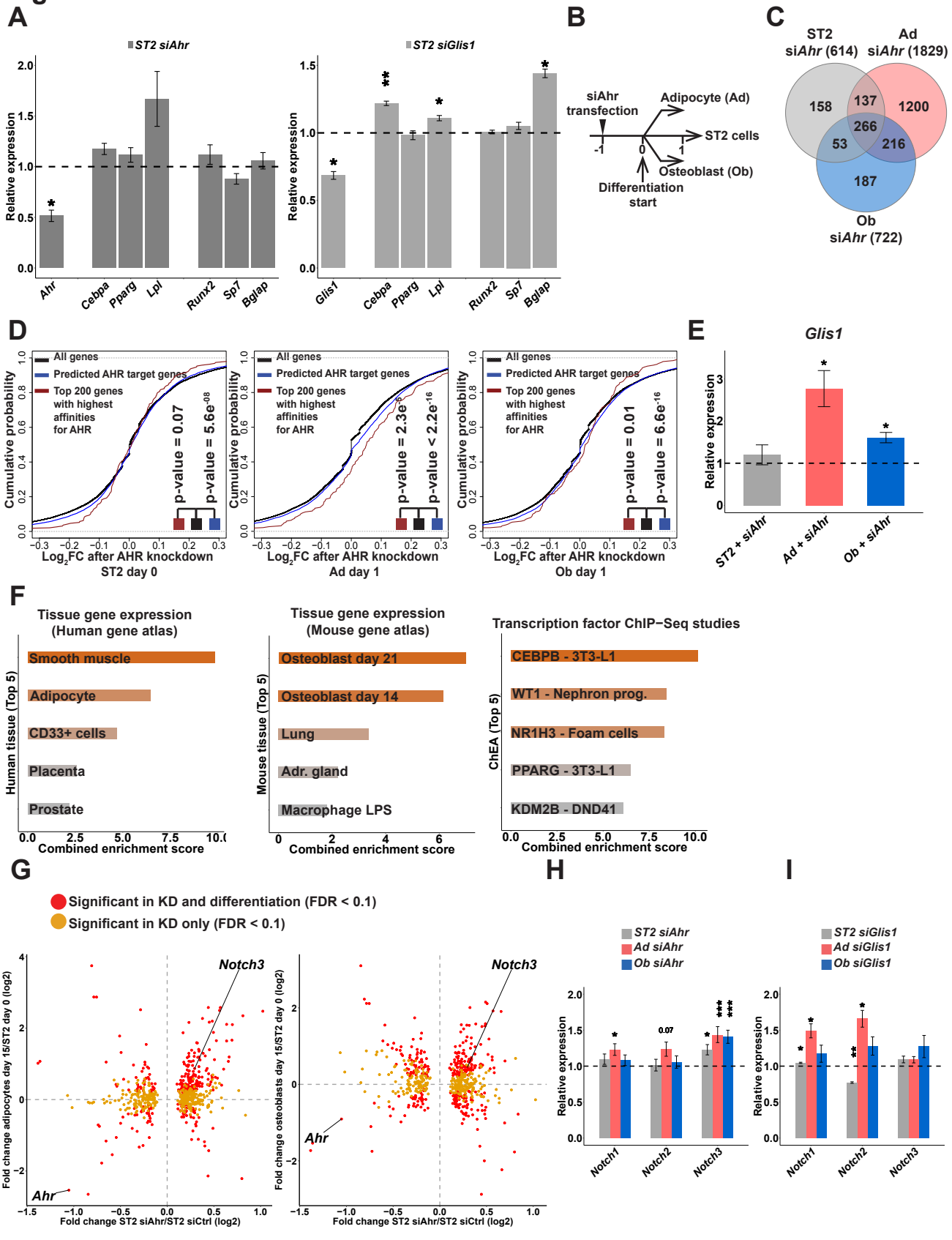

A
ST2
377 kb
H3K4me3
D0
AD1
AD3
AD5
AD9
AD15
H3K36me3
H3K27ac
SE-831
B
OD1
OD3
OD5
OD9
OD15
Glis1
Dmrtb1
C
Adipocyte differentiation
Glis1 mRNA
SE-831 | r = 0.99
Signal relative to undifferentiated ST2 cells
Time [Days]
Expression relative to undifferentiated ST2 cells
D
Osteoblast differentiation
SE-831 | r = 0.95
E
3T3-L1
D7

Figure 7. AHR and GLIS1 KDs lead to induction of lineage-specific genes and confirm EPIC-DREM predicted AHR targets. (A) Relative expression levels of differentiation marker genes following AHR or GLIS1 KD in undifferentiated MSCs are shown (y-axis). The statistical significance for RT-qPCR measurements of KD cells against cells transfected with siControl was determined by a two-tailed Student's t-test (*=p<0.05, **=p<0.01 and ***=p<0.001). Data points represent mean of 3 biological replicates +/- SEM. (B) Schematic representation of the KD RNA-seq experiments. See Supplementary Figure S8 for more details on KD efficiency. (C) Venn diagram comparing the differentially expressed genes from the three different AHR KD conditions (FDR<0.1) identified 266 genes consistently deregulated in all conditions. (D) Confirmation of EPIC-DREM predictions. Cumulative distribution of all expressed genes (black line), all EPIC-DREM predicted AHR target genes at the corresponding condition (blue line), and top 200 targets of AHR (red line) in relation to their $\log_2$ FC upon AHR depletion per condition as indicated. The significance of the increased FC of the AHR targets was confirmed with the Kolmogorov-Smirnov test and the corresponding p-values are indicated. (E) Relative expression levels of *Glis1* mRNA following AHR KD in the three indicated conditions. See panel A for more details on RT-qPCR. (F) Enrichment analysis of the 266 AHR targets identified in panel C for their preferred tissue expression profiles in human and mouse, and proteins identified as shared regulators in existing ChIP-seq studies. Top 5 most enriched hits are shown for each type of enrichment. X-axis indicates the combined enrichment score (Chen et al., 2013) and coloring the enrichment p-value. (G) Scatter plots indicating the transcriptome-wide $\log_2$ FCs as measured by RNA-seq in undifferentiated ST2 cells with reduced AHR levels (x-axis) in comparison to the $\log_2$ FCs of the same transcripts in D15 differentiated adipocytes (left panel) or osteoblasts

(right panel). The colour code separates transcripts expressed differentially only in AHR-KD (yellow), or both in KD and in differentiation (red). (H) *Notch3* is induced upon AHR downregulation in all tested conditions, as measured by RNA-seq, while (I) *Notch1* and *Notch2* are affected by GLIS1 KD. See panel A for more details on RT-qPCR. See also Supplementary Figure S8 for expression dynamics of *Notch* receptors during differentation.

Figure 7


A
ST2 siAhr
ST2 siGlis1
Relative expression
Ahr
Cebpa
Pparg
Lpl
Runx2
Sp7
Bglap
Glis1
B
siAhr transfection
Adipocyte (Ad)
ST2 cells
Osteoblast (Ob)
Differentiation start
C
ST2 siAhr (614)
Ad siAhr (1829)
Ob siAhr (722)
158
137
1200
266
53
216
187
D
All genes
Predicted AHR target genes
Top 200 genes with highest affinities for AHR
Cumulative probability
p-value = 0.07
p-value = 5.6e-08
p-value = 2.3e-5
p-value < 2.2e-16
p-value = 0.01
p-value = 6.6e-16
Log2FC after AHR knockdown
ST2 day 0
Ad day 1
Ob day 1
E
Glis1
Relative expression
ST2 + siAhr
Ad + siAhr
Ob + siAhr
F
Tissue gene expression (Human gene atlas)
Human tissue (Top 5)
Smooth muscle
Adipocyte
CD33+ cells
Placenta
Prostate
Combined enrichment score
Tissue gene expression (Mouse gene atlas)
Mouse tissue (Top 5)
Osteoblast day 21
Osteoblast day 14
Lung
Adr. gland
Macrophage LPS
Transcription factor ChIP−Seq studies
ChEA (Top 5)
CEBPB - 3T3-L1
WT1 - Nephron prog.
NR1H3 - Foam cells
PPARG - 3T3-L1
KDM2B - DND41
G
Significant in KD and differentiation (FDR < 0.1)
Significant in KD only (FDR < 0.1)
Notch3
Ahr
Fold change adipocytes day 15/ST2 day 0 (log2)
Fold change ST2 siAhr/ST2 siCtrl (log2)
Fold change osteoblasts day 15/ST2 day 0 (log2)
H
ST2 siAhr
Ad siAhr
Ob siAhr
Notch1
Notch2
Notch3
I
ST2 siGlis1
Ad siGlis1
Ob siGlis1