Modulation of vitamin D signaling by the pioneer factor CEBPA

Veijo Nurminen, Antonio Neme, Sabine Seuter, Carsten Carlberg
School of Medicine, Institute of Biomedicine, University of Eastern Finland, FI-70211 Kuopio, Finland

ABSTRACT

The myeloid master regulator CCAAT enhancer-binding protein alpha (CEBPA) is known as a pioneer factor. In this study, we report the CEBPA cistrome of THP-1 human monocytes after stimulation with the vitamin D receptor (VDR) ligand 1α,25-dihydroxyvitamin D3 (1,25(OH)2D3) for 2, 8 and 24 h. About a third of the genomic VDR binding sites co-located with those of CEBPA. In parallel, the binding strength of 5% of the CEBPA cistrome, i.e. some 1500 sites, is significantly (p < 0.001) affected by 1,25(OH)2D3. Transcriptome-wise analysis after CEBPA silencing indicated that the pioneer factor enhances both the basal expression and ligand inducibility of 70 vitamin D target genes largely involved in lipid signaling and metabolism. In contrast, CEBPA suppresses 82 vitamin D target genes many of which are related to the modulation of T cell activity by monocytes. The inducibility of the promoter-specific histone marker H3K4me3 distinguishes the former class of genes from the latter. Moreover, prominent occupancy of the myeloid pioneer factor PU.1 on 1,25(OH)2D3-sensitive CEBPA enhancers mechanistically explains the dichotomy of vitamin D target genes. In conclusion, CEBPA supports vitamin D signaling concerning actions of the innate immune system, but uses the antagonism with PU.1 for suppressing possible overreactions of adaptive immunity.

1. Introduction

CCAAT/enhancer-binding proteins (CEBPs) are basic region leucine zipper transcription factors that form a family of six members (α, β, γ, δ, ε and ζ) [1]. They are widely expressed in many tissues and cell types and control a number of cellular processes, such as cellular proliferation and differentiation, inflammation and metabolism [2]. The most prominent member of the CEBP family is CEBPA, which has a key role in fate decisions during myeloid differentiation [3]. For example, patients with acute myeloid leukemia frequently exhibit mutations, deregulation of expression or functional alterations of the CEBPA gene [4]. Moreover, CEBPA-knockout mice die shortly after birth due to numerous developmental defects in the hematopoietic system and other organs [5]. Thus, proper CEBPA regulation is essential for maintaining homeostasis of both embryonic and adult tissues. In the development of hematopoietic lineages CEBPA collaborates as well as antagonizes with several other lineage-specific transcription factors, such as the ETS family member PU.1 [6]. Interestingly, a systematic analysis of gene expression during all stages of hematopoiesis [7] showed that CEBPA, PU.1 as well as the nuclear receptor VDR are tightly co-expressed during monocyte and granulocyte differentiation.

CEBPA and PU.1 both act as pioneer transcription factors, i.e. they suppress possible overreactions of adaptive immunity.

Abbreviations: 1,25(OH)2D3 or 1,25D, 1α,25-dihydroxyvitamin D3; ACSL1, acyl-CoA synthetase long chain family member 1; CD14, CD14 molecule; CDA, cytidine deaminase; CEBP, CCAAT/enhancer binding protein; ChIP-seq, chromatin immunoprecipitation sequencing; CTCF, CCCTC-binding factor; DB, dihydrotransferase 1; SOM, self-organizing map; SSH1, slingshot protein phosphatase 1; TAD, topologically associated domain; TSS, transcription start site; VDR, vitamin D receptor

*Corresponding author at: School of Medicine, Institute of Biomedicine, University of Eastern Finland, POB 1627, FI-70211 Kuopio, Finland.
E-mail address: carsten.carlberg@uef.fi (C. Carlberg).

1 Equal contribution.

2 Present address: Instituto de Investigaciones en Matemáticas y en Sistemas-Mérida, Universidad Nacional Autónoma de México, Sierra Papacal, 97302, Yucatán, Mexico.

2 Present address: Institute for Cardiovascular Physiology, Medical Faculty, Goethe University, D-60590 Frankfurt, Germany.

https://doi.org/10.1016/j.bbagrm.2018.12.004
Received 10 October 2018; Received in revised form 6 December 2018; Accepted 6 December 2018
Available online 11 December 2018
1874-9399/ © 2018 Elsevier B.V. All rights reserved.
can bind genomic DNA even in the co-presence of nucleosomes [8] and help other settler transcription factors, such as VDR, to bind to their targets within chromatin. VDR has a key role in the molecular endocrinology of vitamin D, since it is the exclusive high-affinity receptor for the biologically most active vitamin D metabolite 1,25(OH)\(_2\)D\(_3\) [9,10]. Vitamin D and its receptor are well known for their impact on calcium and phosphorus homeostasis, but they also modulate innate and adaptive immune responses including differentiation of these cell types [11–15]. The combination of two lineage specifying transcription factors with a hormone-sensitive nuclear receptor promises not only detailed insight into the mechanisms of vitamin D-triggered monocytic differentiation [16], but may also be the molecular basis of blood cancer preventive effects of appropriate vitamin D supplementation [17] or even leukemia therapy [18].

Chromatin is the complex of nucleosomes and genomic DNA providing the scaffold for packaging of the whole human genome [19]. Due to the intrinsic repressive potential of chromatin, per cell type, such as THP-1 human monocytes, only some 50–100,000 chromatin regions are accessible to settler transcription factors, as measured by formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) [20]. Chromatin modifying enzymes set post-translational marks to nucleosome-forming histone proteins [21], which represents a kind of chromatin indexing [22]. Some histone marks have well understood functions. For example, H3K27ac labels active chromatin in general [23] and H3K4me3 indicates active transcription start site (TSS) regions [24,25]. Moreover, during cellular differentiation, in particular in embryogenesis, H3K4me3 is a sign of bivalent chromatin, in which genomic regions are marked for both activating and repressing chromatin modifications [26]. In this way, gene expression can respond very fast and flexibly to changing rates in the presence and activity of transcription factors, such as CEBPA, PU.1 and VDR, during differentiation and development [27].

THP-1 cells had been extensively studied in VDR signaling [28]. High quality chromatin immunoprecipitation sequencing (ChiP-seq) data are not only available for genome-wide VDR binding [29,30], but also significant (p < 0.05) effects of 1,25(OH)\(_2\)D\(_3\) on i) binding of PU.1 [31] and the chromatin organizer CCCTC-binding factor (CTCF) [32], ii) the activity of the histone mark H3K27ac [33] and iii) the accessibility of chromatin [34] were reported. These datasets led to a model, in which chromatin loops in the size a few hundred kb, so-called topologically associating domains (TADs), that are flanked by 1,25(OH)\(_2\)D\(_3\)-sensitive CTCF sites define vitamin D signaling units containing at least one vitamin D target gene and one VDR binding site [35].

This study was designed to investigate the impact of CEBPA on vitamin D signaling. In THP-1 cells we demonstrate the sensitivity of the CEBPA cistrome to 1,25(OH)\(_2\)D\(_3\) and its co-localization or overlap with the cistromes for VDR and H3K4me3 marks, respectively. Transcriptome-wide analysis after CEBPA silencing indicates that the pioneer factor enhances in monocytes the activity of 70 vitamin D target genes mainly involved in lipid signaling and metabolism but suppresses 82 genes largely related to the modulation of T cell function. Responses of H3K4me3 marks and co-occupancy of CEBPA enhancers with PU.1 serve as mechanistic explanations for the dichotomy in the gene modulatory effects of CEBPA.

## 2. Material and methods

### 2.1. Cell culture

The human monocytic cell line THP-1 [36] is a well responding and physiologically meaningful model system for the investigation of physiological processes regulated by 1,25(OH)\(_2\)D\(_3\), such as innate immunity and cellular growth [29,37–39]. THP-1 cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM l-glutamine, 0.1 mg/ml streptomycin and 100 U/ml penicillin and were kept at 37°C in a humidified 95% air/5% CO\(_2\) incubator. Prior to RNA or chromatin extraction, cells were first grown overnight in phenol red-free medium supplemented with 10% charcoal-stripped fetal calf serum and then treated with vehicle [0.1% ethanol (EtOH)] or 100 nM 1,25(OH)\(_2\)D\(_3\) (Sigma-Aldrich). Three independent replicate experiments (biological repeats) were performed for obtaining the RNA-seq and ChiPmentation/ChiP-seq datasets. Identical cell growth and treatment conditions had been used in previously published VDR ChIP-seq [30], FAIRE-seq [34], H3K27ac ChIP-seq [33], PU.1 ChIP-seq [31] and RNA-seq [34] experiments.

### 2.2. Dicer substrate siRNA (DsiRNA) silencing

THP-1 cells were transfected with either non-specific control DsiRNA oligomers (NCI) or specific DsiRNAs targeting CEBPA mRNA (Table S1). Prior to transfection, the cells were grown for one day on 6-well plates in phenol red-free RPMI 1640 medium supplemented with 10% charcoal-stripped fetal calf serum and 2 mM l-glutamine. The transfection was performed using the TransiT-siQUEST® reagent (Mirus Bio) according to the instructions by the manufacturer. First, the transfection complexes were formed in RPMI 1640 medium without phenol red, serum and antibiotics using a final concentration of 50 nM scrambled DsiRNA control or mixture of three different DsiRNA against CEBPA mRNA (Table S1). After complex formation they were added to cells. The cells were then incubated in a humidified 95% air/5% CO\(_2\) incubator for 24 h and treated with either vehicle or 100 nM 1,25(OH)\(_2\)D\(_3\) for 24 h. After treatment, the viability of the cells was determined, cells were lysed, RNA was extracted, cDNA synthesized and knockdown efficiency was determined on the mRNA level. Three independent experiments were performed, each of these biological repeats consisted of three technical repeats.

### 2.3. RNA isolation and library preparation

Total RNA was extracted using the Quick-RNA™ Miniprep kit (Zymo Research) according to the manufacturer’s instructions. RNA quality was assessed by Agilent Bioanalyzer and library preparation was performed using NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer’s instructions. After RNA extraction and purity assessment, the rRNA was depleted using NEBNext® rRNA Depletion Kit (New England Biolabs). Briefly, 800 ng of total RNA was used as starting material, to which rRNA depletion master mix containing NEBNext rRNA Depletion Solution and Probe Hybridization Buffer was added and mixed. Probes were hybridized, RNase H digestion was performed for 30 min at 37°C followed by DNase I digestion for 30 min at 37°C. RNA was purified using Agencourt RNAClean XP Beads. After rRNA depletion, libraries were prepared using the NEBNext® Ultra II Directional RNA Library Prep Kit following the manufacturer’s instructions. After adaptor ligation, samples were purified using NEBNext Sample Purification Beads. Adapter ligated DNA was then eluted and amplified by qPCR. The quality of the library was assessed using Agilent Bioanalyzer.

### 2.4. RNA-seq analysis

RNA-seq libraries were sequenced at 50 bp read length on a HiSeq2000 system using standard Illumina protocols at the Gene Core of the EMBL (Heidelberg, Germany). After quality control using afterQC [40], RNA-seq analysis was conducted by processing the high-quality reads using kallisto [41] with parameters –b 100 –single –i 180 –s 20. This software approximates abundance in a fast and efficient way via a pseudo-alignment stage to a reference genome (hg19). Differential gene expression was computed using DESeq2 [42], which implements a negative binomial test over the reads in the two conditions (treated/un-treated), with standard parameters and a p-value cutoff of 0.001.
2.5. ChIP-seq

ChIP assays were performed as described by Zhang et al. [42] with some modifications. After treatment of 2 × 10^7 THP-1 cells, nuclear proteins were cross-linked to genomic DNA by adding formaldehyde directly to the medium to a final concentration of 1% and incubating at room temperature for 10 min on a rocking platform. Cross-linking was stopped by adding glycine to a final concentration of 0.125 M and incubating at room temperature for 10 min on a rocking platform. The cells were collected by centrifugation and washed twice with ice-cold phosphate-buffered saline. The cell pellets were subsequently resuspended twice in 10 ml cell lysis buffer [0.1% SDS, 1 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, protease inhibitors, 50 mM HEPES-KOH (pH 7.5)] and once in 10 ml nuclear lysis buffer [1% SDS, 1 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, protease inhibitors, 50 mM HEPES-KOH (pH 7.5)]. After two washes with cell lysis buffer, the chromatin pellet was resuspended in 700 μl of SDS lysis buffer [1% SDS, 10 mM EDTA, protease inhibitors, 50 mM Tris-HCl (pH 8.1)] and the lysates were sonicated in a Bioruptor Plus (Diagenode) to result in DNA fragments of 200 to 500 bp. Cellular debris was removed by centrifugation. 340 μl aliquots of the lysate were diluted 1:5 in IP dilution buffer [1% Triton X-100, 2 mM EDTA, 150 mM NaCl, protease inhibitors, 20 mM Tris-HCl (pH 7.5)]. Dynabeads Protein G (60 μl, Invitrogen) were coated with 1 μg anti-H3K4me3 antibody (Merck Millipore, 07-473) overnight at 4 °C. The pre-formed bead-antibody complexes were then washed twice with beads wash buffer (0.1% Triton X-100, phosphate-buffered saline, protease inhibitors) and added to the chromatin aliquots. The samples were incubated overnight at 4 °C on a rotating wheel to form and collect the immuno-complexes. The beads were washed sequentially for 5 min on a rotating wheel with 1 ml of the following buffers each: twice cell lysis buffer, once high salt buffer [0.1% SDS, 1% Triton X-100, 1 mM EDTA, 350 mM NaCl, 0.1% sodium deoxycholate, 50 mM HEPES-KOH (pH 7.5)], once ChIP wash buffer (250 mM LiCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl (pH 8.0)) and twice TE buffer [1 mM EDTA, 10 mM Tris-HCl (pH 8.0)]. Then, the immune complexes were eluted using 250 μl ChIP elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 7.5)) at 37 °C for 30 min with rotation. The elution was repeated with a 10 min rotation and the supernatants were combined. The immune complexes were reverse cross-linked at 50 °C for 2 h in the presence of proteinase K (Merckta) in a final concentration of 40 μg/ml. The genomic DNA was isolated with the ChIP DNA Clean&Concentrator Kit (Zymo Research).

2.6. ChIPmentation

ChIPmentation was performed as described by Schmidl et al. [44] with minor modifications. 5 × 10^6 THP-1 cells were used per sample. Cross-linking, cell lysis, sonication and immuno-complex formation were identical as in ChIP. Per sample 35 μl Dynabeads Protein G were coated with 1 μg anti-CEBPA antibody (Diagenode, C15410225) or 1 μg anti-hH3K4me3 antibody. The ChIPmentation reaction was performed by re-suspending the beads into 10 mM Tris-HCl (pH 8) buffer. Beads were then collected on magnetic stand and tagmentation mix containing Tagmentation buffer (10 mM Tris-HCl, 5 mM MgCl2, 10% DMF) and Tn5 transposase (Nextera DNA Library Preparation Kit, Illumina) were added and the samples were incubated for 10 min at 37 °C. Then, Tn5 was inactivated by adding ice-cold cell lysis buffer and incubating for 5 min on ice. The complexes containing the beads and tagmented DNA were then washed sequentially twice with cell lysis buffer and TE buffer. Next, the library was eluted from the beads and reverse-cross-linked by incubation in ChIPmentation elution buffer (10 mM Tris-HCl (pH 8.0), 5 mM EDTA pH 8.0, 0.4% SDS, 300 mM NaCl) containing proteinase K for 1 h at 55 °C and 4 h at 65 °C. Then, the library was isolated with the ChIP DNA Clean&Concentrator Kit (Zymo Research). After library purification, barcoding and library amplification was performed.

2.7. ChIP-seq analysis

CEBPA and H3K4me3 ChIPmentation/ChIP-seq libraries were sequenced at 50 bp read length on a HiSeq2000 system using standard Illumina protocols at the Gene Core of the EMBL (Heidelberg, Germany). ChIP-seq data were aligned with the human reference genome version hg19 using Bowtie software version 1.1.1 [45] with the following essential command line arguments: bowtie -n 1 -m 1 -k 1 -e 70 –best. The aligned input and reads were converted to sorted BAM format using samtools [46] and, after merging the read sets per sample, converted to TDF format using igvtools, in order to allow efficient visualization in the Integrative Genomics Viewer (IGV) genome browser [47]. The quality of the datasets was also checked with the program afterQC and reads with quality below 95% were discarded. Statistically significant ChIP-seq peaks were identified using Model-based Analysis of ChIP-seq data (MACS) version 2 [48] with the following essential command line arguments: macs2 callpeak –bw 150 –keep-dup 1 –g hs –qvalue = 0.01 –m 350 –bdg. Peaks are declared, when enriched regions, compared with the corresponding loci in the input, are found. All data was automatically normalized by MACS so that different datasets, which were derived from different comparisons and offering different numbers of reads, could be compared. Otherwise, default parameters were used. Replicate ChIP-seq data were treated separately and merged at a late stage based on the broad peak output file. For differential binding between datasets of 1,25(OH)2-D3 and EtoH-treated cells, the genomic loci were checked by using PePr software (based on a negative binomial distribution), applying the -diff parameter for differential binding verification and automatic width detection [49]. CEBPA ChIP-seq, H3K4me3 ChIP-seq and RNA-seq raw data are available at Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) at GSE119556. VDR ChIP-seq data was obtained from identically treated THP-1 cells (Neme et al., manuscript in preparation), that of H3K27ac [32] are found at GSE107851, while that of FAIRE-seq and RNA-seq [34] studies are collected at GSE69903.

2.8. Statistics and machine learning tools

For inspecting the datasets traditional hypothesis tests, such as Student’s t-test and Chi-square, as well as machine learning tools [50], such as k-means and random forests, were applied. The k-means algorithm is a well-established method to analyze complex data in general and gene expression in particular [51]. In the method, partitions of data points, which are embedded in a high-dimensional attribute space, form disjoint classes. Each of these classes is chosen, in order to maximize the homogeneity of its elements, while in parallel they present as low resemblance as possible to the other classes, i.e., k-means aims to find clusters in the data. The k-means algorithm is iterative and starts by defining the number of classes (k). Each class is constructed around a centroid, which corresponds to the “center of mass” of the points belonging to that particular class. The locations of the k centroids are calculated based on the distribution of the high-dimensional data points, often referred to as vectors. An error measure of the goodness of the mapping into k classes quantifies how similar are the data points in the same class. The lower the error, the better the mapping. In general, if only one class is allowed, that is, k = 1, the error is maximal, since all elements would be located in the same class. By increasing k, the error will decrease, and in general, a good rule of thumb about the number of classes k is when the error stops decreasing, or when it decreases in a very slow fashion.

The algorithm random forests allow the identification of sets of rules that link attributes with classes. Here, the classes were those obtained by k-means quantifying the impact of CEBPA silencing on gene expression. The set of attributes is defined by markers, such as co-localization of CEBPA with the transcription factors VDR and PU.1 or
Fig. 1. The CEBPA cistrome interferes with VDR signaling. THP-1 cells were stimulated with vehicle (0.1% EtOH) or 100 nM 1,25(OH)2D3 for 2, 8 and 24 h and ChIPmentation for CEBPA and VDR has been performed. For each condition only peaks present in all three biological repeats were considered and the overlaps of the cistromes of CEBPA and VDR are displayed for 1,25(OH)2D3-stimulated cells (A). Stacked bar charts indicate the number of 1,25(OH)2D3-sensitive CEBPA sites at all three time points (B). CEBPA sites overlapping with VDR are indicated in red. A graph indicates the percentage of VDR overlap of 1,25(OH)2D3-sensitive CEBPA sites as a function of time (C). The IGV browser [47] was used to visualize examples of 1,25(OH)2D3-sensitive CEBPA sites overlapping with VDR (shaded in grey) (D). These loci are close to the 1,25(OH)2D3 target genes CD14, ITSN1 and SERINC2.
overlap with the histone markers H3K4me3 and H3K27ac, their strength (fold enrichment, FE) and inducibility (fold change, FC). Ranking variables based on random forests allows the identification of attributes that do not necessarily correlate linearly with the class. In general, random forests is an ensemble method, in which several independent decision trees are generated and trained using subsamples and a fixed number of attributes. Each tree is constructed by partitioning the data by simple binary questions over the relevant attributes. The relevance of each attribute is computed via the distribution entropy so that the answer to the binary question dissipates as much uncertainty as possible. The depth of the tree, i.e. the maximum number of questions about the data, was fixed to a small number, in order to avoid overfitting. Here, 1000 tree predictors with a maximum depth of four levels were created and the relevance of each attribute was calculated simply as the number of times that the attribute appeared in the trees, normalized over the total number of predictors. Random forests analysis was performed by using open source software based on the scikit-learn library over Python 2.7 (https://github.com/antonioneme/VDR_histone_markers).

Exploratory Data Analysis of the Geometry of the Neighborhood Density (EDAGEONed) is a new visualization tool for high-dimensional data (Neme et al., manuscript submitted). In the present study, more common algorithms, such as t-SNE, ISOMAP or PCA, were not able to generate clear clusters of vitamin D target gene classes based on epigenomic markers. Following an alternative route to data analysis, EDAGEONed explores the density of the data in the attribute space, instead of inspecting the actual location, as the majority of traditional methods do. By measuring the changes in density, the algorithm detects differences in density in different classes. EDAGEONed explores the vicinity of genes of the same class in a high-dimensional attribute space, computing among other features the relative density of data points, i.e. objects of different classes present different densities of neighboring data points. The EDAGEONed algorithm counts the number of neighbors of each data point (gene) within a certain distance in the attribute space. Then, the distance is increased and the changes in the number of neighboring vectors are calculated. The ratios of change define the basis for further inspection of the data, but moved to a different space, which is based on the densities of vectors. Thus, objects (genes) of the same class define certain metrics that are different from the corresponding to other classes, if the density of data points is indeed different, regardless of the relative position of those data points in the space [52].

3. Results

3.1. The vitamin D-sensitive CEBPA cistrome of THP-1 cells

With the aim to understand the role of the transcription factor CEBPA in vitamin D signaling, we characterized the vitamin D-sensitive CEBPA cistrome in three biological repeats by treating THP-1 cells for 2, 8 and 24 h with either vehicle (0.1% EtOH) or 100 nM 1,25(OH)2D3, extracting chromatin and performing ChIPmentation with a CEBPA-specific antibody. The ligand-stimulated samples provided 29,868, 31,336 and 28,989 CEBPA binding sites, respectively, at time points 2, 8 and 24 h (FDR 0.1%, Fig. S1A). In total, we identified 38,160 CEBPA peaks that overlapped at 22,546 loci (Table S2). By overlap we refer to binding sites with their extremes being not > 50 bp apart. For comparison, the vehicle-treated samples had a consensus of 22,142 CEBPA sites, 18,761 of which were identical to those of the 1,25(OH)2D3-stimulated samples (Fig. S1B).

The 1,25(OH)2D3-triggered CEBPA cistromes were related to genome-wide VDR binding sites that had been determined by ChIPmentation under identical experimental conditions (Neme, Seuter and Carlberg, manuscript in preparation). After 2 h 1,25(OH)2D3 stimulation only 1063 CEBPA binding loci overlapped with VDR, after 8 h already 1644 CEBPA sites co-located with the nuclear receptor for 1,25(OH)2D3 and after 24 h even 5064 CEBPA sites overlapped (Fig. 1A). These are 3.6, 5.2 and 17.5% of all CEBPA binding sites at the respective time points. There was a prominent rise of VDR binding sites from 2979 via 4110 to 17,134 within the 24 h of the experiment, which had already been described previously with a regular ChIP-seq dataset [30]. However, from the perspective of the VDR cistrome the CEBPA co-location rate was with 35.7%, 40.0% and 29.6% after 2, 8 and 24 h 1,25(OH)2D3 stimulation, respectively, reasonably constant (Table S2).

Since the ChIPmentation experiments were performed in three biological repeats, statistical testing for significant (p < 0.001) effects of 1,25(OH)2D3 stimulation on the strength of CEBPA binding could be performed. Accordingly, 1626, 1163 and 1893 CEBPA sites were found to be affected by 1,25(OH)2D3 treatment for 2, 8 and 24 h, respectively, and 140, 290 and 553 of these sites overlapped with VDR (Fig. 1B). The latter represented 8.6, 24.9 and 29.2% of all ligand-sensitive CEBPA sites, i.e. the VDR overlap rate increased over time (Fig. 1C). Interestingly, the CEBPA-VDR overlap rate did not increase proportionally to the number of VDR sites, i.e. it was more pronounced at early time points, and may indicate that CEBPA-VDR complexes target preferentially primary vitamin D target genes.

In total, 3810 CEBPA sites showed at some time points a significant (p < 0.001) response to 1,25(OH)2D3 stimulation, but there were only 138 CEBPA loci that were affected at all three time points by the VDR ligand (Fig. S1C). The VDR overlap rate (at all three time points) of these most prominently ligand-triggered CEBPA sites was with 46.4% clearly higher than that of any average CEBPA locus (2.1%). Illustrative examples of 1,25(OH)2D3-induced transcription factor binding strength and CEBPA/VDR co-location are shown for enhancer regions regulating the vitamin D target genes CD14, intersectin 1 (ITSN1) and serine incorporator 2 (SERINC2) (Fig. 1D). Interestingly, the enhancer region regulating the SERINC2 gene is composed of three closely located VDR binding sites, two of which co-located with 1,25(OH)2D3-sensitive CEBPA loci.

In summary, in 1,25(OH)2D3-stimulated THP-1 cells the CEBPA cistrome comprises some 30,000 genomic loci, up to 17% of which show co-location with VDR. In parallel, the binding strength of 4–6% of the CEBPA sites is significantly (p < 0.001) affected by the VDR ligand.

3.2. Transcriptome-wide effects of CEBPA silencing on vitamin D target genes

In order to evaluate the functional impact of CEBPA on vitamin D signaling, we performed transcriptome-wide analysis of vitamin D target genes after silencing the CEBPA gene (Fig. 2). THP-1 cells were transfected with a mixture of three DsiRNA oligonucleotides specific for the CEBPA gene, in order to block CEBPA mRNA transcription. As a reference, THP-1 cells were transfected with non-specific control DsiRNA oligonucleotides. 24 h after initiation of transfection the cells were stimulated for another 24 h with either vehicle or 1,25(OH)2D3. RNA was extracted and RNA-seq was performed. The effectiveness of the CEBPA mRNA expression silencing was found to be 89.96% (Fig. S2A). Differential gene expression analysis via the algorithm DESeq2 resulted for the reference sample in 951 genes that were significantly (p < 0.05) affected by 1,25(OH)2D3 stimulation (Fig. S2B). For comparison, in a previously reported RNA-seq analysis, which was based on 1,25(OH)2D3 treatment of THP-1 cells for 2.5, 4 and 24 h, DESeq2 resulted even in 3650 significantly (p < 0.05) modulated genes [30]. However, the unsupervised machine learning approach self-organizing map (SOM) reduced this long gene list to 587 prominent vitamin D targets, 273 of which were identical to the genes identified in the control DsiRNA-transfected cells (Table S3).

The k-means algorithm performed clustering analysis of these 273 prominent vitamin D target genes for the attributes change in i) 1,25(OH)2D3 inducibility (FC) and ii) basal expression with versus without CEBPA silencing and identified four classes (Fig. 2). The application of more than four classes does not further reduce the error, i.e. a distinction into four classes is the optimal description of the dataset.
Fig. 2. CEBPA silencing affects vitamin D target genes. THP-1 cells were transfected with a mixture of three DsiRNA oligonucleotides directed against the CEBPA gene or with the negative control DsiRNA (NC1) and after 24 h they were stimulated with vehicle (0.1% ETOH) or 100 nM 1,25(OH)2D3 (1,2SD) for another 24 h. RNA-seq was performed and 273 prominent known vitamin D target genes [30] were found to be affected by CEBPA silencing. K-means clustering sorted 270 of these genes into four classes characterized by the attributes changes in 1,25(OH)2D3 inducibility and effects on basal gene expression. Experiments had been performed in three biological repeats.

Class 1 contains 70 genes, which are characterized by having both a reduced inducibility and lower basal expression after CEBPA silencing, while the 118 genes of class 2 were neither much affected on the level of response to 1,25(OH)2D3 nor on basal activity. Thus, for the remainder of the study class 2 genes are not further considered. In contrast to class 1, the 67 genes of class 3 showed after CEBPA silencing moderately increased ligand inducibility and basal expression, while the effects on the 15 genes of class 4 were even more prominent. Finally, the genes EPB41L1, CDON and SYT3 were identified as outliers forming each a separate class and were not further considered (Table S3).

Gene ontology analysis using the webtool Enrichr [53] of the 70 genes of class 1 in relation to the 82 genes of classes 3&4 identified a clearly different functional profile (Table S4). Genes of class 1 are involved in phosphatidylinositol 3-kinase activity, response to peptidoglycan-derived muramyl dipeptides, regulation of phospholipid metabolism and lipid kinase activity and the biosynthesis of the eicosanoid lipotxin. In contrast, genes of classes 3&4 represent functions, such as assembly of the adhesion complex podosome, down-regulation of T cell proliferation and regulation of IL10 secretion as well as T cell receptor signaling.

Taken together, transcriptome-wide analysis of prominent vitamin D target genes after CEBPA silencing identified 70 genes with reduced ligand inducibility and basal expression, while a moderate or even strong increase of both attributes was observed with 67 and 15 genes, respectively. Accordingly, CEBPA supports vitamin D target genes largely involved in lipid signaling and metabolism and suppresses many genes related to the regulation of T cell activity.

3.3. CEBPA at promoter regions

To probe further the mechanistic basis of the differential modulation of vitamin D target genes of class 1 versus classes 3&4 by CEBPA, we first investigated the activity of TSS regions on a genome-wide level. For this purpose, we studied the most prominent chromatin marker for promoters, H3K4me3, by ChIPmentation/ChIP-seq in THP-1 cells that had been treated under identical conditions as for CEBPA and VDR, i.e. stimulation with 1,25(OH)2D3 for 2, 8 and 24 h. Based on three biological repeats the experiment resulted in 22,958, 21,582 and 18,532 H3K4me3-specific genomic regions at the three time points with a consensus of 18,309 loci (Fig. 3A). This suggested that > 5000 chromatin regions are specific to time points 2 and/or 8 h, i.e. they showed an early transient response to the vitamin D bolus (100 nM vitamin D).

Fig. 3. Relation of the H3K4me3 cistrome to the CEBPA cistrome. Venn diagrams indicate the overlap between the H3K4me3 cistromes obtained from THP-1 cells that had been treated for 2, 8 and 24 h with 1,25(OH)2D3 (A). All numbers of H3K4me3 marked genomic regions (Table S5) are based on ChIPmentation/ChIP-seq data from three biological repeats. The TSS overlap rates of the different categories are displayed in blue, while the number of overlapping vitamin D target genes TSS regions (TGs) is indicated in red. Genomic regions showing significant (p < 0.001) effects of 1,25(OH)2D3 on H3K4me3 marks were compared (B). The overlap between the cistromes of CEBPA and H3K4me3 is displayed from the perspective of CEBPA and H3K4me3 and the co-location with TSS regions are indicated (in absolute numbers and percentage) (C).
1,25(OH)₂D₃), but only 7.1% of them are located directly on a TSS (Table S5), i.e. they are markers enhancers rather than core promoters. For comparison, the average TSS overlap rate of H3K4me3-marked genomic regions was 55.6, 58.6 and 66.1% after 2, 8 and 24 h 1,25(OH)₂D₃ stimulation, respectively (Fig. 3A). The TSS regions of 213 of the 273 confirmed vitamin D target genes carry H3K4me3 marks (Table S5). 212, 210 and 200 of these H3K4me3-marked TSS regions were found at time points 2, 8 and 24 h, respectively (Fig. 3A).

The design of the ChiPmentation/ChIP-seq experiments allowed statistical testing and 1439, 1576 and 842 H3K4me3 regions were found to be significantly (p < 0.001) affected in their intensity by 1,25(OH)₂D₃ stimulation for 2, 8 and 24 h, respectively (Fig. 3B). In contrast to the transiently present H3K4me3 sites, the TSS overlap rate of these genomic regions was 63.8 to 80.5%, i.e. they primarily represent core promoter promoters. This differential chromatin activity was mainly (90.8 to 99.8%) observed at sites being part of the 18,309 common H3K4me3 sites (Fig. 3A). From the TSS regions of 213 vitamin D target genes, which carried H3K4me3 marks, 37, 77 and 15 showed differential histone modification abundance at time points 2, 8 and 24 h, respectively. However, only at nine genomic regions at all three stimulation time points differential H3K4me3 activity was detected (Fig. 3B).

Interestingly, 4063, 3906 and 2609 of the H3K4me3 regions after 1,25(OH)₂D₃ treatment for 2, 8 and 24 h, i.e. 14.1 to 17.7% of all, overlapped with CEBPA sites (Fig. 3C). Since H3K4me3 peaks are mostly far wider than transcription factor peaks (compare peak widths in Tables S2 and S5), this represented 7019, 7032 and 2778 CEBPA sites, i.e. 9.6 to 23.5% of all. The TSS overlap rate of these joint CEBPA-H3K4me3 sites was 36.3 to 51.3% (Fig. 3C), which, however, is lower than the average rate for this histone marker. In contrast, from the perspective of CEBPA sites the TSS overlap rate was 7.9 to 16.4%, which is clearly higher than the average rate of 2.3% for all CEBPA loci (Table S2) or even the 0.5 to 0.6% of the non-H3K4me3 overlapping CEBPA sites (Fig. 3C).

In summary, 60–69% of the H3K4me3 cistrome indicate TSS regions and some 5–7% show sensitivity to 1,25(OH)₂D₃ stimulation. 10–24% of the CEBPA cistrome overlaps with H3K4me3 markers.

3.4. Linking transcriptomic and epigenomic datasets by machine learning approaches

The vitamin D-triggered ChiPmentation/ChIP-seq time course datasets for CEBPA, VDR and H3K4me3 comprised for each site many attributes, such as binding strength (FE) and inducibility (FC). In order to identify the most relevant attributes distinguishing class 1 vitamin D target genes from those of classes 3&4, we applied the supervised machine learning algorithm random forests for H3K4me3-marked genomic regions, accessible chromatin and binding of the transcription factors VDR, CEBPA and PU.1 in a distance of up to 100 bp, 1 kb, 10 kb, 50 kb, 100 kb, 200 kb and 500 kb of the TSS of the in total 152 genes (Fig. 4A). This characterization created a data matrix (Fig. 4E) by the attributes found by random forests (Fig. 4C). Accordingly, the dichotomy of vitamin D target genes concerning their co-regulation by CEBPA was indicated by average inducibility (FC) of H3K4me3 signals at their TSS regions, average inducibility (FC) of PU.1 binding in a distance of 50 kb to their TSS, average VDR peak number in vicinity of the TSS and average strength (FE) of H3K4me3 signals in a distance of 50 kb to their TSS. Accordingly, 1105, 751 and 5330 H3K4me3 regions, i.e. 28.8% of all after 24 h stimulation, overlapped with VDR (Fig. 3A). The TSS overlap rate of these genomic loci ranged from 45.9 to 76.8%, which in average is the regular quota for H3K4me3 regions.

Some 800 bp downstream of the TSS of the class 1 vitamin D target gene sterol O-acyltransferase 1 (SOAT1) persistent CEBPA binding and H3K4me3 marking was observed (Fig. S4A). The wide H3K4me3 region overlapped with a VDR site directly at the TSS and another co-locating with the CEBPA site. At both loci prominent VDR binding occurred only after 24 h vitamin D stimulation. Two persistent CEBPA sites, which are located at the TSS and 900 bp downstream of the class 2 gene GLI1 pathogenesis related 1 (GLIPR1), overlapped with a double-peaked H3K4me3 region and VDR binding (Fig. S4B). The latter was again found only after 24 h vitamin D treatment. Close to the TSS of the class 3 gene NFKB inhibitor alpha (NFKBIA) three VDR binding regions were observed after 24 h vitamin D stimulation, one directly on the TSS and the others 1.0 and 2.4 kb downstream of it (Fig. S4C). A persistent CEBPA site overlapped with the central VDR site and a triple-peaked H3K4me3 region. At the TSS of the class 4 gene fructose-bisphosphatase 1 (FBP1), differential H3K4me3 marking was found at all three stimulation time points, but CEBPA binding only after 8 h and 24 h and VDR binding after 24 h vitamin D stimulation (Fig. S4D). In general, from 213 vitamin D target genes displaying H3K4me3 marks at their TSS regions (Table S5) some 50 overlapped with CEBPA sites at each time point (Fig. S3B) and even 122 co-located with VDR loci after 24 h stimulation (Fig. S3C). Interestingly, none of the VDR overlaps were specific to time points 2 and 8 h, i.e. most efficient response of vitamin D target genes is obtained after 24 h stimulation.

Taken together, in a distance of 50 kb from TSS regions the inducibility and strength of H3K4me3 sites as well as the inducibility of PU.1 and number of VDR sites close to the TSS showed to be the main attribute to distinguish class 1 and classes 3&4 vitamin D target genes. H3K4me3-marked target genes often show persistent CEBPA binding close to or at their TSS and overlap with VDR sites preferentially occurring 24 h after vitamin D stimulation.

3.5. CEBPA on enhancer regions

The drastic increase of VDR binding sites after 24 h of 1,25(OH)₂D₃ treatment (Fig. 1A) and the presence of VDR on TSS regions of more than half of all confirmed vitamin D target genes at 24 h (Fig. S3C) suggested focusing on the role of CEBPA on enhancer regions at this time point. Importantly, datasets of identically performed 1,25(OH)₂D₃ stimulation experiments were available for FAIRE-seq [34] and ChiP-seq for the marker of active chromatin, H3K27ac [33], and the pioneer factor PU.1 [31] (Table S2). The integration of these datasets indicated that from the 28,989 sites of the CEBPA cistrome at 24 h vitamin D stimulation 10,852 (37.4%) locate within inactive, non-accessible chromatin, since no overlap with FAIRE or H3K27ac sites was detected (Fig. 5A). In contrast, 10,435 of all CEBPA loci overlapped with FAIRE loci, 16,728 with H3K27ac regions and 14,856 with PU.1 sites. These are 36.0, 57.7 and 51.3% of all CEBPA sites. From these in total 21,004 CEBPA binding sites only 583 overlapped with TSS regions, i.e. there are > 20,000 enhancer regions in seven different configurations in open chromatin.

The enhancer category formed by 6811 CEBPA sites co-locating with accessible (FAIRE) and active (H3K27ac) chromatin as well as with PU.1 was the largest. Moreover, these genomic regions showed
with 42.7% the highest VDR overlap rate suggesting main relevance for 
mediating vitamin D signaling. Second ranking in VDR overlap (26.3%) 
were 2015 CEBPA sites that also co-located with accessible and active 
chromatin but lacked PU.1 binding. The enhancer regions of class 1 
vitamin D target genes patatin like phospholipase domain containing 1 
(PNPLA1, Fig. S5A) and cytidine deaminase (CDA, Fig. S5B) were ex-
amples of the latter category and did not show any overlap with PU.1, 
while the class 3 genes slingshot protein phosphatase 1 (SSH1, Fig. S5C) 
and Kelch domain containing 8B (KLHDC8B, Fig. SSD) represented the 
former category and showed PU.1 occupancy (for gene classification see 
Fig. 2 and Table S3). Interestingly, at all four example enhancer regions a 24 h stimulation with 1,25(OH)2D3 resulted in significantly 
(p < 0.001) increased CEBPA occupancy. Moreover, the CEBPA-
FAIRE-H3K27ac enhancer regions with and without PU.1 co-occupancy 
showed with 21.9 and 25.3% a significantly (p < 10−25, Chi-squared test) higher overlap with H3K4me3 marks than other CEBPA enhancer 
categories (Fig. 5A). A more detailed analysis of both CEBPA enhancer 
categories indicated that 17.6 and 11.2% overlapped only with 
H3K4me3 and VDR (Fig. S7). The class 1 gene insulin receptor (INSR) 
was applied (D). The method computes the density of data points in a 
high-dimensional space. Each point represents a locus, and each dimension of that 
space is an attribute. The density is calculated by counting the number of data points of the same class within a 
High-dimensional space defined by the most 
relevant attributes 

Fig. 4. Segregating CEBPA-modulated vitamin D target genes by epigenomic attributes. In order to distinguish class 1 vitamin D target genes from those of classes 3& 
4, which are characterized by different epigenomic attributes around their TSS, such as transcription factor binding and histone markers (different shapes) (A), the attributes are assembled in a data matrix (B). The machine learning approach random forests ranks the attributes by relevance (C). Since traditional mappings of this high-dimensional space to two dimensions via the algorithms SOM, t-SNE or ISOMAP did not produce clear clusters for the classes, the new algorithm EDAGEONeD was applied (D). The method computes the density of data points in a high-dimensional space. Each point represents a locus, and each dimension of that space is an attribute. The density is calculated by counting the number of data points of the same class within a fixed distance. By increasing the distance, changes in density are inspected. Data points of different classes tend to have different changes in the compute density. The different density measures (in arbitrary units) allow the distance segregation of class 1 genes from those of classes 3&4 (E) by the attributes found by random forests (C).

4. Discussion

In the multi-layered regulation of hematopoiesis, transcription fac-
tors play a key role. Since CEBPA works together with PU.1 and VDR in 
myeloid differentiation [7], we used THP-1 human monocytes to 
characterize the functional interaction of CEBPA with vitamin D sig-
nealing. In this cellular model, the CEBPA cistrome comprises some 
30,000 loci, which is in the same order as described previously for other 
monocytic cell lines [54]. Within 24 h after a 1.25(OH)2D3 bolus 
(100 nM) the VDR cistrome drastically increases in numbers from some 
1000 to 17,000 loci and > 5000 of these co-locate with CEBPA sites, i.e. 
about a third of genome-wide VDR loci are affected by the pioneer 
factor. For comparison, even some 60% of the VDR cistrome overlaps 
with PU.1 sites, but one has to consider that with some 110,000 sites 
the PU.1 cistrome in THP-1 cells [31] is 3–4 times larger than that of 
CEBPA.

Interestingly, in average some 5% of the CEBPA cistrome is sensitive
to $1,25(\text{OH})_2\text{D}_3$ treatment, i.e. genome-wide the binding intensity of the transcription factor to some 1500 sites changes in dependence of vitamin D stimulation. This means that some of the activities of the myeloid lineage-determining transcription factor CEBPA are modulated by vitamin D. Accordingly, the well-known differentiation-inducing potential of $1,25(\text{OH})_2\text{D}_3$ on monocytes and its precursors [55] could be explained, at least in part, by the effects of vitamin D on CEBPA activity. Interestingly, vitamin D showed also similar effects with PU.1 [31] and the chromatin organizer CTCF [32]. Moreover, also 5–10% of accessible chromatin sites and histone activity had been shown to depend on vitamin D [33,34]. Confirming this general observation, also 5–7% of the H3K4me3 cistrome described in this study showed vitamin D sensitivity. This indicates that not only epigenome changes, such as increased chromatin accessibility and binding of pioneer factors, affect VDR binding and vitamin D signaling, but that also in turn $1,25(\text{OH})_2\text{D}_3$ treatment affects the epigenome and pioneer factor activity.

The physiologically most important question of this study was, to what extent vitamin D signaling, i.e. vitamin D target gene expression, is modulated by CEBPA. On a transcriptome-wide perspective CEBPA silencing affects the basal expression as well as the $1,25(\text{OH})_2\text{D}_3$ inducibility of more than half (152 of 273) of all vitamin D target genes in THP-1 cells. CEBPA interferes with vitamin D signaling in two contrary ways: 70 target genes (class 1) are supported by the pioneer factor in their response to vitamin D, while possibly overboarding reactions of 82 genes (classes 3&4) to the VDR ligand are suppressed by CEBPA.

An important finding of this study is that the two sets of vitamin D target genes represent different biological processes: while CEBPA-supported genes preferentially mediate lipid signaling and metabolism, are many CEBPA-repressed genes in monocytes involved in controlling T cell responses. This dichotomy reflects well-known functions of vitamin D on the immune system, such as supporting the defense against microbe infection via innate immunity [56] and suppressing possible autoimmune reactions of the adaptive immune system [57]. Thus, CEBPA pushes in monocytes those vitamin D target genes that are needed for an effective response to bacterial infection but in parallel the pioneer factor helps to avoid that some other vitamin D target genes, via which monocytes stimulate the response of T cells, are not over-active.

A mechanistic explanation for the suppressive function of CEBPA on classes 3&4 vitamin D target genes appears to be an antagonism between CEBPA and PU.1, where CEBPA blocks the VDR supporting action of PU.1 on enhancers controlling these genes. The antagonism between the two myeloid pioneer factors has been described before in the context of myeloid differentiation [6]. Based on the results of this
study we suggest that the CEBPA-PU.1 antagonism is also the molecular basis for the differential reaction of monocytes in immune responses. While the PU.1 occupancy of CEBPA enhancers regulating vitamin D target genes of classes 3 & 4 is already obvious when inspecting the regulatory scenarios of these genes, a second mechanism explaining the target genes of classes 3 & 4 is already obvious when inspecting the H3K4me3 cistrome is more than double as high than for the later time point (24 h). This indicates a focus on the regulation of primary vitamin D target genes.

In general, H3K4me3 modifications do not only indicate active promoters, but they also represent a memory function of the epigenome [58, 59], in order to accelerate transcriptional responses. Essential for the efficient function of the innate immune system is a very rapid reaction to bacterial infections. Not only H3K4me3 marks to TSS regions of key target genes but also to chromatin regions of neighboring genes within the same TAD may memorize the information that is essential for a rapid and massive response. In this study, we showed that these H3K4me3-marked regions are not only core promoters but also CEBPA-binding enhancers.

In conclusion, CEBPA modulates vitamin D signaling by two rather different mechanisms. Urgent vitamin D-triggered reactions of the innate immune system are supported via fast (2 h) responding CEBPA-H3K4me3-marked regions are not only core promoters but also CEBPA-binding enhancers (1997) 569

341.


