Temporal epigenomic profiling identifies AHR as a dynamic super-enhancer controlled regulator 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) as a mesenchymal key TF controlled by a dynamic cluster of MSC-specific super-enhancers that become repressed in both lineages. AHR represses differentiation-induced genes such as *Notch3* and we propose AHR to function as a guardian 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 (Consortium 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 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 (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 (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 MSCs towards both lineages. 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 identify aryl hydrocarbon receptor (AHR) as a central regulator of multipotent MSCs under dynamic control from 4 adjacent SEs that become repressed in the differentiated cells. The AHR repression allows upregulation of many adipocyte- and osteoblast-specific genes, including *Notch3*, a conserved developmental regulator.

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. Importantly, at genome-wide level all the histone modifications showed good correlation with the RNA-seq data across the time points (Pearson correlation co-efficients of approximately 0.5), further arguing for the reproducibility of the obtained results (Becker *et al*. in preparation). 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 5156 significantly differentially expressed genes ($\log_2FC \geq 1$, FDR<0.05) compared to the MSCs (Figure 1C; Supplementary Table S1). During osteoblast differentiation 2072 genes were dfifferentially expressed. 1401 of these genes were affected in both lineages. However, as illustrated by the top 100 genes

with highest variance across the time points and 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) (Schmidt et al., 2017). Our approach uses a novel 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 2). 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 2).

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 applied EPIC-DREM to the analysis for each timepoint of the two differentiation time series. 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 show expression in the differentiation series and associated to genes within 50 kb 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 (Gerard et al., 2017).

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 (Figure 2 and 3). Knowing the time point-specific TF-target gene interactions allows to associate split points and paths with the key TFs regulating 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 total number of TFs controlling the individual paths are indicated with the top TFs per path listed (based on their fold change (FC) and split score; see Supplementary Table S2 for all TFs per split).

Inspection of the identified key TFs confirmed many known positive (e.g. CEBPA, CEBPB, PPARG, and STAT5A/STAT5B) and negative (e.g. HES1 and GATA2) regulators of adipocyte differentiation (Farmer, 2006), to be among the TFs controlling the genes in the induced paths, while becoming up- or down-regulated themselves, respectively (Figure 3A). Interestingly, while not changing in their expression, TFs like E2F4 were annotated among those with highest split score during the first day of adipogenesis, consistently with E2F4's role as early repressor of adipogenesis whose activity is controlled by its co-repressor p130 (Farmer, 2006). At

the same time E2F7 and E2F8 seem to be associated with regulation of genes first repressed at early adipogenesis, and then upregulated during the following days (Figure 3A).

The results of the regulatory network of osteoblastogenesis confirms many known regulators such as DLX3 (Hassan et al., 2004) , DDIT3 (CHOP) (Shirakawa et al., 2006), ATF4, and FOXO1 (Long, 2011), while revealing many other factors that have not been previously associated to osteoblast differentiation (Figure 3B).

The analysis identified several TFs that could play a role in both lineages. In keeping with previous findings using the same ST2 cell line, ID4 was identified by EPIC-DREM as an activator of upregulated genes during osteoblastogenesis, while in adipocytes ID4 was controlling downregulated genes, thereby favouring the osteoblast lineage (Tokuzawa et al., 2010). Among other TFs implicated in several different splits in both lineages we found for example VDR, GLIS1, ARNT2 and AHR.

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 a list of tens of TFs. In order to further prioritize the identified main regulators, we hypothesized that the key TFs of the differentiation processes would be 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 the time. The normalized read counts under the merged SE region were collected and the relative SE signals normalized to D0 are shown in 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 S3). 120 and 79 merged SEs showed a dynamic profile ($\log_2FC \geq 1$ in at least one time point) in adipocyte and osteoblast differentiation, respectively (Figure 4C-4E). 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 S3) while most osteoblast SEs could be assigned into one of two simple profiles that either increase or decrease in signal over time (Figure 4D).

25 dynamic SEs showed changes in both differentiation processes (Figure 4E). Based on the proximity, presence of H3K4me3-marked TSSs and correlation with the expression profiles, these SEs could be assigned to 18 separate genes, several of which have already been implicated in MSCs or cell types derived from them (Figure 4E) (e.g. (Ambrosi et al., 2017) (Gu et al., 2007) (Cui et al., 2007)). However, only three of the genes, *aryl hydrocarbon receptor* (*Ahr*), *GLIS family zinc finger 1* (*Glis1*), and *Meningioma 1* (*Mn1*), encode TFs and could serve as central nodes in GRNs. Interestingly, both AHR and GLIS1 were already predicted by EPIC-DREM as regulators of multiple bifurcation events in both lineages (Figure 3), while MN1 motif is not included in the motif collection currently used by EPIC-DREM. Moreover, *Ahr* is associated with four separate merged SEs, more than any other TF in our analysis. Therefore, we next focused on *Ahr* regulation.

Ahr is controlled by multiple SEs in MSCs and repressed with lineage-specific dynamics

The four adjacent merged SEs (SE_{283}, SE_{284}, SE_{285}, and SE_{286}) at *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). All four SE regions follow similar dynamics across the time points (Figure 5A-5D) although SE_{283} did not pass 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. Similarly, the signal from all SE regions is 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. Finally, in both lineages the repression is accompanied by a decreased signal of H3K36me3 in the gene body of *Ahr*, confirming repression at transcriptional level.

While the newly identified SE cluster at the *Ahr* locus is also flanked by another active gene, *Snx13*, it is unlikely that the SEs contribute to its regulation; during the 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 of 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 S4) ("The 3D Genome Browser,").

Interestingly, the SEs controlling *Ahr* appear to be specific for the multipotent MSCs as only weak 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, (Mikkelsen et al., 2010)). Importantly, the large SE domains downstream of the *Ahr* gene could be identified also in human MSCs, but not in other inspected human cell types, suggesting that the complex regulation of *Ahr* expression in these multipotent cells could be conserved and relevant also in human development (Supplementary Figure S5).

AHR regulates mesenchymal multipotency through repression of lineage-specific genes such as Notch3

Based on the above results, we reasoned that *Ahr* could play an important role 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 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, Singh, & Bennett, 2014). We therefore tested whether the endogenous activity of AHR is important for the maintenance of the appropriate transcriptional program of the MSCs by knocking down AHR in undifferentiated cells, or those differentiated for one day towards either lineage (Figure 6A, Supplementary Figure S6). Two days following the KD total RNA was extracted and subjected to RNA-seq analysis.

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 6B, 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, were clearly shifted towards more upregulation at each condition, arguing for functional relevance of the EPIC-DREM predictions and for AHR's role as transcriptional repressor.

The number of genes significantly affected by the AHR-KD 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 6B-6D, Supplementary Table S4). The higher extent of genes affected in the differentiated cells, and especially in the adipocyte lineage, is well in keeping with the

ratio of genes normally changing in the early osteoblasto- and adipogenesis. This suggests that early changes induced by the knock-down in the undifferentiated cells are amplified in the differentiating cells. To further elucidate the role of AHR in the undifferentiated cells, we asked how the genes deregulated were behaving later in the normally differentiated D15 adipocytes and osteoblasts. The scatter plots in Figure 6C 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 colour 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 also often induced in the differentiated cell types (Figure 6C). 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, as already suggested by the EPIC-DREM analysis (Figure 3).

To better understand the function of 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 6D). 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 6E). 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) .

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 6C). In both differentiation time courses *Notch3* expression showed an anti-correlating profile compared *Ahr*, with *Notch3* becoming induced while *Ahr* levels decreased (Figure 6F). 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 S6). 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-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, that cover

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 two TFs with potentially significant roles in both lineages, *Ahr* and *Glis1* (Figure 4). In addition, dynamic SE profiling supports a role for *Mn1*. 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 data not shown), suggesting that they could play an important role in maintaining MSCs. On contrary, the SE downstream of *Mn1* shows an increase in its signal in both lineages (data not shown), suggesting it could be a general driver of differentiation. Indeed, MN1 has already

been shown to be targeted by VDR in osteoblasts and to be required for proper osteoblast differentiation (Meester-Smoor et al., 2005; Sutton et al., 2005; X. Zhang et al., 2009). Our results suggest a similar role also in adipocytes.

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). Further work will be needed to elucidate the role of GLIS1 in MSCs.

Unlike for GLIS1 and MN1, 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 (Shimb 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 maintanance 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. Analysing 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).

AHR depletion in the undifferentiated and early stage differentiated cells confirmed many of the EPIC-DREM predictions but also revealed an enrichment for lineage-specific genes among AHR targets, including *Notch3* (Figure 6). Notch signaling has been implicated in numeours developmental processes with highly diverse outcomes (Bray, 2016) and also in MSCs *Notch3* regulation is accompanied by changes in other *Notch* genes (Supplementary Figure S6). Interestingly, AHR has been recently linked to regulation of Notch signaling in mouse lymphoid cells and testis (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 (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 as a likely guardian of mesenchymal multipotency 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) 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 S5.

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 $\Delta\Delta Ct$ 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 S5.

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^{TM}-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, ("FastQC,")]. 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, ("Picard,")]. 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 ("FastQC,"). 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). 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:

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

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 2. TEPIC is available online at ("TEPIC"), DREM can be downloaded at ("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 S6. 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.

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 $\geq$ 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.

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. Scripts can be requested from authors directly.

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 performed the Western blot analysis. 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

Aguirre, A., Montserrat, N., Zacchigna, S., Nivet, E., Hishida, T., Krause, Marie N., Kurian, L., Ocampo, A., Vazquez-Ferrer, E., Rodriguez-Esteban, C., Kumar, S., Moresco, James J., Yates Iii, John R., Campistol, Josep 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:https://doi.org/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 (Pt 22)*, 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 Aging Impairs Stem Cell-Based Hematopoietic and Bone Regeneration. *Cell Stem Cell, 20*(6), 771-784.e776. doi:10.1016/j.stem.2017.02.009

Baruzzo, G., Hayer, K. E., Kim, E. J., Di Camillo, B., FitzGerald, G. A., & Grant, G. R. (2017). Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods, 14*(2), 135-139. doi:10.1038/nmeth.4106

Boitano, A. E., Wang, J., Romeo, R., Bouchez, L. C., Parker, A. E., Sutton, S. E., Walker, J. R., Flaveny, C. A., Perdew, G. H., Denison, M. S., Schultz, P. G., & Cooke, M. P. (2010). Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science, 329*(5997), 1345-1348. doi:10.1126/science.1191536

Bray, S. J. (2016). Notch signalling in context. *Nat Rev Mol Cell Biol, 17*(11), 722-735. doi:10.1038/nrm.2016.94

Cawthorn, W. P., Scheller, E. L., Learman, B. S., Parlee, S. D., Simon, B. R., Mori, H., Ning, X., Bree, A. J., Schell, B., Broome, D. T., Soliman, S. S., DelProposto, J. L., Lumeng, C. N., Mitra, A., Pandit, S. V., Gallagher, K. A., Miller, J. D., Krishnan, V., Hui, S. K., Bredella, M. A., et al. (2014). Bone marrow adipose tissue is an endocrine organ that contributes to increased circulating adiponectin during caloric restriction. *Cell Metab, 20*(2), 368-375. doi:10.1016/j.cmet.2014.06.003

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

Consortium, F., the, R. P., Clst, 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., et al. (2014). A promoter-level mammalian expression atlas. *Nature, 507*(7493), 462-470. doi:10.1038/nature13182

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

Cui, W., Cuartas, E., Ke, J., Zhang, Q., Einarsson, H. B., Sedgwick, J. D., Li, J., & Vignery, A. (2007). CD200 and its receptor, CD200R, modulate bone mass via the differentiation of osteoclasts. *Proceedings of the National Academy of*

Sciences of the United States of America, 104(36), 14436-14441. doi:10.1073/pnas.0702811104

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

DREM [http://sb.cs.cmu.edu/drem/] 10 September 2016

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

FastQC [http://www.bioinformatics.babraham.ac.uk/projects/fastqc] 05 December 2014

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

Galhardo, M., Berninger, P., Nguyen, T. P., Sauter, T., & Sinkkonen, L. (2015). Cell type-selective disease-association of genes under high regulatory load. *Nucleic Acids Res, 43*(18), 8839-8855. doi:10.1093/nar/gkv863

Galhardo, M., Sinkkonen, L., Berninger, P., Lin, J., Sauter, T., & Heinaniemi, M. (2014). Integrated analysis of transcript-level regulation of metabolism reveals

disease-relevant nodes of the human metabolic network. *Nucleic Acids Res, 42*(3), 1474-1496. doi:10.1093/nar/gkt989

Gasiewicz, T. A., Singh, K. P., & Bennett, J. A. (2014). The Ah receptor in stem cell cycling, regulation, and quiescence. *Ann N Y Acad Sci, 1310*, 44-50. doi:10.1111/nyas.12361

Gerard, D., Schmidt, F., Ginolhac, A., Schmitz, M., Halder, R., Ebert, P., Schulz, M., Sauter, T., & Sinkkonen, L. (2017).Data from: Temporal epigenomic profiling identifies AHR as a dynamic super-enhancer controlled regulator of mesenchymal multipotency Dryad Digital Repository database. doi:10.5061/dryad.r32t3

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie, A. L. M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., Challinor, M., Wallace, K., Gilmour, J., Assi, S. A., Cauchy, P., Hoogenkamp, M., Westhead, D. R., Lacaud, G., Kouskoff, V., Gottgens, B., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell, 36*(5), 572-587. doi:10.1016/j.devcel.2016.01.024

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics, 27*(7), 1017-1018. doi:10.1093/bioinformatics/btr064

Gu, D., Yu, B., Zhao, C., Ye, W., Lv, Q., Hua, Z., Ma, J., & Zhang, Y. (2007). The effect of pleiotrophin signaling on adipogenesis. *FEBS Letters, 581*(3), 382-388. doi:https://doi.org/10.1016/j.febslet.2006.12.043

Gusmao, E. G., Allhoff, M., Zenke, M., & Costa, I. G. (2016). Analysis of computational footprinting methods for DNase sequencing experiments. *Nat Meth, 13*(4), 303-309. doi:10.1038/nmeth.3772

Hassan, M. Q., Javed, A., Morasso, M. I., Karlin, J., Montecino, M., van Wijnen, A. J., Stein, G. S., Stein, J. L., & Lian, J. B. (2004). Dlx3 transcriptional regulation of osteoblast differentiation: temporal recruitment of Msx2, Dlx3, and Dlx5 homeodomain proteins to chromatin of the osteocalcin gene. *Mol Cell Biol, 24*(20), 9248-9261. doi:10.1128/mcb.24.20.9248-9261.2004

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell, 38*(4), 576-589. doi:10.1016/j.molcel.2010.05.004

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell, 155*(4), 934-947. doi:10.1016/j.cell.2013.09.053

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E., & Young, R. A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell, 58*(2), 362-370. doi:10.1016/j.molcel.2015.02.014

Huang, B., Butler, R., Miao, Y., Dai, Y., Wu, W., Su, W., Fujii-Kuriyama, Y., Warner, M., & Gustafsson, J. A. (2016). Dysregulation of Notch and ERalpha signaling in AhR-/- male mice. *Proc Natl Acad Sci U S A, 113*(42), 11883-11888. doi:10.1073/pnas.1613269113

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

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

Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes, 5*, 337. doi:10.1186/1756-0500-5-337

Long, F. (2011). Building strong bones: molecular regulation of the osteoblast lineage. *Nat Rev Mol Cell Biol, 13*(1), 27-38. doi:10.1038/nrm3254

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., & Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature, 474*(7350), 225-229. doi:10.1038/nature10106

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011, 17*(1). doi:10.14806/ej.17.1.200

Meester-Smoor, M. A., Vermeij, M., van Helmond, M. J., Molijn, A. C., van Wely, K. H., Hekman, A. C., Vermey-Keers, C., Riegman, P. H., & Zwarthoff, E. C. (2005). Targeted disruption of the Mn1 oncogene results in severe defects in

development of membranous bones of the cranial skeleton. *Mol Cell Biol, 25*(10), 4229-4236. doi:10.1128/mcb.25.10.4229-4236.2005

Mikkelsen, T. S., Xu, Z., Zhang, X., Wang, L., Gimble, J. M., Lander, E. S., & Rosen, E. D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. *Cell, 143*(1), 156-169. doi:10.1016/j.cell.2010.09.006

Naruse, M., Ishihara, Y., Miyagawa-Tomita, S., Koyama, A., & Hagiwara, H. (2002). 3-Methylcholanthrene, which binds to the arylhydrocarbon receptor, inhibits proliferation and differentiation of osteoblasts in vitro and ossification in vivo. *Endocrinology, 143*(9), 3575-3581. doi:10.1210/en.2002-220003

Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., Maurano, M. T., Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M., Roach, V., Dunn, D., et al. (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.

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

Picard [http://broadinstitute.github.io/picard] 04 July 2017.

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. L., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., The, F. C., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., Petretto, E., Forrest, A. R. 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

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., Nordström, K., Barann, M., Sinha, A., Fröhler, S., Xiong, J., Dehghani Amirabad, A., Behjati Ardakani, F., Hutter, B., Zipprich, G., Felder, B., Eils, J., Brors, B., Chen, W., et al. (2017). Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. *Nucleic Acids Research, 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.

Shirakawa, K., Maeda, S., Gotoh, T., Hayashi, M., Shinomiya, K., Ehata, S., Nishimura, R., Mori, M., Onozaki, K., Hayashi, H., Uematsu, S., Akira, S., Ogata, E., Miyazono, K., & Imamura, T. (2006). CCAAT/enhancer-binding protein homologous protein (CHOP) regulates osteoblast differentiation. *Mol Cell Biol, 26*(16), 6105-6116. doi:10.1128/mcb.02429-05

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

Sutton, A. L., Zhang, X., Ellison, T. I., & Macdonald, P. N. (2005). The 1,25(OH)2D3-regulated transcription factor MN1 stimulates vitamin D receptor-mediated transcription and inhibits osteoblastic cell proliferation. *Mol Endocrinol, 19*(9), 2234-2244. doi:10.1210/me.2005-0081

TEPIC [https://github.com/SchulzLab/TEPIC] 21 June 2017

The 3D Genome Browser [http://3dgenome.org] 06 May 2017

Tokuzawa, Y., Yagi, K., Yamashita, Y., Nakachi, Y., Nikaido, I., Bono, H., Ninomiya, Y., Kanesaki-Yatsuka, Y., Akita, M., Motegi, H., Wakana, S.,

Noda, T., Sablitzky, F., Arai, S., Kurokawa, R., Fukuda, T., Katagiri, T., Schonbach, C., Suda, T., Mizuno, Y., et al. (2010). Id4, a new candidate gene for senile osteoporosis, acts as a molecular switch promoting osteoblast differentiation. *PLoS Genet, 6*(7), e1001019. doi:10.1371/journal.pgen.1001019

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

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).

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, X., Dowd, D. R., Moore, M. C., Kranenburg, T. A., Meester-Smoor, M. A., Zwarthoff, E. C., & MacDonald, P. N. (2009). Meningioma 1 is required for appropriate osteoblast proliferation, motility, differentiation, and function. *J Biol Chem, 284*(27), 18174-18183. doi:10.1074/jbc.M109.001354

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

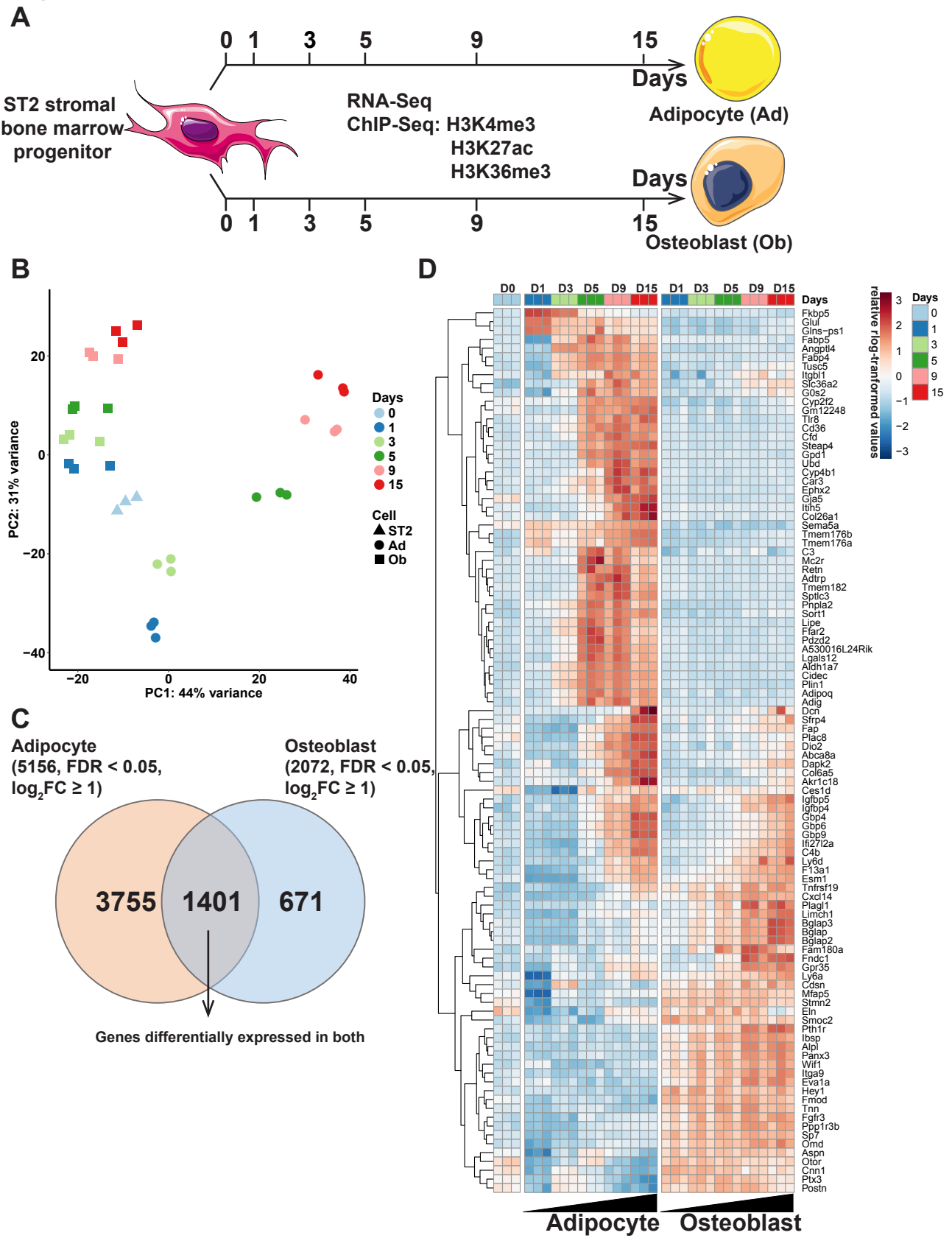
A
0 1 3 5 9 15
Days
ST2 stromal bone marrow progenitor
RNA-Seq
ChIP-Seq: H3K4me3
H3K27ac
H3K36me3
Adipocyte (Ad)
Osteoblast (Ob)
B
PC2: 31% variance
PC1: 44% variance
Days
0
1
3
5
9
15
Cell
ST2
Ad
Ob
C
Adipocyte
(5156, FDR < 0.05,
log2FC ≥ 1)
Osteoblast
(2072, FDR < 0.05,
log2FC ≥ 1)
3755
1401
671
Genes differentially expressed in both
D
D0 D1 D3 D5 D9 D15 D1 D3 D5 D9 D15
Days
relative rlog-tranformed values
Adipocyte
Osteoblast

Figure 2

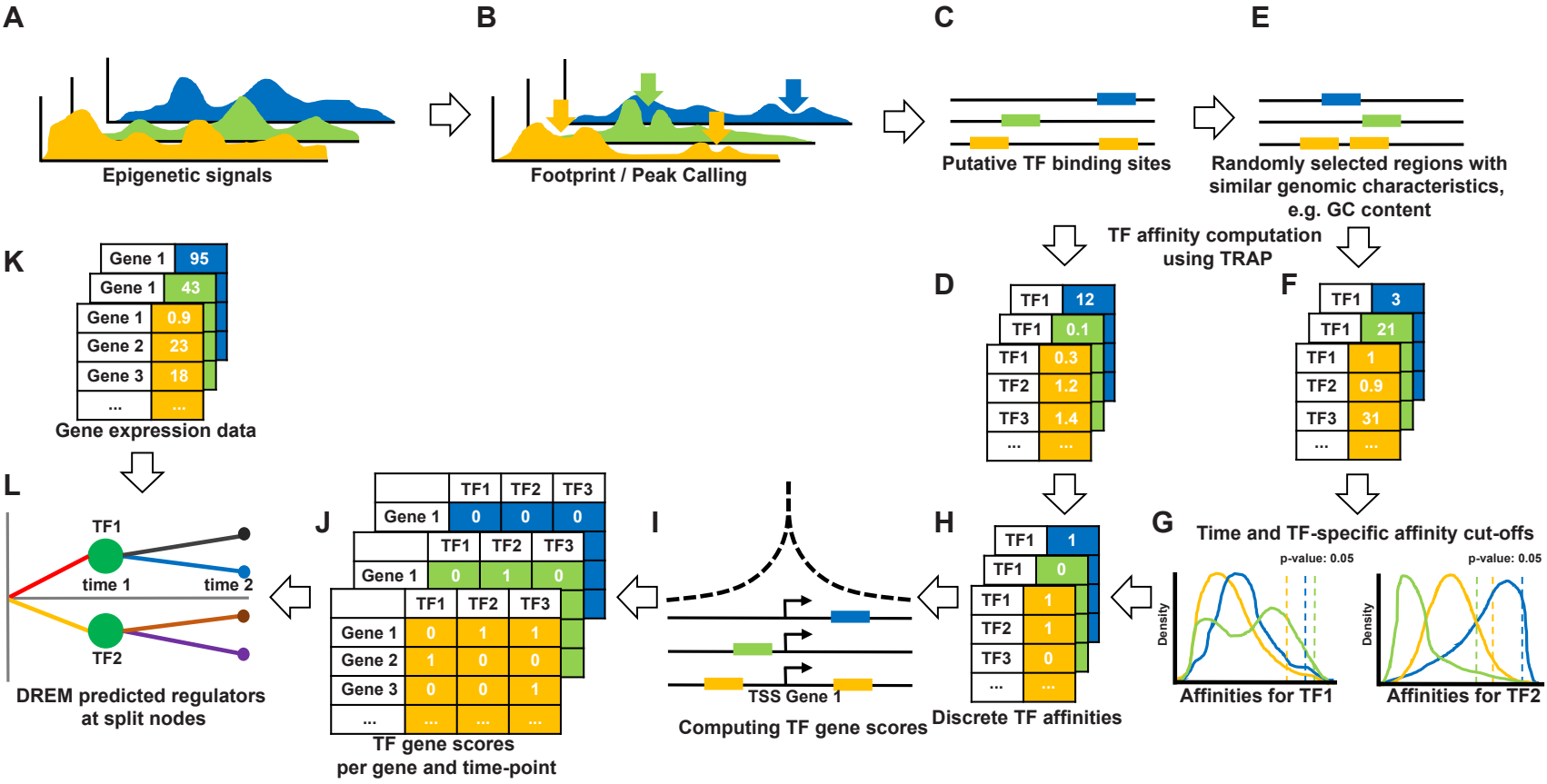
A
Epigenetic signals
B
Footprint / Peak Calling
C
Putative TF binding sites
E
Randomly selected regions with similar genomic characteristics, e.g. GC content
TF affinity computation using TRAP
K
Gene 1 95
Gene 1 43
Gene 1 0.9
Gene 2 23
Gene 3 18
... ...
Gene expression data
D
TF1 12
TF1 0.1
TF1 0.3
TF2 1.2
TF3 1.4
... ...
F
TF1 3
TF1 21
TF1 1
TF2 0.9
TF3 31
... ...
L
TF1
time 1
time 2
TF2
DREM predicted regulators at split nodes
J
TF1 TF2 TF3
Gene 1 0 0 0
Gene 1 0 1 0
Gene 1 0 1 1
Gene 2 1 0 0
Gene 3 0 0 1
... ...
TF gene scores per gene and time-point
I
TSS Gene 1
Computing TF gene scores
H
TF1 1
TF1 0
TF1 1
TF2 1
TF3 0
... ...
Discrete TF affinities
G
Time and TF-specific affinity cut-offs
p-value: 0.05
Density
Affinities for TF1
p-value: 0.05
Density
Affinities for TF2

Figure 3

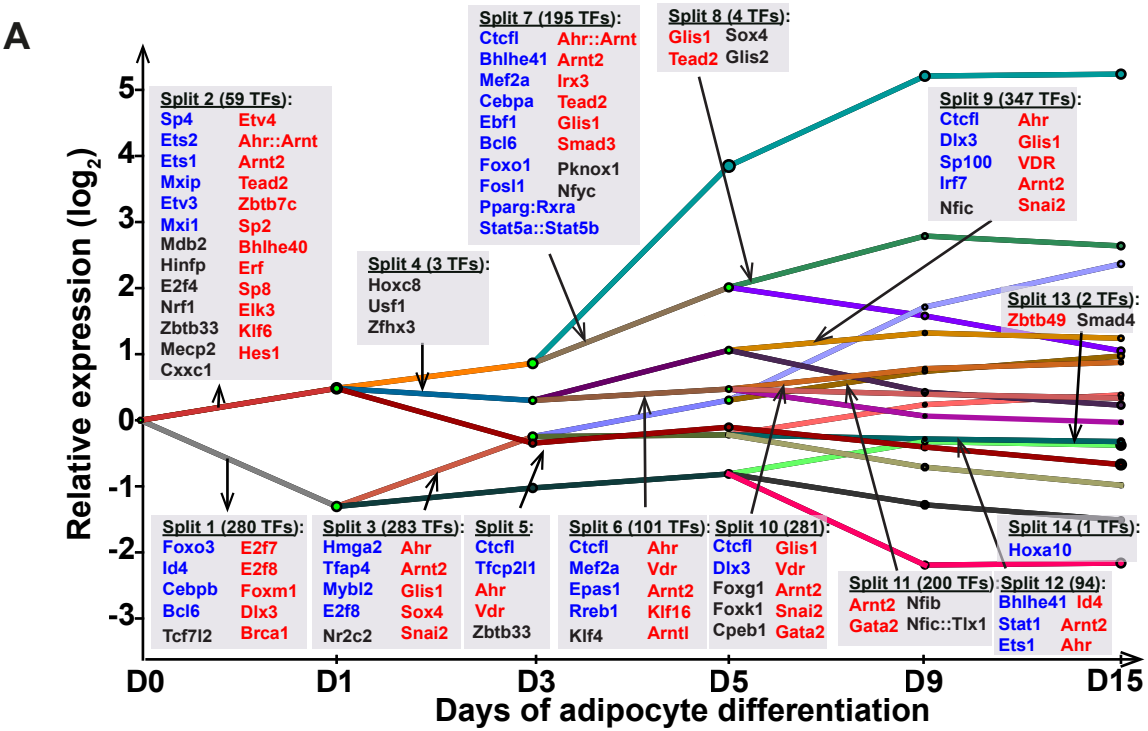
A
Split 1 (280 TFs):
Foxo3 E2f7
Id4 E2f8
Cebpb Foxm1
Bcl6 Dlx3
Tcf7l2 Brca1
Split 2 (59 TFs):
Sp4 Etv4
Ets2 Ahr::Arnt
Ets1 Arnt2
Mxip Tead2
Etv3 Zbtb7c
Mxi1 Sp2
Mdb2 Bhlhe40
Hinfp Erf
E2f4 Sp8
Nrf1 Elk3
Zbtb33 Klf6
Mecp2 Hes1
Cxxc1
Split 3 (283 TFs):
Hmga2 Ahr
Tfap4 Arnt2
Mybl2 Glis1
E2f8 Sox4
Nr2c2 Snai2
Split 4 (3 TFs):
Hoxc8
Usf1
Zfhx3
Split 5:
Ctcfl
Tfcp2l1
Ahr
Vdr
Zbtb33
Split 6 (101 TFs):
Ctcfl Ahr
Mef2a Vdr
Epas1 Arnt2
Rreb1 Klf16
Klf4 Arntl
Split 7 (195 TFs):
Ctcfl Ahr::Arnt
Bhlhe41 Arnt2
Mef2a Irx3
Cebpa Tead2
Ebf1 Glis1
Bcl6 Smad3
Foxo1 Pknox1
Fosl1 Nfyc
Pparg:Rxra
Stat5a::Stat5b
Split 8 (4 TFs):
Glis1 Sox4
Tead2 Glis2
Split 9 (347 TFs):
Ctcfl Ahr
Dlx3 Glis1
Sp100 VDR
Irf7 Arnt2
Nfic Snai2
Split 10 (281):
Ctcfl Glis1
Dlx3 Vdr
Foxg1 Arnt2
Foxk1 Snai2
Cpeb1 Gata2
Split 11 (200 TFs):
Arnt2 Nfib
Gata2 Nfic::Tlx1
Split 12 (94):
Bhlhe41 Id4
Stat1 Arnt2
Ets1 Ahr
Split 13 (2 TFs):
Zbtb49 Smad4
Split 14 (1 TFs):
Hoxa10
Relative expression (log2)
5
4
3
2
1
0
-1
-2
-3
D0
D1
D3
D5
D9
D15
Days of adipocyte differentiation
B
Split 1 (335 TFs):
Arnt2
Id2
Foxf2
Id4
Dlx3
Gata2
Hoxc13
Vdr
Itgb2
Osr1
Etv1
Arntl
Sp8
Tcf4
Etv5
Glis1
Nfkb1
Bach1
Tead2
Klf6
Split 2 (350 TFs):
Arnt2
Id2
Id4
Foxf2
Dlx1
Nr1d1
Vdr
Itgb2
Osr1
Arntl
Etv1
Sp8
Sox4
Glis1
Nr2f2
Gmeb1
Cenpb
Pbx1
Split 3 (18 TFs):
Arnt2
Glis1
Glis2
Zkscan3
Ccnt2
Klf4
Srf
Atf2
Ctcf
Glis3
Rara(var2)
Zbtb18
Tbx15
Split 4 (347 TFs):
Arnt2 Prdm4
Id2 Gata6
Vdr Foxm1
Itgb2 Hmgn3
Split 5 (3 TFs):
E2f1
Nfyc
Nfkb2
Split 6 (8 TFs):
Mybl2
Setdb1
Hltf
Creb3l1
Split 7 (31 TFs):
Atf4 Zbtb18
Nr3c1 Pitx1
Lhx6 Zscan26
Tcf12 Tcf3
Split 8 (159 TFs):
Foxf2 Vdr
Nr1d1 Hoxa6
Nfe2l1 Itgb2
Ddit3 Glis1
Bhlhe41 Ahr
Rbpj Bach1
Atf4 Hoxa5
Foxo1 Nr2f2
Split 9 (12 TFs):
Cenpb Thap1
E2f1 Mybl1
Mtf1 Sp2
Rora
Split 10 (2 TFs):
Irf1
Rara(var2)
Split 11 (13 TFs):
Tcfp2l1
Setdb1
Foxm1
Six5
Split 12 (9 TFs):
Maff
Gli3
Stat1::Stat2
Split 13 (1 TF):
Zbtb3
Split 14 (34 TFs):
Maff E2f8
Cenpb Bach1
Relative expression (log2)
5
4
3
2
1
0
-1
-2
D0
D1
D3
D5
D9
D15
Days of osteoblast differentiation

Figure 4

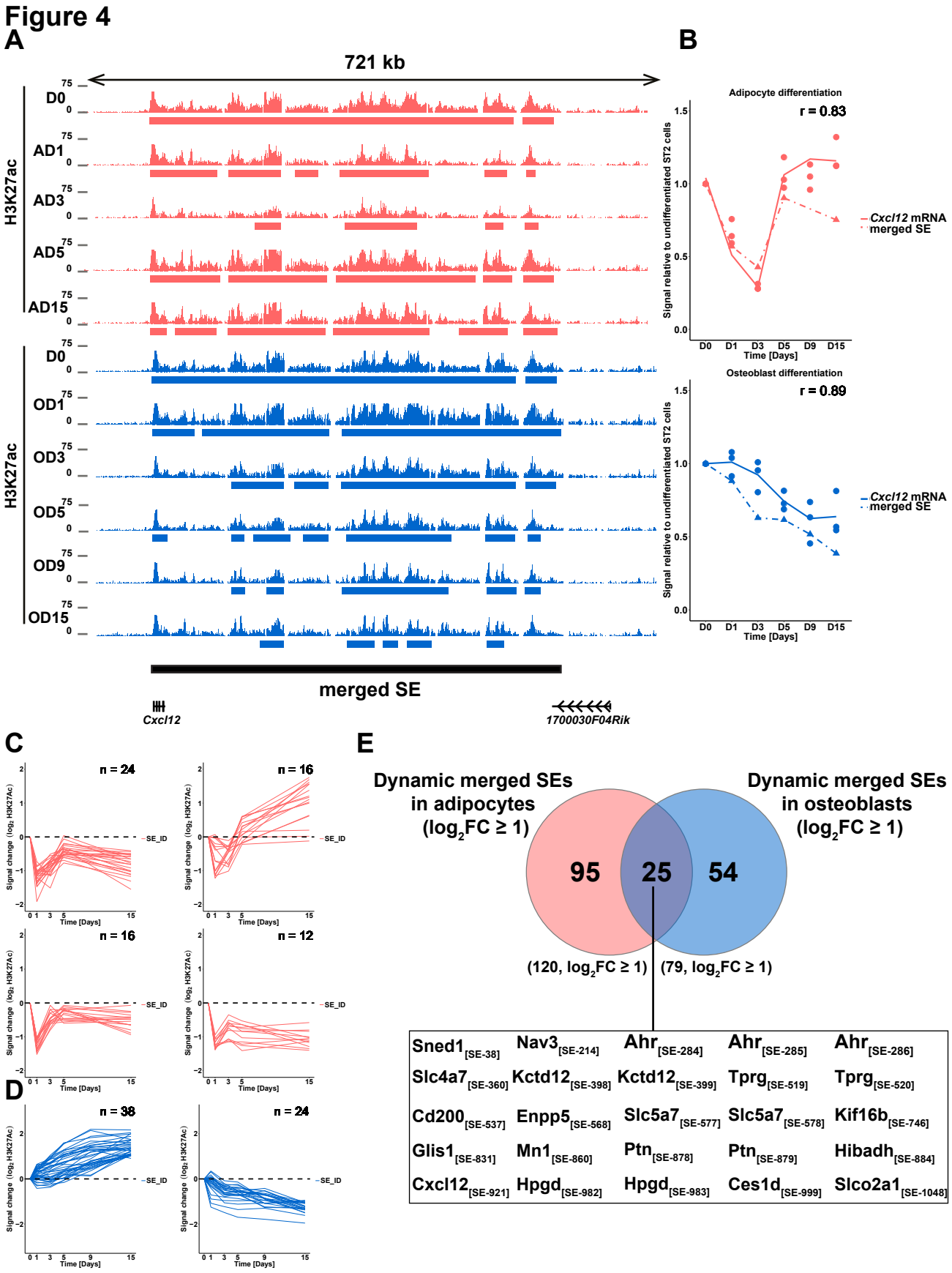
A
721 kb
H3K27ac
D0
AD1
AD3
AD5
AD15
OD1
OD3
OD5
OD9
OD15
merged SE
Cxcl12
1700030F04Rik
B
Adipocyte differentiation
r = 0.83
Osteoblast differentiation
r = 0.89
Signal relative to undifferentiated ST2 cells
Time [Days]
Cxcl12 mRNA
merged SE
C
n = 24
n = 16
n = 12
D
n = 38
SE_ID
Signal change (log2 H3K27Ac)
E
Dynamic merged SEs in adipocytes (log2FC ≥ 1)
Dynamic merged SEs in osteoblasts (log2FC ≥ 1)
95
25
54
(120, log2FC ≥ 1)
(79, log2FC ≥ 1)
Sned1[SE-38]
Nav3[SE-214]
Ahr[SE-284]
Ahr[SE-285]
Ahr[SE-286]
Slc4a7[SE-360]
Kctd12[SE-398]
Kctd12[SE-399]
Tprg[SE-519]
Tprg[SE-520]
Cd200[SE-537]
Enpp5[SE-568]
Slc5a7[SE-577]
Slc5a7[SE-578]
Kif16b[SE-746]
Glis1[SE-831]
Mn1[SE-860]
Ptn[SE-878]
Ptn[SE-879]
Hibadh[SE-884]
Cxcl12[SE-921]
Hpgd[SE-982]
Hpgd[SE-983]
Ces1d[SE-999]
Slco2a1[SE-1048]

Figure 5

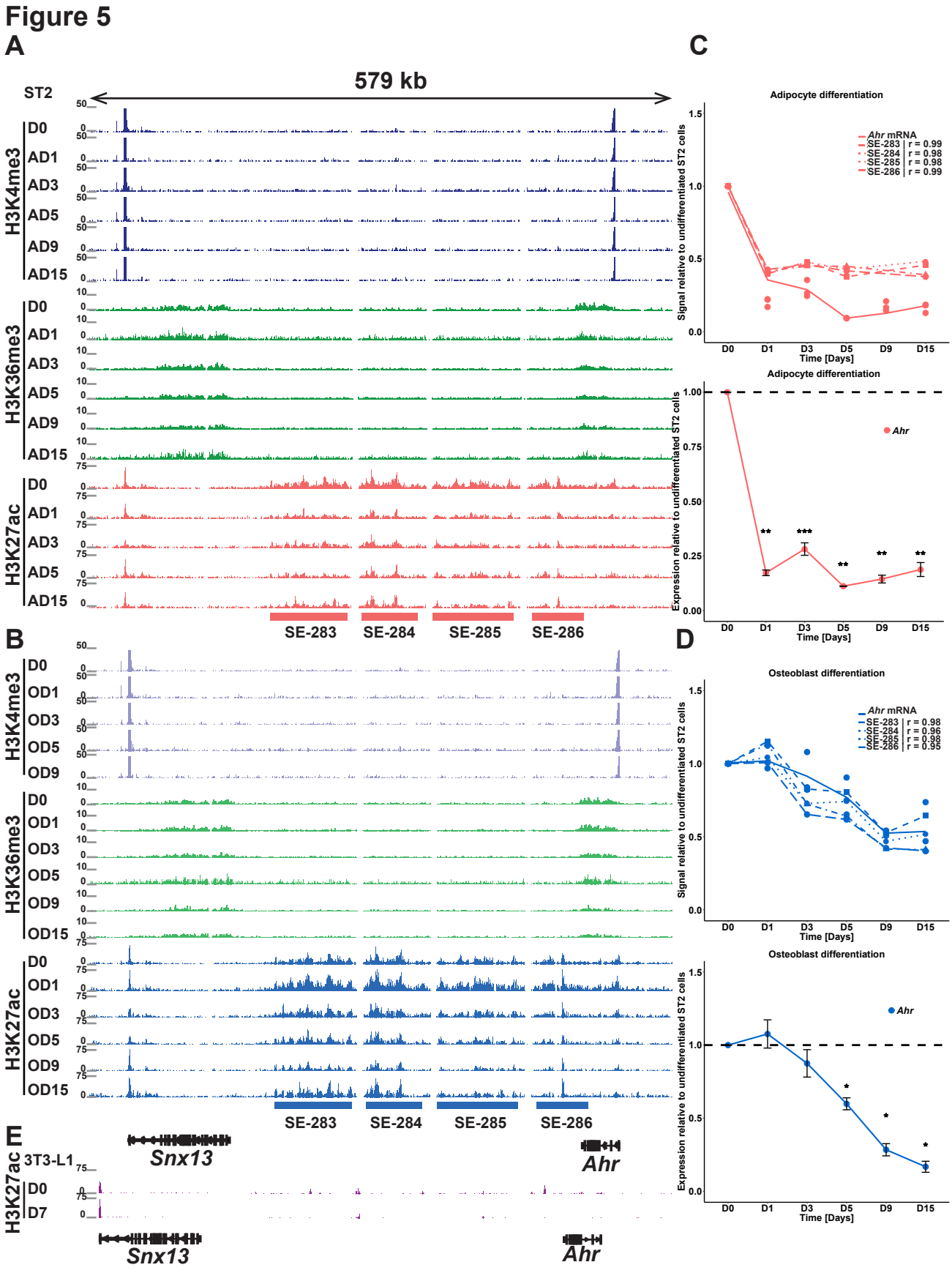
A
ST2
579 kb
H3K4me3
H3K36me3
H3K27ac
D0
AD1
AD3
AD5
AD9
AD15
SE-283
SE-284
SE-285
SE-286
B
OD1
OD3
OD5
OD9
OD15
C
Adipocyte differentiation
Ahr mRNA
SE-283 | r = 0.99
SE-284 | r = 0.98
SE-285 | r = 0.98
SE-286 | r = 0.99
Signal relative to undifferentiated ST2 cells
Time [Days]
Ahr
Expression relative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

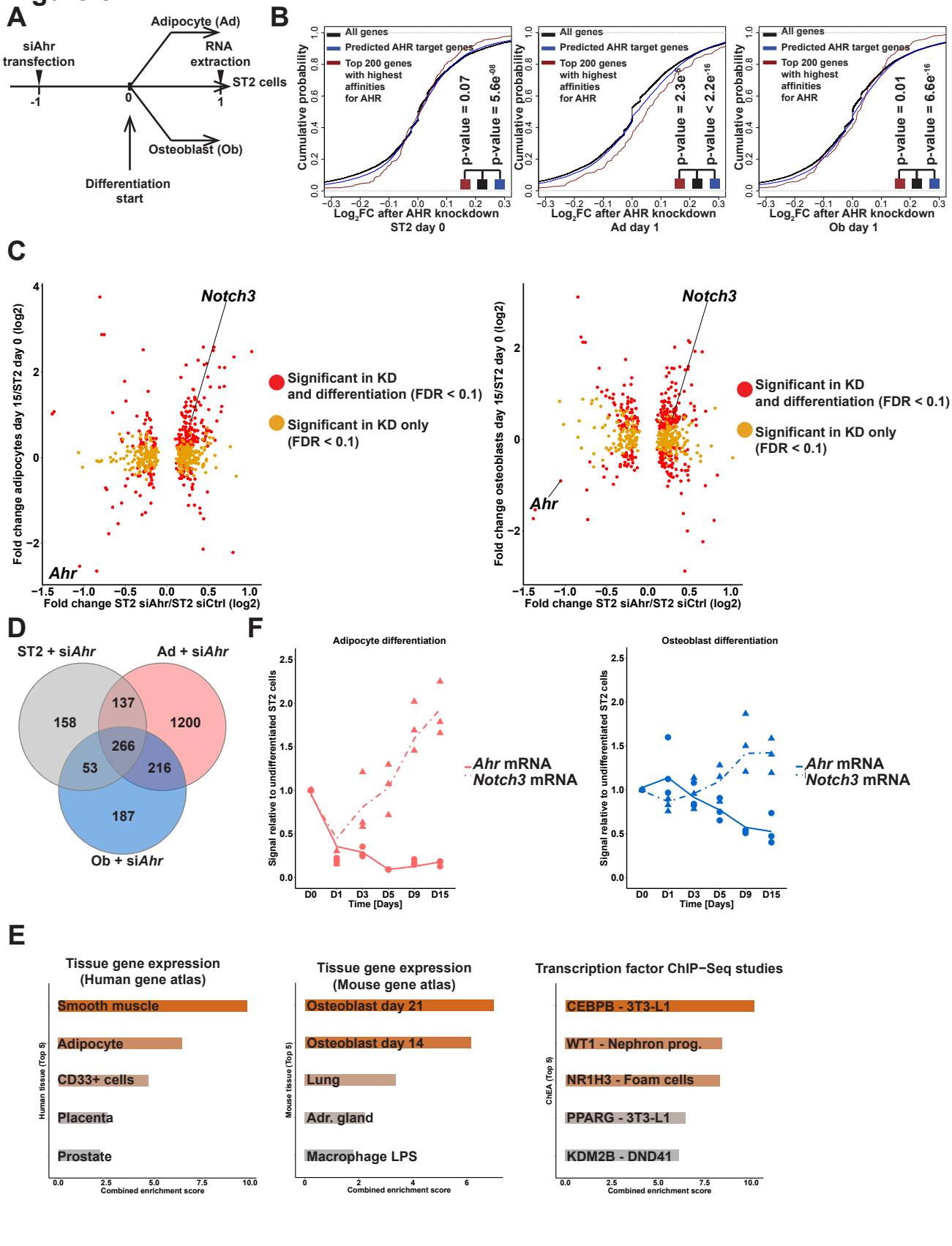
A
Adipocyte (Ad)
siAhr transfection
RNA extraction
ST2 cells
-1
0
1
Osteoblast (Ob)
Differentiation start
B
All genes
Predicted AHR target genes
Top 200 genes with highest affinities for AHR
Cumulative probability
p-value = 0.07
p-value = 5.6e-08
Log2FC after AHR knockdown
ST2 day 0
p-value = 2.3e-05
p-value < 2.2e-16
Ad day 1
p-value = 0.01
p-value = 6.6e-16
Ob day 1
C
Notch3
Ahr
Fold change adipocytes day 15/ST2 day 0 (log2)
Fold change ST2 siAhr/ST2 siCtrl (log2)
Significant in KD and differentiation (FDR < 0.1)
Significant in KD only (FDR < 0.1)
Fold change osteoblasts day 15/ST2 day 0 (log2)
D
ST2 + siAhr
Ad + siAhr
Ob + siAhr
158
137
1200
266
53
216
187
F
Adipocyte differentiation
Osteoblast differentiation
Signal relative to undifferentiated ST2 cells
Ahr mRNA
Notch3 mRNA
D0
D1
D3
D5
D9
D15
Time [Days]
E
Tissue gene expression (Human gene atlas)
Smooth muscle
Adipocyte
CD33+ cells
Placenta
Prostate
Human tissue (Top 5)
Combined enrichment score
Tissue gene expression (Mouse gene atlas)
Osteoblast day 21
Osteoblast day 14
Lung
Adr. gland
Macrophage LPS
Mouse tissue (Top 5)
Transcription factor ChIP−Seq studies
CEBPB - 3T3-L1
WT1 - Nephron prog.
NR1H3 - Foam cells
PPARG - 3T3-L1
KDM2B - DND41
ChEA (Top 5)

Figure 1. Time-series RNA-seq and ChIP-seq profiling of adipocytes and osteoblass 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 2: **Workflow of the EPIC-DREM approach.** (A-D) 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. (E-F) 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. (G) 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. (H) Applying the threshold on the TF affinities computed in (D) results in a set of discrete TF affinities per TF and timepoint. (I-J) Using TEPIC, the discrete TF affinities are integrated into discrete, timepoint specific TF-gene association scores. (J-K) 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. In panels A-J,K different time points are indicated by different colours.

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. The paths where the predicted TF-target gene interactions cover at least 30% of the genes in the upstream split point with a split score $\leq$ 0.01 are shown. The top TFs based on their time point-specific FCs and split scores are annotated in the grey boxes. The top TFs are colored depending whether they are induced (blue), repressed (red), or unchanged (black). The total number of TFs predicted to control each path is indicated and the corresponding TFs are listed in Supplementary Table S2.

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 S3. y-axis indicates the signal change in $\log_2$-scale. (E) Venn analysis of the dynamic merged SEs from both lineages identifies 25 shared dynamic SEs with $\log_2 FC \geq 1$. The identified SEs and their proximal expressed genes with the highest correlation across time are listed in the box. Adipocyte D9 sample for H3K27ac was not included in the above analysis due to low number of mappable high quality reads.

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 S5.

Figure 6. AHR-KD confirms EPIC-DREM predictions and leads to induction of Notch3 and other lineage-specific genes. (A) Schematic representation of the experimental set-up. See Supplementary Figure S6 for the efficiency of the AHR downregulation. (B) Confirmation of EPIC-DREM predictions. Cumulative distribution of all expressed genes (black line), all EPIC-DREM predicted AHR target genes (blue line), and top 200 targets of AHR (red line) in relation to their 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. (C) Scatter plots indicating the transcriptome-wide $\log_2$-FCs as measured by RNA-seq in undifferentiated ST2 cells with reduced AHR levels 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). (D) Venn diagram comparing the differentially expressed genes from the three different KD conditions (FDR<0.1) identifies 266 genes consistently deregulated in all conditions. (E) Enrichment analysis of the 266 identified AHR targets their preferred tissue expression profiles in human and mouse, and proteins identified as shared regulators in existing ChIP-seq studies. Top5 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. (F) *Notch3* is induced in parallel with *Ahr* downregulation. The *Ahr* (intact line) and *Notch3* (dashed line) mRNA levels as measured by RNA-seq are depicted across the time-series of adipocyte (left panel) and osteoblast (right panel) differentiation. The values for all biological replicates are indicated separately. See also Supplementary Figure S6 for other *Notch* receptors.