The impact of the vitamin D-modulated epigenome on VDR target gene regulation

Veijo Nurminen¹, Antonio Neme¹, Sabine Seuter², Carsten Carlberg⁎

School of Medicine, Institute of Biomedicine, University of Eastern Finland, FI-70211 Kuopio, Finland

ABSTRACT

The micronutrient vitamin D significantly modulates the human epigenome via enhancing genome-wide the rate of accessible chromatin and vitamin D receptor (VDR) binding. This study focuses on histone marks of active chromatin at promoter and enhancer regions and investigates, whether these genomic loci are sensitive to vitamin D. The epigenome of THP-1 human monocytes contains nearly 23,000 sites with H3K4me3 histone modifications, 550 of which sites are significantly (p < 0.05) modulated by stimulation with the VDR ligand 1α,25-dihydroxyvitamin D3 (1,25(OH)2D3). H3K27ac histone modifications mark active chromatin and 2473 of 45,500 sites are vitamin D sensitive. The two types of ligand-dependent histone marks allow to distinguish promoter and enhancer regulation by vitamin D, respectively. Transcription start site overlap is the prime attribute of ligand-dependent H3K4me3 marks, while VDR co-location is the top ranking parameter describing 1,25(OH)2D3-sensitive H3K27ac marks at enhancers. A categorization of 1,25(OH)2D3-sensitive histone marks by machine learning algorithms - using the attributes overall peak strength and ligand inducibility - highlights 260 and 287 regions with H3K4me3 and H3K27ac modifications, respectively. These loci are found at the promoter regions of 59 vitamin D target genes and their associated enhancers. In this way, ligand-dependent histone marks provide a link of the effects of 1,25(OH)2D3 on the epigenome with previously reported mRNA expression changes of vitamin D target genes. In conclusion, the human epigenome responds also on the level of histone modifications to 1,25(OH)2D3 stimulation. This allows a more detailed understanding of vitamin D target gene regulation.

1. Introduction

Chromatin is a complex of nucleosomes around which genomic DNA is wrapped and provides the scaffold for the packaging of the entire human genome [1]. In general, chromatin has an intrinsic repressive potential, in order to conserve the epigenetic landscape of a differentiated cell. Thus, by default the epigenome largely restricts the access of transcription factors to promoter and enhancer regions leaving only some 50–100,000 accessible chromatin regions per cell type [2]. Chromatin modifiers are a large family of > 100 enzymes that are primarily located in the nucleus, where they perform post-translational modifications, such as acetylation and methylation, of nucleosome-forming histone proteins [3]. These histone marks represent a kind of chromatin indexing, which is often summarized as the histone code [4]. Some of the best understood histone modifications are H3K4me3, which indicates active transcription start site (TSS) regions [5,6], and H3K27ac, which labels active chromatin in general [7]. Some of these histone marks stay stable during cell divisions and can last for multiple generations, while others are highly transient [8]. Chromatin modifiers do not bind directly to genomic DNA but use transcription factors, such as VDR, to be directed to genomic regions, at which they change the local histone modification pattern.

Abbreviations: 1,25(OH)2D3 or 1,25D, 1α,25-dihydroxyvitamin D3; CD14, CD14 molecule; ChIP, Chromatin immunoprecipitation; ChIP-seq, ChIP sequencing; CLMN, Calmin; CTCF, CCCTC-Binding factor; FAIRE-seq, Formaldehyde-assisted isolation of regulatory elements sequencing; FC, Fold change; FE, Fold enrichment; FDR, False discovery rate; GABPA, GA binding protein transcription factor alpha subunit; IGV, Integrative Genomics Viewer; LPO, Lactoperoxidase; LD, Ligand-dependent; MACS, Model-based Analysis of ChIP-Seq data; NCOA, Nuclear receptor co-activator; NCOR1, Nuclear receptor co-repressor 1; PBS, Phosphate-buffered saline; RNA-seq, RNA sequencing; SOM, Self-organizing map; TAD, Topologically associating domain; TMEM37, Transmembrane protein 37; TMC06, Transmembrane and coiled-coil domains 6; 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: Institute for Cardiovascular Physiology, Medical Faculty, Goethe-University, D-60590 Frankfurt, Germany.

http://doi.org/10.1016/j.bbagrm.2018.05.006
Received 13 February 2018; Received in revised form 21 May 2018; Accepted 21 May 2018
Available online 03 July 2018
1874-9399/ © 2018 Elsevier B.V. All rights reserved.
The micronutrient vitamin D has pleiotropic physiological functions, since it not only regulates calcium and phosphorus homeostasis but also modulates innate and adaptive immune responses [9–11]. For example, the transcription factor VDR, the high-affinity nuclear receptor for the biologically most active vitamin D metabolite 1,25(OH)2D3 [12,13], is together with the pioneer transcription factor PU.1 a regulator of monocyte differentiation [14]. In the physiological background of the latter, high quality genome-wide data are available from 1,25(OH)2D3-stimulated THP-1 human monocytes on i) the binding of VDR [15], PU.1 [16] and the 3D chromatin organizer CTCF [17] via chromatin immunoprecipitation sequencing (ChIP-seq), ii) chromatin accessibility as measured by formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) [18] and iii) gene expression as assessed via RNA-sequencing (RNA-seq) [18]. Interestingly, at a few thousand sites these large-scale epigenomic datasets are significantly (p < 0.05) modulated by 1,25(OH)2D3 stimulation. All these data contributed to the “chromatin model of vitamin D signaling” [19], in which topologically associating domains (TADs), i.e. chromatin loops in the size a few hundred kb, were defined by flanking 1,25(OH)2D3-sensitive CTCF binding sites [17]. These TADs form vitamin D signaling units that contain at least one vitamin D target gene and a prominent VDR binding site, the activity of which is fine-tuned by PU.1. The combined epigenomic changes then result in looping of vitamin D inducible enhancers towards accessible TSS regions on the same TAD and finally stimulate the expression of vitamin D target genes in the respective chromatin unit. In addition, within these TADs also ligand-modulated sites of chromatin accessibility are observed that do not co-locate with VDR, PU.1 or CTCF sites. This indicates that additional epigenomic mechanisms may apply in the regulation of vitamin D target genes, such as the involvement of additional pioneer factors like GABPA [20] or the activity for chromatin modifiers [21].

In general, vitamin D and its receptor VDR were found to have a number of different effects on the epigenome [22]. For example, VDR interacts with members of the NCOA family of histone acetyltransferases [23] and with co-repressor proteins, such as NCoR1 [24], that mediate the contact with histone deacetylases. The present study was designed to get genome-wide insight on the effects of vitamin D on the epigenome by studying 1,25(OH)2D3-triggered changes in H3K4me3 and H3K27ac modification patterns, i.e. it focuses on dynamic effects of the VDR ligand on both types of histone modifications. Interestingly, from the 23,000 regions carrying H3K4me3 marks and the 45,500 sites with H3K27ac marks, 550 and 2473, respectively, are significantly (p < 0.05) modulated by 1,25(OH)2D3, demonstrating that the human epigenome responds to vitamin D also on the level of histone modifications. These ligand-dependent histone marks are found at promoter and enhancer regions of 59 vitamin D target genes allowing a refinement of the chromatin model of vitamin D signaling.

2. Material and methods

2.1. Cell culture

The human acute monocytic leukemia cell line THP-1 [25] is a well responding and physiologically meaningful model system for the investigation of 1,25(OH)2D3-triggered physiological processes, such as innate immunity and cellular growth [26–29]. 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% CO2 incubator. Prior to chromatin extraction, cells were first grown overnight in phenol red-free medium supplemented with 5% charcoal-stripped fetal calf serum and then treated with vehicle (0.1% ethanol (EtOH)) or 100 nM 1,25(OH)2D3 (Sigma-Aldrich). Three independent replicate experiments (biological repeats) were performed for obtaining the ChIP-seq datasets. Identical cell growth and treatment conditions had been used in previously published in VDR ChIP-seq [15], FAIRE-seq [18] and RNA-seq [18] experiments.

2.2. ChIP

ChIP assays were performed as described by Zhang et al. [30] with some modifications. After treatment of 2 x 107 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 glycin 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 (PBS). 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) or 1 µl anti-H3K27ac antibody (Merck Millipore, 07-449) overnight at 4°C. The pre-formed bead-antibody complexes were then washed twice with beads wash buffer (0.1% Triton X-100, PBS, 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 (Fermentas) in a final concentration of 40 µg/ml. The genomic DNA was isolated with the ChIP DNA Clean&Concentrator Kit (Zymo Research).

2.3. ChIP-seq analysis

2–10 ng of each ChIP DNA template was used for library preparation with the NEEBNext Ultra II kit (New England Biolabs) and 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). The number of read counts is indicated in Table S1. ChIP-seq data were aligned with the human reference genome version hg19 using Bowtie software version 1.1.1 [31] with the following essential command line arguments: bowtie -n 1 -m 1 -k 1 -e 70 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 [33]. Statistically significant ChIP-seq peaks were identified using Model-based Analysis of ChIP-Seq data (MACS) version 2 [34] with the following essential command line arguments: macs2 callpeak –bw 150 –keepdup 1 –g hs –qvalue = 0.01 –m 5 50 –bdg. Peaks are declared, when enriched regions, compared with the corresponding loci in the input, are found.
Otherwise, default parameters were used. Replica 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)₂D₃ 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 [35]. H3K4me3 and H3K27ac ChIP-seq raw data are available at GEO (www.ncbi.nlm.nih.gov/geo) at GSE107851. The raw data of previously published VDR ChIP-seq [15] in identically treated THP-1 cells are available at GSE89431, while that of FAIRE-seq and RNA-seq [18] studies is collected at GSE69303. The identification of vitamin D-sensitive TAD regions was published by Neme et al. [17]. All data are assembled in Table S2.

2.4. Statistics and machine learning tools

Descriptive statistics as well as machine learning tools were applied to examine the datasets. Linear regression helped to correlate the changes in histone marks and the mRNA of the nearby genes. The difference in the distribution of vitamin D target gene subsets was tested by applying a Wilcoxon-Mann-Whitney test (U test), i.e. non-parametric hypothesis testing, so that for this test the criteria of normality of data distribution is not required.

Ranking variables based on their relevance is important, in order to differentiate parameters that can explain the studied phenomena in contrast to non-informative attributes. Linear regression is a common attribute ranking method, which is based on maximizing the variables with the highest R² values, but the method fails to capture non-linear relations. Therefore, the random forests approach was applied [36], in order to rank the relative relevance of attributes describing histone marks. Random forests constitute a machine learning ensemble method, in which several independent decision trees or predictors are created and trained using subsets of the samples 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 was computed via the Gini correlation index 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 over fitting. Here, 800 tree predictors with a maximum depth of four levels were created and the relevance of each attribute was calculated as the number of times that an 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 sklearn library over Python 2.7 (https://github.com/antonioneme/VDR_histone_marks).

2.5. Vector quantization

Genomic regions carrying ligand-dependent histone marks were characterized by several attributes, such as peak strength and overlap with VDR binding sites, accessible chromatin (as measured by FAIRE-seq) or a TSS (Table S2). These attributes define high-dimensional descriptions that require particular algorithms for a meaningful and comprehensive characterization. Vector quantization algorithms, such as self-organizing map (SOM) [37], find average descriptions of similar data points, and via an iterative process, create clusters of the high-dimensional data, which allow both a visual as well as statistical inspections. SOM is an unsupervised algorithm that allows the identification of representative profiles of the inspected data. The output of a SOM is a summary of the hundreds or thousands of input data points in a two-dimensional lattice, where each node is a class or category. Here, a 3 × 3 SOM was created accounting for nine different categories. In this way, hundreds of possible changes of peak enrichment after ligand treatment are summarized into subsets of similar profiles. SOMs are created on open source software based on Python 2.7 (https://github.com/antonioneme/VDR_histone_marks).

3. Results

3.1. Epigenome-wide view on TSS regions and active chromatin in THP-1 cells

For a genome-wide monitoring of H3K4me3 histone marks, three independent ChIP-seq experiments were performed with THP-1 cells that had been treated for 24 h with either vehicle (EtOH) or 1,25(OH)₂D₃. The exclusive focus on genomic regions providing a significant (false discovery rate (FDR) < 0.1%) enrichment of the H3K4me3 signal over input control (in all three biological repeats as well as at both treatment conditions, i.e. in six independent samples) highlighted 22,998 genomic regions (Table S2). In parallel, under identical conditions a triplicate ChIP-seq experiment was conducted for H3K27ac histone modifications and resulted in 45,578 loci (Table S2). The datasets of both types of histone marks overlapped in 19,512 genomic regions, i.e. in 84.4% of all H3K4me3 sites (Fig. 1A). Interestingly, 12,794 of these overlapping genomic regions (65.6% of all) were assigned with a TSS. In contrast, the TSS overlap rate of the 3486 and 26,066 loci carrying exclusively H3K4me3 and H3K27ac marks, respectively, was only 37.6% and 1.8%.

For comparison, the overlap of all H3K4me3 marks with sites of accessible chromatin, as determined previously via triplicate FAIRE-seq in THP-1 cells treated for 24 h with 1,25(OH)₂D₃ [18], was with 18,692 genomic regions (84.4% of all, Fig. 1B) slightly lower than observed with H3K27ac modifications. Moreover, also the respective TSS overlap rate was with 62.1% (11,613 sites) a bit reduced in reference to H3K27ac modifications. For comparison, only 66.0% of chromatin regions with H3K27ac marks (30,060 sites) overlapped with accessible chromatin and just 37.5% (11,283 sites) of these sites co-located with a TSS region (Fig. S1A).

For a more detailed analysis of the effects of 1,25(OH)₂D₃ on histone modifications, the overlap of all H3K4me3 marks with VDR binding
sites was determined, as obtained previously from triplicate ChIP-seq in THP-1 cells treated for 24 h with 1,25(OH)2D3 [15]. On 4071 genomic regions H3K4me3 marks overlapped with VDR binding sites, 2490 (61.2%) of which co-located with a TSS (Fig. 1C). For comparison, 6653 regions with a H3K27ac mark (14.6% of all) also bound VDR, but only 38.1% (2532 sites) of them were found on a TSS (Fig. S1B). Importantly, the majority (58.9%) of all VDR binding sites in THP-1 cells are located within active chromatin. The complete set of 11,657 known VDR binding loci in THP-1 cells had previously been segregated into 510 persistent sites ("present at all time points"), 2109 are transient sites ("present not at all time points") and 9038 sites occurring only after 24 h ligand treatment ("24 h only") [15]. In total, 6924 loci of the VDR cistrome overlap with any of the two histone marks, 293 of which on persistent sites, 1213 are transient sites and 5418 sites occur only after 24 h ligand treatment (Table S2). On 93 of the persistent sites and 150 of the transient sites a stimulation of THP-1 cells with 1,25(OH)2D3 had a significant (p < 0.05) effect on VDR binding, i.e. in total 5661 (81.8%) of the overlapping VDR sites were affected by 1,25(OH)2D3 treatment.

In summary, the chromatin of THP-1 cells carries nearly 23,000 H3K4me3 marks and some 45,500 H3K27ac marks. H3K4me3 marks overlap with active chromatin (84.4% of all) and accessible chromatin (81.3%) and showed >62% overlap with TSS regions. At >4000 of the H3K4me3 regions also VDR was found.

3.2. 1,25(OH)2D3 significantly affects histone mark strength

Three independent replicate ChIP-seq experiments - both for ligand- and vehicle-treated cells - allowed the identification of genomic regions, in which histone mark enrichment was statistically significantly (p < 0.05) affected by 1,25(OH)2D3 stimulation. In total, 550 regions with H3K4me3 marks and 2473 sites with H3K27ac marks were found to be modulated by the VDR ligand (Fig. 2A). Surprisingly, only at 26 loci ligand-dependent H3K4me3 and H3K27ac sites overlapped. This is far less than the 84.4% and 42.8% co-location of the complete set of both makers (Fig. 1A) suggesting that the ligand dependence of each of the two histone marks represents different biological functions. Interestingly, 252 of the 550 ligand-dependent H3K4me3 marks (45.9%), but only 151 of the 2473 ligand-dependent H3K4me3 marks (6.1%) overlapped with a TSS (Fig. 2A). These percentages are clearly lower than the TSS rates of the respective complete sets of both histone modifications (61.3% and 29.1%).

Furthermore, only 51 of the 550 ligand-dependent H3K4me3 marks (9.3%) overlapped with 1,25(OH)2D3-sensitive accessible chromatin regions [18] (Fig. 2B). Similarly, only 163 the 2473 ligand-dependent H3K27ac marks (6.6%) overlapped with ligand-dependent accessible chromatin regions (Fig. S2A). In contrast, 147 of the 550 regions with ligand-dependent H3K4me3 marks bound VDR (Fig. 2C), i.e. the VDR overlap rate increased from 17.7% to 26.7%. In parallel, ligand-dependent H3K27ac marks showed an increase in VDR overlap rate from 14.6% to 24.5% (Fig. S2B).

Taken together, 2.4% and 5.4% of the genomic regions marked by H3K4me3 and H3K27ac modifications, are significantly modulated by 1,25(OH)2D3 treatment. Compared to the complete sets of both histone marks, the overlap between their ligand-dependent subsets or with 1,25(OH)2D3-sensitive accessible chromatin as well as with TSS regions was clearly reduced, while the rate of VDR co-location increased.

3.3. Categorization of genomic regions carrying ligand-dependent histone marks

In order to identify the most important parameters describing ligand-dependent H3K4me3 and H3K27ac marks, the machine learning approach random forests relevance ranking was performed by using the following ten attributes: overlap with i) TSS, binding sites for the transcription factors ii) VDR, iii) PU.1 and iv) CTCF, v) all and vi) ligand-dependent accessible chromatin, vii) VDR binding motifs, localization viii) in a cluster of histone marks or ix) within a ligand-dependent TAD and x) distance to the closest histone mark. For H3K4me3 marks random forests indicated as top ranking attributes overlap with i) TSS regions, ii) accessible chromatin and iii) VDR (Fig. 3A), while for H3K27acmarks VDR overlap was top ranked followed by that with accessible chromatin and TSS regions (Fig. 3B). Accordingly, from the 550 regions with ligand-dependent H3K4me3 marks those 252 loci overlapping with a TSS region seemed to be more important than the 141 sites showing VDR co-location. In contrast, from the 2743 ligand-dependent H3K27ac marks those 606 sites, which are co-localizing with VDR, were more critical than the 151 loci overlapping with a TSS.

A genome-wide view on the accumulated peak enrichment of the regions containing ligand-dependent H3K4me3 marks (Fig. S3A) and the sites with ligand-dependent H3K27ac marks (Fig. S3B) indicated for both types of histone modifications no obvious clustering to any chromosomal region. However, a wide peak strength distribution range was observed, which was further investigated by the machine learning approach SOM using the attributes accumulated peak enrichment in the presence and absence of VDR binding as well as their fold change. A 2 × 2 lattice (Fig. 4A) categorized the 550 ligand-dependent H3K4me3 peaks into i) SOM group 1 containing 48 rather prominent peaks being clearly induced by ligand (in average 2.6-fold), ii) 120 less prominent peaks showing only an average 1.5-fold induction (SOM group 2), iii)
3.4. Linking ligand-dependent histone marks to vitamin D target genes

The TSS regions of 461 of 587 vitamin D target genes, which had previously been identified by triplicate RNA-seq [18], overlapped with H3K4me3 and/or H3K27ac marks (Table S2). However, only 28 and 29 vitamin D target genes, respectively, have exclusively either a ligand-dependent H3K4me3 or H3K27ac mark at their TSS region, while the genes CD14 molecule (CD14) and lactoperoxidase (LPO) were highlighted even by both 1,25(OH)2D3-dependent histone marks on their promoters (Table S2, see also examples provided in Fig. 5). Interestingly, these in total 59 genes were exclusively up-regulated by vitamin D stimulation and 57 of them are located in vitamin D-modulated TADs [15]. 42 of these genes are primary vitamin D targets and 17 secondary. Thus, the average inducibility of genes carrying ligand-dependent histone marks was higher (4.4-fold for both histone marks) than for vitamin D target genes that do not have ligand-dependent histone marks on their TSS regions (2.0-fold for 402 genes, Fig. S4). Furthermore, analysis of the relation between the inducibility of vitamin D target genes and the fold change of ligand-dependent histone modifications on their TSS regions described for both histone modifications a trend of a positive correlation, which was for H3K27ac (r² = 0.4, Fig. S5B) more convincing than for H3K4me3 (r² = 0.09, Fig. S5A).

The identification of ligand-dependent histone marks and their classification by SOM significantly extended the understanding of regulatory scenarios of vitamin D target genes. From the 30 regions of ligand-dependent H3K4me3 marks that overlap with the TSS regions of vitamin D target genes, 14 and 10 were in SOM groups 1 and 4, but only 4 and 2 in groups 2 and 3 (Table S2). Moreover, 195 (75.0%) of the regions with ligand-dependent H3K4me3 marks of SOM groups 1 and 4 were located within vitamin D-triggered TADs and from the 1,25(OH)2D3-dependent H3K27ac marks in SOM groups 1 and 2 this applied even to 226 loci (78.7%, Table S2). Only 11 of the latter overlapped with TSS regions, while the remaining 215 sites may represent important vitamin D-triggered enhancers being located within 172 different TAD regions, i.e. a few TADs contain more than one vitamin D-sensitive, H3K27ac-marked enhancer.

The SOM categorization of histone marks suggested at least three different gene regulatory scenarios at TSS regions i) both ligand-dependent H3K4me3 and H3K27ac marks, ii) only ligand-dependent H3K4me3 marks or iii) exclusively ligand-dependent H3K27ac marks, as illustrated by the following examples. The 83.7-fold induced CD14 gene carries both ligand-dependent H3K4me3 and H3K27ac marks (being members of their respective SOM groups 1) on its 4.8 kb wide promoter region and its TSS co-locates with a ligand-inducible VDR site (Fig. 5A). The CD14 gene is regulated by an enhancer region some 25 kb downstream of its TSS carrying ligand-dependent H3K27ac marks (SOM group 1) and two ligand-induced VDR sites. The genomic region of the CD14 gene is the only vitamin D-sensitive, H3K27ac-marked enhancer within the respective TAD (id486 [17], Table S2). For comparison, the TSS of the vitamin D target gene transmembrane and coiled-coil domains 6 (TMCO6), which locates only 6 kb upstream of the TSS of the CD14 gene, is marked only by ligand-insensitive histone modifications and the gene is induced only 2.1-fold. A different gene regulatory scenario is represented by the 13.8-fold induced transmembrane 37 (TMEM37) gene, which has an 8.1 kb wide promoter region marked by ligand-inducible H3K4me3 (SOM group 4) and two ligand-induced VDR sites that, however, do not overlap with the TSS (Fig. 5B). Some 30 kb upstream of its TSS a narrow enhancer region is marked by ligand-dependent H3K27ac marks (SOM group 1), the only vitamin D-sensitive, H3K27ac-marked enhancer within the TAD cluster (id325/326)) and a single ligand-induced VDR site. A third example is the 28.7-fold induced calmin (CLMN) gene, which has a rather short promoter (2.1 kb) carrying ligand-dependent H3K27ac marks (SOM...
group 1) but no VDR site (Fig. 5C). The gene is regulated by an enhancer region located some 6 kb downstream of its TSS, which is marked by ligand-dependent H3K27ac modifications (SOM group 2, the closer of two vitamin D-sensitive, H3K27ac-marked enhancers within the respective TAD (id178)) and two ligand-induced VDR sites.

Taken together, SOM allowed categorizing ligand-dependent H3K4me3 and/or H3K27ac marks at promoter regions of 59 vitamin D target genes as well as 215 vitamin D-sensitive, H3K27ac-marked enhancers within 172 TADs. Ligand-dependent histone marks correlated with differential expression of vitamin D target genes.

4. Discussion

This study describes in high accuracy - enabled by the consensus of each six ChIP-seq datasets - the location of the histone modifications H3K4me3 and H3K27ac within the epigenome of THP-1 monocytes. More than 60% of the nearly 23,000 sites carrying H3K4me3 modifications mark genomic regions with a TSS. This proves previous reports [6] on the biological significance of this histone modification. In addition, the approximately 45,000 sites with H3K27ac modifications label active chromatin, such as 32,000 putative enhancers and some 13,000 active promoter/TSS regions. Although H3K4me3 modifications are rather specific indicators for active TSS regions, its co-occurrence with H3K27ac marks increases the accuracy in the prediction of
Fig. 5. Histone marks define regulatory scenarios of vitamin D target genes. THP-1 cells were treated for 24 h with vehicle (EtOH) or 1,25(OH)$_2$D$_3$ (1,25D) and ChIP-seq for H3K4me3 (purple), H3K27ac (green) and VDR (red [15]) and FAIRE-seq (blue [18]) were performed in three biological repeats. The IGV browser [33] was used to visualize the enhancer and promoter regions of the primary vitamin D target genes CD14 (A), TMEM37 (B) and CLMN (C). Gene structures are shown in blue, their change in gene expression after 24 h 1,25(OH)$_2$D$_3$ treatment was measured by triplicate RNA-seq [18]. Promoter and enhancer regions are shaded in grey and their respective SOM classification is indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
actively transcribed genes. This confirms the expectation that both histone modifications provide more specific information about active chromatin than previous assessments of accessible chromatin via FAIRE-seq highlighting in THP-1 cells nearly 67,000 genomic regions [18]. Nevertheless, accessible chromatin confirms H3K4me3 marks nearly as accurate as H3K27marks do, i.e. there is no major difference between active and accessible chromatin in labeling TSS regions. Although histone modification ChIP-seq data as well as FAIRE-seq data can be found for a number of cellular models of the ENCODE project [2], only a few studies have yet taken advantage of these data. One of them, which is directly comparing histone modifications with accessible chromatin, had been performed in breast cancer cells and confirms the association of FAIRE-seq data with marks of active chromatin [38].

An important finding of this study is the observation that subsets of 550 sites with H3K4me3 modifications and of 2743 regions with H3K27ac marks are significantly (p < 0.05) modulated in their intensity by 1,25(OH)2D3, i.e. they are dependent on VDR ligand treatment. Under identical stimulation conditions, a similar observation was made concerning a significant change in the accessibility of chromatin [18]. This means that FAIRE-seq data do not only confirm histone marks of active chromatin but also their ligand inducibility. To our knowledge, this study is the first reporting that a cellular perturbation, such as a stimulation with 1,25(OH)2D3, has effects on the epigenome both on the level of chromatin accessibility and histone modifications.

Although the example of the TSS region of the prominent vitamin D target gene CD14 indicates an overlap of both ligand-dependent H3K4me3 and H3K27ac marks, this applies only to 26 genomic loci. Thus, there is no statistically relevant indication that an overlap of both ligand-dependent histone modifications marks most up-regulated genes. From the 4587 previously described ligand-modulated sites of accessible chromatin in THP-1 cells [18], only 51 and 163, respectively, overlapped with ligand-sensitive H3K4me3 and H3K27ac marks, as exemplified by the promoter and enhancer region of the TMEM37 gene. A similar observation on the ligand-dependent increase of H3K4me3 marks on TSS regions of vitamin D target genes had already been made in SK-BR-3 human breast cancer cells [39]. This means that the majority of the ligand-dependent sites of open chromatin seem to label different aspects of the 1,25(OH)2D3 sensitivity of the human genome, such as binding sites for PU.1, CTCF and other transcription factors. Thus, both 1,25(OH)2D3-sensitive histone modifications label different epigenetic modules, such as being a promoter or enhancer region.

In mouse models of bone tissue, the Pike team used a set of histone modifications, including H3K4me3, as marks for epigenomic changes during perturbation of cells with signaling molecules, such as nuclear receptor ligands, and by cellular differentiation [40–42]. Some of these studies used stimulation periods far longer than 24 h and lack a rigorous statistical analysis, but they show similar principles as this study, i.e. the molecular effects of a stimulation with 1,25(OH)2D3 can be monitored via changes of the epigenome, such as histone modifications.

In accordance with these observations, relevance ranking assigned the attribute “TSS overlap” as the most relevant description of ligand-dependent H3K4me3 marks, while “VDR overlap” was found as top characteristic of ligand-dependent H3K27ac marks. For both types of chromatin modifications co-location with accessible chromatin showed to be second ranking and overlap with either VDR or TSS was third ranking, respectively. Thus, ligand-dependent H3K4me3 marks label 252 promoter regions as being sensitive to vitamin D, while ligand-dependent H3K27ac marks highlight 606 enhancer regions that are modulated by 1,25(OH)2D3. Although 49 of these regions overlap with a TSS in the presence (see CD14 gene) or absence (see CLMN gene) of a ligand-modulated H3K4me3 mark, the majority of them label classical enhancer regions many kb upstream or downstream of the vitamin D target gene that they regulate. SOM classification allowed focusing on 260 regions with H3K4me3 marks and 287 H3K27ac sites. From the latter, 215 are traditional enhancers located within 172 different vitamin D-triggered TADs [17]. This means that ligand-sensitive histone modifications provide important information for understanding the gene regulatory scenarios of vitamin D target genes.

Since ChIP-seq peaks of histone modifications represent histones of many nucleosomes in a row, they are far wider and complex than those of transcription factors, such as VDR, or accessible chromatin determined by FAIRE-seq. Accordingly, histone modification peaks label whole promoter and enhancer regions that each may contain many transcription factor binding sites. In this way, ligand-dependent H3K4me3 and H3K27ac marks provide a different quality to genome-wide assessment of the epigenome than that of single transcription factors. In total 59 vitamin D genes were identified carrying 1,25(OH)2D3-sensitive histone marks on their TSS region. All of these genes are up-regulated in their expression after 1,25(OH)2D3 stimulation suggesting that ligand-sensitive histone modifications are specific to up-regulated genes. Moreover, 57 of the 59 genes are located within 1,25(OH)2D3-sensitive TADs [17]. In combination with the above mentioned 215 enhancer regions, the presence of regions with ligand-dependent histone modifications turns out to be a more accurate descriptor of the regulatory scenarios of vitamin D target genes than the previously reported presence of persistent VDR binding sites [15].

In conclusion, the observation of ligand-dependent histone modifications adds a new level to the current chromatin model of vitamin D signaling [19]. These histone marks label whole regions, such as promoters and enhancers, and not individual transcription factor binding sites, i.e. they provide additional information to refine the existing model.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2018.05.006.

Conflict of interests

The authors declare to have no conflict of interests.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgement

Kind thanks to the Gene Core Facility at the EMBL in Heidelberg, Germany, for massively parallel sequencing services.

Funding

This work was supported by the Academy of Finland (#267067) and a stipend of the Doctoral School in Molecular Medicine of the University of Eastern Finland to VN.

References


