Accepted Manuscript

Selective regulation of biological processes by vitamin D based on the spatio-temporal cistrome of its receptor

Antonio Neme, Sabine Seuter, Carsten Carlberg

PII: S1874-9399(17)30136-0
DOI: doi:10.1016/j.bbagrm.2017.07.002
Reference: BBAGRM 1164

To appear in: *BBA - Gene Regulatory Mechanisms*

Received date: 29 April 2017
Revised date: 22 June 2017
Accepted date: 12 July 2017

Please cite this article as: Antonio Neme, Sabine Seuter, Carsten Carlberg, Selective regulation of biological processes by vitamin D based on the spatio-temporal cistrome of its receptor, *BBA - Gene Regulatory Mechanisms* (2017), doi:10.1016/j.bbagrm.2017.07.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Selective regulation of biological processes by vitamin D based on the spatio-temporal cistrome of its receptor

Antonio Neme, Sabine Seuter and Carsten Carlberg[#]

School of Medicine, Institute of Biomedicine, University of Eastern Finland,

FI-70211 Kuopio, Finland

[#]**Corresponding author**: Prof. Carsten Carlberg

School of Medicine

Institute of Biomedicine

University of Eastern Finland

POB 1627

FI-70211 Kuopio

Tel.: +358-40-355-3062

E-mail: carsten.carlberg@uef.fi

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

The transcription factor vitamin D receptor (VDR) is the exclusive nuclear target of the biologically active form of vitamin D (1,25(OH)$_2$D$_3$). In THP-1 human monocytes we obtained a highly accurate VDR cistrome after 2 and 24 h ligand stimulation comprising more than 11,600 genomic loci, 78% of which were detected exclusively after 24 h. In contrast, a group of 510 persistent VDR sites occurred at all conditions and some 2,100 VDR loci were only transiently occupied. Machine learning and statistical analysis as well as a comparison with the re-analyzed B cell VDR cistrome indicated a subgroup of 339 highly conserved persistent VDR sites that were suited best for describing vitamin D-triggered gene regulatory scenarios. The 1,25(OH)$_2$D$_3$-dependent transcriptome of THP-1 cells comprised 587 genes, 311 of which were primary targets with main functions in the immune system. More than 97% of the latter genes were located within 1,25(OH)$_2$D$_3$-modulated topologically associated domains (TADs). The number of persistent and transient VDR sites was found to be the main discriminator for sorting these TADs into five classes carrying vitamin D target genes involved in distinct biological processes. In conclusion, specific regulation of biological processes by vitamin D depends on differences in time-dependent VDR binding.
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABBREVIATIONS

1,25(OH)$_2$D$_3$ or 1,25D	1α,25-dihydroxyvitamin D$_3$
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP sequencing
CTCF	CCCTC-binding factor
DR3	direct repeat spaced by 3 nucleotides
FAIRE-seq	formaldehyde-assisted isolation of regulatory elements sequencing
FC	fold change
FDR	false discovery rate
GO	gene ontology
GOrilla	Gene Ontology enRIchment anaLysis and visuaLizAtion
HOMER	Hypergeometric Optimization of Motif EnRichment
IGV	Integrative Genomics Viewer
MACS	Model-based Analysis of ChIP-Seq data
RNA-seq	RNA sequencing
SOM	self-organizing map
TAD	topologically associated domain
TSS	transcription start site
VDR	vitamin D receptor

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

INTRODUCTION

Vitamin D is well known for its role in the regulation of calcium and phosphorus homeostasis controlling bone formation, but the nuclear hormone also modulates innate and adaptive immune responses [1-3]. Sub-nanomolar concentrations of the biologically active form of vitamin D, 1α,25-dihydroxyvitamin D_3 (1,25(OH)$_2$D$_3$), bind and activate the nuclear receptor VDR [4]. VDR is the exclusive nuclear target of 1,25(OH)$_2$D$_3$, i.e. the molecular endocrinology of vitamin D is based on this unique transcription factor. The canonical model of nuclear receptor signaling suggests that VDR binds as a heterodimer with retinoid X receptor to sites formed by a direct repeat of two RGKTSA (R = A or G, K = G or T, S = C or G) motifs spaced by three nucleotides (DR3) [5, 6]. In fact, the method chromatin immunoprecipitation (ChIP) combined with massive parallel sequencing (ChIP-seq) demonstrated on a genome-wide level that below the summits of VDR peaks DR3-type sequences are the most prominent binding motifs [7-11].

In human cellular systems VDR ChIP-seq has been reported for i) B cells (GM10855 and GM10861) [7], ii) monocytes (THP-1) [8], iii) macrophages (lipopolysaccharide-differentiated THP-1) [9], iv) colon cancer cells (LS180) [10] and v) hepatic stellate cells (LX2) [11]. A harmonized re-analysis of these datasets indicated more than 23,000 unique VDR binding loci throughout the whole human genome and demonstrated that ligand stimulation increases the number of DNA-bound VDR proteins in average by a factor of 2.5 [9]. In parallel, the method Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq) determined in THP-1 cells 62,000 accessible chromatin loci, nearly 9,000 of which are significantly ($p < 0.05$) modulated by 1,25(OH)$_2$D$_3$ [12]. Although the exact molecular mechanisms of these vitamin D-triggered epigenome changes are not fully understood, it is obvious that they are secondary consequences of genome-wide VDR binding.

The multi-zinc finger transcription factor CCCTC-binding factor (CTCF) [13, 14] is the key protein within insulator regions that separate loops of genomic DNA in the size of hundreds of kilobases to a few megabases, referred to as TADs [15], from each other. CTCF is highly

ACCEPTED MANUSCRIPT

conserved between tissues and cell types and organizes the human genome into at least 2,000 active and inactive TADs [16]. In THP-1 cells the binding of CTCF is significantly enhanced by $1,25(OH)_2D_3$ stimulation at more than 1,300 genomic loci [17], more than half of which mark one or both anchors of 587 TADs that contain at least one VDR binding site and one vitamin D target gene. The information that the latter CTCF sites are involved in DNA looping and form TAD borders is based on CTCF chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) data from K562 cells [18] having a genome-wide CTCF binding profile that is 95% identical to THP-1 cells [17].

VDR ChIP-seq results have already been described for THP-1 cells after a ligand treatment of 40 min [8], which is a very short time frame compared to other experimental systems, such as B cells, that had been stimulated for up to 36 h [7]. Therefore, aiming for a better characterization of VDR, we performed in this study triplicate ChIP-seq after 2 and 24 h $1,25(OH)_2D_3$ stimulation of THP-1 cells and obtained a highly accurate VDR cistrome with more than 11,600 genomic loci. From this high total number of sites we found that a subgroup of 339 highly conserved persistent VDR loci is best suited for describing vitamin D gene regulatory scenarios. Almost all primary vitamin D target genes in THP-1 cells are located within $1,25(OH)_2D_3$-modulated TADs and 168 of them are controlled by conserved persistent VDR sites, whereas 120 genes are close to transiently occupied VDR sites. This suggests that the time-dependent binding profile of a few hundred VDR loci is sufficient for regulating most primary vitamin D target genes and through this the physiology of vitamin D.

METHODS

Cell culture

The human acute monocytic leukemia cell line THP-1 [19] is a well responding and physiologically meaningful model system for the investigation of $1,25(OH)_2D_3$-triggered physiological processes, such as those related to innate immunity [8, 20-22]. The cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine,

ACCEPTED MANUSCRIPT

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 chromatin extraction, cells were first grown overnight in phenol red-free medium supplemented with charcoal-stripped fetal calf serum and then treated with vehicle (0.1% ethanol (EtOH)) or 100 nM 1,25(OH)$_2$D$_3$.

ChIP

ChIP assays were performed as described by Zhang *et al.* [23] with some modifications. After treatment of 20 x 10^6 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 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). 25 µl anti-VDR antibody (Cell Signaling Technology, #12550) was coated to 60 µl Dynabeads Protein G (Invitrogen) in an overnight incubation 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 for 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

ACCEPTED MANUSCRIPT

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 1 ml 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. Immune complexes were reverse cross-linked at 50-65 °C for 2-4 h in the presence of proteinase K (Fermentas) in a final concentration of 40 µg/ml. Genomic DNA was isolated with the ChIP DNA Clean&Concentrator Kit (Zymo Research).

ChIP-seq analysis

ChIP DNA templates were sequenced at 50 bp read length using standard manufacturer protocols at the Gene Core of the EMBL (Heidelberg, Germany). All ChIP-seq data were (re)analyzed at harmonized settings: alignment with the human reference genome version hg19 using Bowtie software version 1.1.1 [24] with the following essential command line arguments: bowtie -n 1 -m 1 -k 1 -e 70 --best. The aligned input and VDR reads were converted to sorted BAM format using samtools [25] 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 [26]. Statistically significant ChIP-seq peaks were identified using Model-based Analysis of ChIP-Seq data (MACS) version 2 [27] with the following essential command line arguments: macs2 callpeak –bw 150 –keep-dup 1 –g hs --qvalue=0.01 –m 5 50 --bdg. Otherwise, default parameters were used.

Functional characterization of VDR loci

In total more than 65,000 VDR ChIP-seq peaks were identified in at least one of the three biological repeats of THP-1 cells and in at least one time point and one treatment. Cells were

ACCEPTED MANUSCRIPT

treated for 2 and 24 h with ligand or vehicle and the false discovery rate (FDR) was fixed to 0.1%. This stringent cutoff was applied to each condition (replicate, time and treatment) separately to check for stability in a later stage (**Table S1**). A significant modulation of VDR ChIP-seq peaks by 1,25(OH)$_2$D$_3$ was computed by comparing the intensity of peaks obtained from ligand-treated THP-1 cells to the co-located peak intensity obtained from vehicle-treated cells. The three independent replicate experiments allowed the calculation of statistical significance in a paired Student's *t*-test.

The summit regions (± 100 bp) of the VDR peak loci were searched *de novo* for transcription factor binding motifs via the webtool Hypergeometric Optimization of Motif EnRichment (HOMER) (http://homer.ucsd.edu/homer) [28]. A χ^2 test was computed to check whether the proportion of peaks containing a given attribute (DR3 motifs, overlap with transcription start sites (TSSs) or FAIRE-seq peaks etc.) was different in pairs of groups. For each pair of groups, the proportion of peaks containing the attribute of interest was computed by dividing the number of peaks satisfying the condition over the total number of peaks in the group. In all cases, the degree of freedom was 1. The area under the χ^2 curve was further inspected to compute the p-value. Two peaks overlap if they share at least one bp, or if the distance from the starting/ending point of one of them to the closest point of the second peak does not exceed 50 bp.

Dimensionality reduction and vector quantitation

Both VDR loci and CTCF-defined TADs were characterized by several attributes, such as binding strength, presence of markers, overlap with a TSS and occurrence of binding motifs. These attributes define high-dimensional descriptions that require particular algorithms for a meaningful and comprehensive characterization. Dimensionality reduction algorithms create low-dimensional representations (normally two-dimensional) that allow visual as well as statistical inspections. Self-organizing map (SOM) belongs to the most commonly used dimensionality reduction algorithms, especially when data is non-uniformly distributed and assumptions about it complexity hold [29]. SOM is an unsupervised algorithm that allows the

ACCEPTED MANUSCRIPT

identification of representative profiles of the inspected data. An iterative process captures in a discrete space, defined by a two-dimensional lattice, the topology inherent in the high-dimensional space defined by the usually large number of attributes. SOMs have several advantages over more traditional dimensionality reduction methods, such a principal component analysis, since the former can capture high-order momentum of the data distribution, whereas the latter can only approximate first and second order momentum. SOM also shows vector quantization properties. These allow an unbiased separation of large number of data into a (much) smaller number of classes that define, on average, the specific properties of the data defining that particular class. K-means is an alternative algorithm that separates high-dimensional data into a (small) number of classes. The method involves an iterative process in which data is partitioned into k different centroids, a process which is referred in the machine learning field as a vector quantization. Each centroid defines the center of the corresponding class. The location of the k centroids is computed taking into account the distribution of the high-dimensional data. Vector quantization algorithms construct a partitioning of the space, referred to as Voronoi tessellation. Through the tessellation each data point is "served" by the closest center, also known as the "post-office problem". Data points within a given class show a high similarity between them, whereas data in different classes tend to present a high dissimilarity. The quantization error of the partition is the average (Euclidean) distance between the center of a class and the description of the data points assigned to that class. In general, this error varies with the number k of centers, so the appropriate number of classes k has to be identified by iteration.

Re-analysis of RNA-seq data

RNA sequencing (RNA-seq) reads from THP-1 cells that had been treated for 2.5, 4 and 24 h with ligand or vehicle (GSE69303) [12] were mapped to the reference genome (hg19) using STAR alignment [30]. Then, raw counts and fragments per kilobase million (FKPM) were obtained for profiling gene expression. Finally, differential expression was calculated with DESeq2 [31] under the assumption of negative binomial distribution for the treated and

ACCEPTED MANUSCRIPT

untreated conditions and using a cutoff FDR of 0.5%. In order to take into account differential expression as a function of time, we applied SOM and created in this way an external layer over DESeq2. In this way, from the rather heterogeneous group of 3,650 vitamin D target genes a small number of profiles were obtained. The webtool Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) (http://cbl-gorilla.cs.technion.ac.il) [32] was used for gene ontology (GO) enrichment analysis of vitamin D target genes. GOrilla identifies, via a commonly applied hypergeometric model, the enriched terms for the target list of genes against a background list. The algorithm offers adequate p-values as a consequence of the robust assumptions about the underlying probability distribution. In common gene set enrichment analysis the p-value is estimated via permutations, i.e. the p-value is a function of the number of permutations. In contrast, in GOrilla a lower number of permutations are needed, in order to achieve stability in the estimated p-value.

RESULTS

Features of the VDR cistrome in human monocytes

With the aim to obtain a highly accurate, time-dependent cistrome of VDR in human monocytes, we performed in three biological repeats VDR ChIP-seq with THP-1 cells that had been stimulated for 2 and 24 h with vehicle (EtOH) or $1,25(OH)_2D_3$ (**Table S2**). In vehicle-treated cells we found 854 and 1,714 VDR binding sites (FDR < 0.1%) at time points 2 and 24 h, respectively, while in the presence of $1,25(OH)_2D_3$ even 1,685 and 11,287 VDR loci were detected (**Fig. 1A**). The number of overlapping VDR loci in ligand treated and untreated cells was 636 at 2 h stimulation and even 1,495 at 24 h (**Fig. S1A**). For 15 (2.4%) and 421 (28.2%) of these sites, respectively, the VDR binding strength was significantly ($p < 0.05$) modulated by $1,25(OH)_2D_3$ (**Fig. 1A**). Furthermore, 1,546 (91.8%) of the 1,685 VDR loci obtained after 2 h stimulation with $1,25(OH)_2D_3$ were identical to those observed after 24 h ligand stimulation (**Fig. S1A**), 397 (25.7%) of which were significantly ($p < 0.05$) modulated by ligand (**Fig. 1A**). This implies that only 139 VDR sites occurred exclusively at

ACCEPTED MANUSCRIPT

the early time point (2 h), while 9,741 loci were detected only at the late time point (24 h) (**Fig. S1A**).

The complete VDR cistrome of 11,657 sites that were obtained early (2 h) and late (24 h) in the absence and presence of 1,25(OH)$_2$D$_3$ (**Table S2**) can be sub-divided into i) 510 “persistent” sites that occurred at all four conditions, ii) 2,109 “transient” loci that were obtained at 1-3 conditions and not exclusively at the ligand-stimulated late time point and iii) 9,038 VDR sites that were detected only after 24 h in 1,25(OH)$_2$D$_3$-stimulated cells (**Fig. 1B**). Below the summits of 333 (65.3%) persistent VDR loci HOMER screening (score 7) detected a DR3-type motif. Moreover, these sites also scored highest for their average VDR binding strength (19.3). From the transient VDR sites 1,128 (53.5%) carried a DR3-type motif and had an average receptor binding strength of 10.4. For the “24 h 1,25(OH)$_2$D$_3$ only” VDR loci the DR3 rate was only 45.3% and the average binding strength just 8.7. The mean VDR binding strength of the respective sites not carrying a DR3-type motif was with 13.3, 8.1 and 7.0 always lower than in the presence of the canonical receptor binding site. The overall DR3 rate of all 11,657 VDR sites is 47.6%. In all three VDR peak categories DR3 motifs are most prominent, but we also detected enrichment ($p < 0.00001$) for binding motifs of the transcription factors ETV1, GABPA, ERG, STAT5, ELF5, PU.1, BMYB and ZNF711 below the peak summits (**Table S3**). Interestingly, STAT5 was not significantly enriched below “24 h 1,25(OH)$_2$D$_3$ only” VDR loci and persistent VDR peaks are specific for ZNF711 sites.

For further characterization of the 510 persistent VDR loci we applied a SOM approach with the attributes i) VDR binding strength at each of the four conditions, ii) overlap with a TSS, overlap with accessible chromatin as measured by FAIRE-seq [12] and iii) rate of DR3-type motifs below the peak summits as discriminators (**Fig. 1C**). This analysis demonstrated that the persistent sites are a heterogenous group of VDR loci that can be subdivided into nine more homogenous subgroups. The 313 VDR sites belonging to subgroups 2, 3, 5 or 6 were characterized by a high level of chromatin accessibility paired with the presence of DR3-type motifs. Furthermore, subgroups 6, 8 and 9 (121 members in total) showed very high VDR binding strength after 24 h ligand treatment, i.e. they represent the most prominent VDR sites.

ACCEPTED MANUSCRIPT

In contrast, the 54 members of subgroup 7 displayed only average levels for all assessed attributes and the in total 87 members of subgroups 1 and 4 were low in all attributes and probably represent VDR loci of lower functionality. The VDR binding sites 30 kb downstream of the *SEPT3* TSS, 18 kb downstream of the *NINJ1* TSS, 185 kb downstream of the *NCOR2* TSS and 3.5 kb upstream of the *NFE2* TSS are representative examples for members of subgroups 9, 8, 7 and 6 (in descending order of binding strength), respectively (**Fig. S1B**). At all four sites VDR binding strength significantly increased after ligand stimulation, all loci overlapped with accessible chromatin and 3 of the 4 sites carried a DR3-type motif.

The genome-wide peak strength distribution of the 510 persistent VDR sites (**Fig. 1D**), color-coded for the nine SOM subgroups, showed a rather even allocation, i.e. there is no obvious clustering of VDR loci within the human genome. The average distance between persistent VDR sites was 2.12 Mb (ranging from 622 bp to 19.9 Mb), while the average distance of these sites to all VDR loci was only 0.35 Mb (ranging from 495 bp to 6.06 Mb, **Fig. S1C**). In concordance with the examples shown in **Fig. S1B** the mean VDR binding strength of the 510 persistent loci increased significantly ($p < 10^{-8}$) after ligand treatment at both time points as well as with increased stimulation time from 2 to 24 h (**Fig. S1D**).

The only other hematopoietic cell systems, in which VDR ChIP-seq results are publically available, are the lymphoblastoid cell lines GM10855 and GM10861 that were either unstimulated or treated for 36 h with $1,25(OH)_2D_3$ [7]. We performed harmonized re-analysis of the data using identical peak calling settings as for THP-1 cells and used only those peaks that occurred in both of the duplicates available for the B cell models. This provided large sets of some 14,000 and 15,500 VDR loci for unstimulated GM10855 and GM10861 cells and even 22,000 and 39,300 sites for stimulated cells, which, however, overlapped only for some 6,600 and 16,200 positions, respectively (**Figs. S2A and B**). Only 11.0% (94 of 854) and 12.2% (209 of 1,714) of the VDR sites observed in THP-1 cells that were treated with vehicle for 2 and 24 h, respectively, overlapped with the common VDR sites of unstimulated lymphoblastoids (**Fig. S2A**). In parallel, 34.2% (576 of 1,685) and 23.3% (2,632 of 11,287) of

ACCEPTED MANUSCRIPT

VDR loci in THP-1 cells that were stimulated with 1,25(OH)$_2$D$_3$ for 2 and 24 h, respectively, overlapped with the common VDR sites of the B cell clones (**Fig. S2B**).

From the subgroup of 510 persistent VDR sites in THP-1 cells 12.7% (65) and 45.3% (231) overlapped with VDR loci common in unstimulated and stimulated B cells, respectively (**Fig. S3A**). The corresponding percentages for the 2,109 transient and the 9,038 "24 h 1,25D only" VDR sites were significantly ($p < 0.0000003$; $\chi^2 = 99.2$) lower than that of the persistent loci, as they were only 10.2% (215) and 29.5% (623) (**Fig. S3B**) and 7.6% (700) and 20.2% (1,827) (**Fig. S3C**), respectively. Interestingly, the nine SOM subgroups of the persistent VDR sites differed clearly in their overlap rate with B cell VDR ChIP-seq data ranging from 2.5 and 19.1% for subgroups 1 and 4 up to 70.5 and 81.8% for subgroups 8 and 9 (**Fig. 1E**). For comparison, the overlap with VDR ChIP-seq data that we obtained previously from THP-1 cells that were stimulated for 40 min with 1,25(OH)$_2$D$_3$ [8] showed a very comparable SOM subgroup ranking ranging from 0 and 6.3% for subgroups 4 and 1 and 95.4 and 100% for subgroups 6 and 8 (**Fig. 1F**). The VDR binding sites 0.5 kb upstream of the *CAMP* TSS, 12 kb downstream of the *DENND6B* TSS, 1.8 kb downstream of the *CYTH4* TSS and 37 kb downstream of the *SEMA4D* TSS are representative examples for members of subgroups 9, 8, 7 and 6, respectively (**Fig. S4**).

In summary, the VDR cistrome in THP-1 cells comprises more than 11,600 sites, 78% of which were detected only after 24 h ligand stimulation. In contrast, a group of 510 persistent VDR loci occurred at all four conditions and was characterized by very high DR3 motif rate and VDR binding strength. In direct comparison with B cell VDR ChIP-seq data the persistent VDR loci showed a significantly higher overlap rate than other VDR groups in THP-1 cells. Moreover, the SOM subgroups 6-9 within the persistent VDR sites are confirmed to a high rate in B cells, while subgroups 1, 2 and 4 displayed only low conservation between monocytes and B cells.

ACCEPTED MANUSCRIPT

$1,25(OH)_2D_3$-triggered transcriptome of human monocytes

RNA-seq data from THP-1 cells that had been treated for 2.5, 4 and 24 h with $1,25(OH)_2D_3$ were initially analyzed by Bowtie alignment and Cuffdiff statistics for differential expression and indicated in total 1,284 vitamin D target genes [12]. When we re-analyzed the raw data by using more powerful, recent algorithms, such as STAR for the alignment and DESeq2 for the detection of differential gene expression, we found in total 3,650 genes that were significantly ($p < 0.05$) modulated in their expression by $1,25(OH)_2D_3$ (**Table S4**). From these, 300 genes responded already after 2.5 h stimulation and 440 after 4 h making a total of 517 primary $1,25(OH)_2D_3$ target genes (14.2% of all, **Fig. 2A**). A genome-wide display of the vitamin D target genes at the three time points demonstrated their rather even distribution, i.e. there was no obvious clustering at any chromosomal region (**Fig. S5A**).

Next we used a SOM approach to sort the rather heterogenous group of 3,650 vitamin D target genes into nine more homogenous subgroups (**Fig. 2B**). The basal state of the 3,650 genes and their change in expression (at 2.5, 4, and 24 h) are four attributes representing four dimensions, which were transformed by dimensionality reduction and vector quantization into the presented expression profiles (**Fig. 2B**). SOM mapped the expression of the 3,650 genes into a lattice of nine (3x3) units, each of which was associated to a subgroup of similar genes. The more distant the units, the more dissimilar the profiles are (compare subgroups 1 and 9, the most distant ones). The 11 and 49 members of subgroups 1 and 4, representing strongly and moderately down-regulated genes, as well as the 17, 109 and 401 members of the strongly, highly and moderately up-regulated subgroups 9, 8 and 6 clearly differed from the 162, 1,028 and 483 weakly down-regulated members of subgroups 2, 3 and 7 as well as from the 1,390 up-regulated genes in subgroup 5, respectively, that were in average less than 2-fold modulated by $1,25(OH)_2D_3$. Since in addition the 3,063 members of subgroups 2, 3, 5 and 7 showed less change as a function of time, we filtered them out. From the remaining 587 genes 311 (53.0%) were primary vitamin D targets, i.e. SOM filtering resulted in their 3.7-fold enrichment (**Fig. 2C**). The genome-wide display of the 217 genes responding to $1,25(OH)_2D_3$

ACCEPTED MANUSCRIPT

already after 2.5 h, the 289 genes reacting after 4 h and all 587 genes being responsive after 24 h again showed no obvious signs of spatial clustering (**Fig. S5B**).

Taken together, the application of the most recent RNA-seq data analysis tools indicated 3,650 significantly 1,25(OH)$_2$D$_3$ responsive genes. Using SOM subgrouping and a threshold level of more than 2-fold average regulation reduced this list to 587 genes, 53% of which are primary vitamin D targets.

VDR type defined TADs relate to distinct physiological functions

Recently, we observed significantly ($p < 0.05$) 1,25(OH)$_2$D$_3$-sensitive CTCF binding sites and found by comparison with ChIA-PET data from K562 cells that the majority of them are anchors of 587 TADs [17]. This creates for the very most VDR sites and their target genes a defined chromosomal environment, i.e. we have a good estimation which target genes are controlled by which VDR loci in their vicinity. The genomic environment of vitamin D target genes is heterogeneous and we assumed that this heterogeneity provides the tools to unveil the pleiotropic properties of vitamin D. Therefore, we first characterized the vitamin D-sensitive TADs exclusively by their VDR-related attributes, such as the number of persistent and transient VDR sites, total VDR binding strength and average fold change (FC) of VDR binding. K-means suggested the categorization of the TADs into five most distinct classes (**Table S5**). We also tested the categorization into 3, 4, 6 and 7 TAD classes but obtained with 5 the lowest quantization error. The TADs in each of the five classes contain a set of 34 to 156 primary and secondary vitamin D target genes. GO analysis using GOrilla indicated for each of the five TAD classes a non-overlapping set of enriched biological processes, in which the respective vitamin D target genes are involved (**Table S5**). We used a Voronoi tessellation to display for each of the five TAD classes the three most prominently enriched biological processes (**Fig. 3**). Each of the k centers found by the K-means algorithm corresponds to a specific class of TADs, and each of the classes specifies a rather different VDR activity, quantified as the number of transient or persistent peaks, binding strength and other VDR-related attributes. The separation of genes based on their VDR-defined environment shows a

ACCEPTED MANUSCRIPT

clear pleiotropy: the 126 genes contained in TADs of class 1 are specifically enriched in processes related to immune function, the 34 genes of TAD class 2 are connected with differentiation and adhesion, the 83 genes of TAD class 3 are found in membrane-based signal transduction pathways, while the function of the 140 genes of TAD class 4 is rather diverse ranging from antigen-receptor signaling, RAS kinase signaling and lipid storage and finally the 156 genes of TAD class 5 are related to cellular localization and transport. This suggests that the rather broad functional profile of vitamin D can be subdivided into modules being characterized by different types of TADs that can be distinguished by their VDR binding profiles. A further important observation of the K-means segregation of TAD classes was that the most prominent attributes for their distinction are the number of persistent and transient VDR sites.

In summary, exclusively VDR-related attributes were sufficient to sort vitamin D-sensitive TADs into five classes that carry vitamin D target genes involved with clearly distinct biological processes. Main discriminators for this TAD class sorting approach were the numbers of persistent and transient VDR sites.

Gene regulatory scenarios of primary vitamin D target genes

From the 311 primary $1,25(OH)_2D_3$ target genes 302 (97.1%) are located within a vitamin D sensitive TAD, while this was a bit less likely (85.9%) for the 276 secondary target genes (**Table S6**). A Circos plot displayed the location of persistent VDR sites, primary and secondary vitamin D target genes and vitamin D-sensitive TADs indicating their rather even genome-wide distribution (**Fig. 4**). This represents an alternative approach to the Voronoi tessellation (**Fig. 3**) and allowed us to distinguish four types of TADs that contain either i) only primary vitamin D target genes, ii) only secondary vitamin D target genes, iii) primary and secondary vitamin D target genes and iv) no vitamin D target genes. Interestingly, 95 of the 172 (55.3%) TADs with only primary vitamin D target genes contain at least one of the 339 prominent persistent VDR sites, while this is the case only for 29 of the 120 (24.2%) TADs carrying exclusively secondary vitamin D target genes. In total, 168 of the 311 (54.0%)

ACCEPTED MANUSCRIPT

primary vitamin D target genes are co-located with a prominent persistent VDR locus (**Fig. 4**). In contrast, secondary vitamin D target genes are significantly less likely associated with persistent VDR loci ($p = 0.00022$; $\chi^2 = 13.58$).

Examples of TADs containing only primary vitamin D target genes are that of *ARHGEF16* controlled by a persistent VDR site 0.6 kb downstream of its TSS (**Fig. S6A**) and *GOLPH3* with even two persistent VDR loci 0.3 kb downstream and 57 kb upstream of its TSS (**Fig. S6B**). These two example TADs each contain one additional transient VDR binding site. From the 134 remaining primary vitamin D target genes that are located within a TAD lacking a persistent VDR locus, 120 carry at least one transient VDR site within their regulatory locus (**Table S6**). Examples of these types of primary vitamin D target genes are *KCNK5* with a transient VDR site 36 kb downstream of its TSS (**Fig. 5A**) and a TAD with the genes *CCR1* and *CCRL2* possibly controlled by the same transient VDR site 13 and 185 kb upstream of their TSS, respectively (**Fig. 5B**). This leaves 62 primary target genes (19.9% of all) that cannot be explained by these gene regulatory scenarios.

GO enrichment analysis of all 311 primary 1,25(OH)$_2$D$_3$ target genes using GOrilla listed "regulation of immune system process" and "immune system process" among the most abundant terms (**Table S6**). This highlights that regulation of the immune system is a key physiological consequence of primary vitamin D target genes. In contrast, when the GO enrichment analysis was focused on the 168 primary vitamin D target genes that are controlled by a persistent VDR site, five rather general terms (threshold $p < 0.0001$), such as "cell surface receptor signaling pathway" and "positive regulation of response to stimulus" were identified (**Fig. S7**). For comparison, under the same constraints the 120 primary targets under the control of a transient VDR site were associated with 12 more specific terms, such as "regulation of immune system process" and "regulation of osteoclast differentiation".

Taken together, from the 311 primary vitamin D target genes in human monocytes 168 are controlled by persistent VDR sites and 120 by transient VDR loci, i.e. we understand the regulation of more than 92% of the genes. These two classes of primary vitamin D target

ACCEPTED MANUSCRIPT

genes can be functionally distinguished by either regulating rather general cellular processes or immune system-specific functions.

DISCUSSION

The key to the understanding of the pleiotropic action of vitamin D is the spatio-temporal binding profile of its nuclear receptor VDR. Therefore, we investigated in this study the VDR cistrome at time points 2 and 24 h after the stimulation of human monocytes with $1{,}25(OH)_2D_3$. These two time points are of high relevance to dissect early and late response to $1{,}25(OH)_2D_3$. The use of three biological repeats did not only increase the accuracy of the genome-wide binding pattern of VDR, but in addition allowed hypothesis testing. A first impressive result of this study was the massive increase of VDR binding loci to more than 11,000 sites after 24 h ligand stimulation. This 6.6-fold boost in relation to vehicle-treated cells is significantly higher than the 2.0-fold increase after 2 h or the 1.3-fold increment after 40 min described in our earlier reports [8, 9]. For comparison, our previous FAIRE-seq analysis in THP-1 cells had demonstrated that a 2 h stimulation with $1{,}25(OH)_2D_3$ results at more than 3,300 genomic loci in significant changes in chromatin accessibility, after 24 h even at more than 4,500 sites, while after 48 h only some 2,400 regions respond to vitamin D [12]. This provides additional evidence that a 24 h stimulation with $1{,}25(OH)_2D_3$ causes maximal (epi)genome-wide effects. Interestingly, in the same cellular system a 24 h vitamin D stimulation induced at more than 1,300 genomic loci significant changes in CTCF binding [17] or at more than 5,600 sites the binding of the pioneer transcription factor PU.1 [33]. Taken together, within 24 h after the onset of $1{,}25(OH)_2D_3$ stimulation the epigenome-wide outcomes are i) VDR binding at more than 10,000 additional sites, ii) chromatin opening at some 4,500 loci, iii) increased PU.1 pioneer factor binding at more than 5,000 regions and iv) changes in CTCF-based TAD anchors affecting more than 500 chromatin loops. Moreover, within 24 h these vitamin D-triggered effects on the epigenome affect first 311 primary target genes and later 276 secondary target genes.

ACCEPTED MANUSCRIPT

Our observation that vitamin D-sensitive TADs [17] can be classified exclusively by VDR cistrome-related data into five classes, the vitamin D target genes of which are involved in non-overlapping biological processes, demonstrates that the epigenome-wide binding pattern of VDR is sufficient for describing the molecular basis of the physiological function of vitamin D. The conserved persistent VDR loci presented in this study suggest for more than half of all primary vitamin D target genes (168 of 311) a rather simple gene regulatory scenario. A target gene is controlled by one or more conserved persistent VDR sites that are located within the same TAD, i.e. at this locus VDR is activated via ligand binding and then affects the structure of its chromatin neighborhood, so that its accessibility is increased, which allows the binding of PU.1 proteins as well as of further VDR molecules including an increased association with the original locus. For about a third of the primary vitamin D target genes (120 of 311) a transiently occupied VDR locus seems to be the main regulator of the respective TAD. In both scenarios, a key condition is the presence of VDR already in the absence of ligand or 2 h after ligand stimulation, in order to take the role as prime target for vitamin D within the given TAD.

We assume that after 24 h stimulation the very most of the effects of vitamin D are of secondary nature. However, in THP-1 cells there are 510 persistent VDR loci, 339 of which are highly conserved. These 339 VDR sites do not only occur at all four conditions and are confirmed by B cell VDR ChIP-seq datasets, but they also have an over-proportionally high binding strength and rate of DR3-type motifs. The equal distribution of these sites over the human genome suggests that they may be strategically positioned, in order to provide the whole genome with sensitivity to vitamin D. The similarly equal genomic distribution of primary vitamin D target genes supports this concept. Therefore, we interpret the conserved persistent VDR sites as “nodes” in vitamin D signaling, at which not only primary contacts of VDR ligands with the genome are established, but also functional consequences of vitamin D induction are coordinated throughout the whole stimulation period. Accordingly, this class of primary vitamin D target genes under the control of conserved persistent VDR sites seems to mediate rather ubiquitous cell activation processes. In contrast, the second class of primary

ACCEPTED MANUSCRIPT

targets that are controlled by transient VDR sites are more specifically related to the immune-regulatory functions of vitamin D. Transient VDR sites are characterized by the property of not always being bound by the receptor, i.e. to be regulated by “on/off” association of VDR with its chromatin loci. This regulatory mechanism is more suited for a tight control of immune functions than the constant presence of VDR to persistent VDR sites.

The clustering of vitamin D target genes obtained by the similitudes of their corresponding TAD landscapes leads to an interesting observation. Genes in TADs described by similar VDR-related attributes tend to participate in enriched pathways that are exclusive for that group of TADs. This offers specific evidence on how the VDR activity tends to orchestrate the pleiotropic effects of $1,25(OH)_2D_3$. That is, genes in regions with similar VDR activity tend to participate in specific pathways, which in turn are different to the pathways enriched for genes in domains with different VDR activity.

In conclusion, within a time span of 24 h the stimulation of a cell with $1,25(OH)_2D_3$ provokes at more than 10,000 genomic loci VDR binding events. Key nodes in vitamin D signaling seem to be conserved persistent VDR sites that control ubiquitous cell activating functions. Moreover, a few hundred transient VDR sites mediate more tissue-specific primary functions of vitamin D, such as immune system regulation in the THP-1 monocyte model system. The very most of the remaining VDR sites seem to be involved in secondary effects of vitamin D.

ACKNOWLEDGEMENTS

This work was supported by the Academy of Finland (#267067). Kind thanks to the Gene Core Facility at the EMBL in Heidelberg, Germany, for massively parallel sequencing services.

ACCEPTED MANUSCRIPT

REFERENCES

[1] A. Verstuyf, G. Carmeliet, R. Bouillon, C. Mathieu, Vitamin D: a pleiotropic hormone, Kidney international, 78 (2010) 140-145.

[2] D. Feldman, A.V. Krishnan, S. Swami, E. Giovannucci, B.J. Feldman, The role of vitamin D in reducing cancer risk and progression, Nat Rev Cancer, 14 (2014) 342-357.

[3] C. Carlberg, The physiology of vitamin D-far more than calcium and bone, Front Physiol, 5 (2014) 335.

[4] M.R. Haussler, C.A. Haussler, L. Bartik, G.K. Whitfield, J.C. Hsieh, S. Slater, P.W. Jurutka, Vitamin D receptor: molecular signaling and actions of nutritional ligands in disease prevention, Nutrition reviews, 66 (2008) S98-112.

[5] K. Umesono, K.K. Murakami, C.C. Thompson, R.M. Evans, Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D_3 receptors, Cell, 65 (1991) 1255-1266.

[6] C. Carlberg, I. Bendik, A. Wyss, E. Meier, L.J. Sturzenbecker, J.F. Grippo, W. Hunziker, Two nuclear signalling pathways for vitamin D, Nature, 361 (1993) 657-660.

[7] S.V. Ramagopalan, A. Heger, A.J. Berlanga, N.J. Maugeri, M.R. Lincoln, A. Burrell, L. Handunnetthi, A.E. Handel, G. Disanto, S.M. Orton, C.T. Watson, J.M. Morahan, G. Giovannoni, C.P. Ponting, G.C. Ebers, J.C. Knight, A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution, Genome research, 20 (2010) 1352-1360.

[8] S. Heikkinen, S. Väisänen, P. Pehkonen, S. Seuter, V. Benes, C. Carlberg, Nuclear hormone 1α,25-dihydroxyvitamin D_3 elicits a genome-wide shift in the locations of VDR chromatin occupancy, Nucleic Acids Res, 39 (2011) 9181-9193.

[9] P. Tuoresmäki, S. Väisänen, A. Neme, S. Heikkinen, C. Carlberg, Patterns of genome-wide VDR locations, PLoS ONE, 9 (2014) e96105.

[10] M.B. Meyer, P.D. Goetsch, J.W. Pike, VDR/RXR and TCF4/beta-catenin cistromes in colonic cells of colorectal tumor origin: impact on c-FOS and c-MYC gene expression, Mol Endocrinol, 26 (2012) 37-51.

[11] N. Ding, R.T. Yu, N. Subramaniam, M.H. Sherman, C. Wilson, R. Rao, M. Leblanc, S. Coulter, M. He, C. Scott, S.L. Lau, A.R. Atkins, G.D. Barish, J.E. Gunton, C. Liddle, M. Downes, R.M. Evans, A vitamin D receptor/SMAD genomic circuit gates hepatic fibrotic response, Cell, 153 (2013) 601-613.

[12] S. Seuter, A. Neme, C. Carlberg, Epigenome-wide effects of vitamin D and their impact on the transcriptome of human monocytes involve CTCF, Nucleic Acids Res, 44 (2016) 4090-4104.

[13] J.E. Phillips, V.G. Corces, CTCF: master weaver of the genome, Cell, 137 (2009) 1194-1211.

ACCEPTED MANUSCRIPT

[14] R. Ghirlando, G. Felsenfeld, CTCF: making the right connections, Genes Dev, 30 (2016) 881-891.

[15] T. Ali, R. Renkawitz, M. Bartkuhn, Insulators and domains of gene expression, Curr Opin Genet Dev, 37 (2016) 17-26.

[16] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions, Nature, 485 (2012) 376-380.

[17] A. Neme, S. Seuter, C. Carlberg, Vitamin D-dependent chromatin association of CTCF in human monocytes, Biochim Biophys Acta, 1859 (2016) 1380-1388.

[18] ENCODE-Project-Consortium, B.E. Bernstein, E. Birney, I. Dunham, E.D. Green, C. Gunter, M. Snyder, An integrated encyclopedia of DNA elements in the human genome, Nature, 489 (2012) 57-74.

[19] S. Tsuchiya, M. Yamabe, Y. Yamaguchi, Y. Kobayashi, T. Konno, K. Tada, Establishment and characterization of a human acute monocytic leukemia cell line (THP-1), Int J Cancer, 26 (1980) 171-176.

[20] J.M. Matilainen, T. Husso, S. Toropainen, S. Seuter, M.P. Turunen, P. Gynther, S. Ylä-Herttuala, C. Carlberg, S. Väisänen, Primary effect of 1α,25(OH)$_2$D$_3$ on IL-10 expression in monocytes is short-term down-regulation, Biochim Biophys Acta, 1803 (2010) 1276-1286.

[21] P. Gynther, S. Toropainen, J.M. Matilainen, S. Seuter, C. Carlberg, S. Väisänen, Mechanism of 1α,25-dihydroxyvitamin D$_3$-dependent repression of interleukin-12B, Biochim Biophys Acta, 1813 (2011) 810-818.

[22] S. Seuter, S. Heikkinen, C. Carlberg, Chromatin acetylation at transcription start sites and vitamin D receptor binding regions relates to effects of 1α,25-dihydroxyvitamin D$_3$ and histone deacetylase inhibitors on gene expression, Nucleic Acids Res, 41 (2013) 110-124.

[23] J. Zhang, H.M. Poh, S.Q. Peh, Y.Y. Sia, G. Li, F.H. Mulawadi, Y. Goh, M.J. Fullwood, W.K. Sung, X. Ruan, Y. Ruan, ChIA-PET analysis of transcriptional chromatin interactions, Methods, 58 (2012) 289-299.

[24] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol, 10 (2009) R25.

[25] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools, Bioinformatics, 25 (2009) 2078-2079.

[26] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat Biotechnol, 29 (2011) 24-26.

[27] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based analysis of ChIP-Seq (MACS), Genome Biol, 9 (2008) R137.

ACCEPTED MANUSCRIPT

[28] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, C.K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol Cell, 38 (2010) 576-589.

[29] T. Kohonen, Self-organizing maps, Series Inform. Sci., 30 (1995).

[30] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, 29 (2013) 15-21.

[31] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol, 15 (2014) 550.

[32] E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists, BMC bioinformatics, 10 (2009) 48.

[33] S. Seuter, A. Neme, C. Carlberg, Epigenomic PU.1-VDR crosstalk modulates vitamin D signaling, Biochim Biophys Acta, 1860 (2017) 405-415.

AUTHOR CONTRIBUTIONS

A.N., S.S. and C.C. conceived and designed this study. A.N. did all data analysis and created the figures, S.S. conducted the experiments and C.C. wrote the manuscript. All authors reviewed and revised the manuscript.

ADDITIONAL INFORMATION

Accession codes: VDR ChIP-seq raw data are available at Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) at GSE89431.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

FIGURE LEGENDS

Figure 1: The VDR cistrome in monocytes. In three independent experiments THP-1 cells were stimulated with either vehicle (0.1% EtOH) or ligand (100 nM 1,25(OH)$_2$D$_3$, 1,25D) for 2 and 24 h and VDR ChIP-seq was performed. For each condition, only peaks present in all three biological repeats were considered. VDR peak counts for the individual conditions as well as for their overlap are indicated (**A**). The in total 11,657 VDR sites are segregated into 510 persistent sites, 2,109 transient loci and 9,038 sites found exclusively in cells treated for 24 h with 1,25(OH)$_2$D$_3$ (**B**). A SOM was performed with all 510 persistent VDR sites based on VDR binding strength (BS), TSS overlap, chromatin accessibility and DR3-type motif presence. Nine more homogenous sub-groups are differentiated (**C**). The most discriminatory attributes are indicated in orange and red, less important attributes are in blue and light blue. Bars indicate average numbers with red and green lines as maximal and minimal ranges, respectively. The number of DR3 motif-containing sites in relation to DR3-free sites and their respective average VDR binding strength are determined. Genome-wide view of the sum of the VDR binding strength (at the 4 conditions) for the 510 persistent VDR sites, color-coded for the nine SOM subgroups (**D**). Weak VDR sites (mainly members of SOM groups 1, 2 and 4) are located below the horizontal grey line. For B cell lines (**E** [7]) as well as for a previously published THP-1 VDR ChIP-seq dataset (**F** [8], 40 min 1,25(OH)$_2$D$_3$ treatment) the overlap with persistent VDR sites from this study are differentiated for the nine SOM subgroups.

Figure 2: Differential analysis of 1,25(OH)$_2$D$_3$-dependent transcriptome in monocytes. The overlap of triplicate RNA-seq datasets from THP-1 cells that had been treated for 2.5, 4 and 24 h with 1,25(OH)$_2$D$_3$ is displayed by Venn diagrams [12]. Re-analysis using STAR alignment and DESeq2 statistics provided 3,650 significantly ($p < 0.05$) vitamin D-responsive genes (**A**) that were filtered by SOM analysis (**B**) to 587 most prominently responding genes (mean FC ≥ 2) (**C**). In **B** for each group the average FC of gene expression (red) was

ACCEPTED MANUSCRIPT

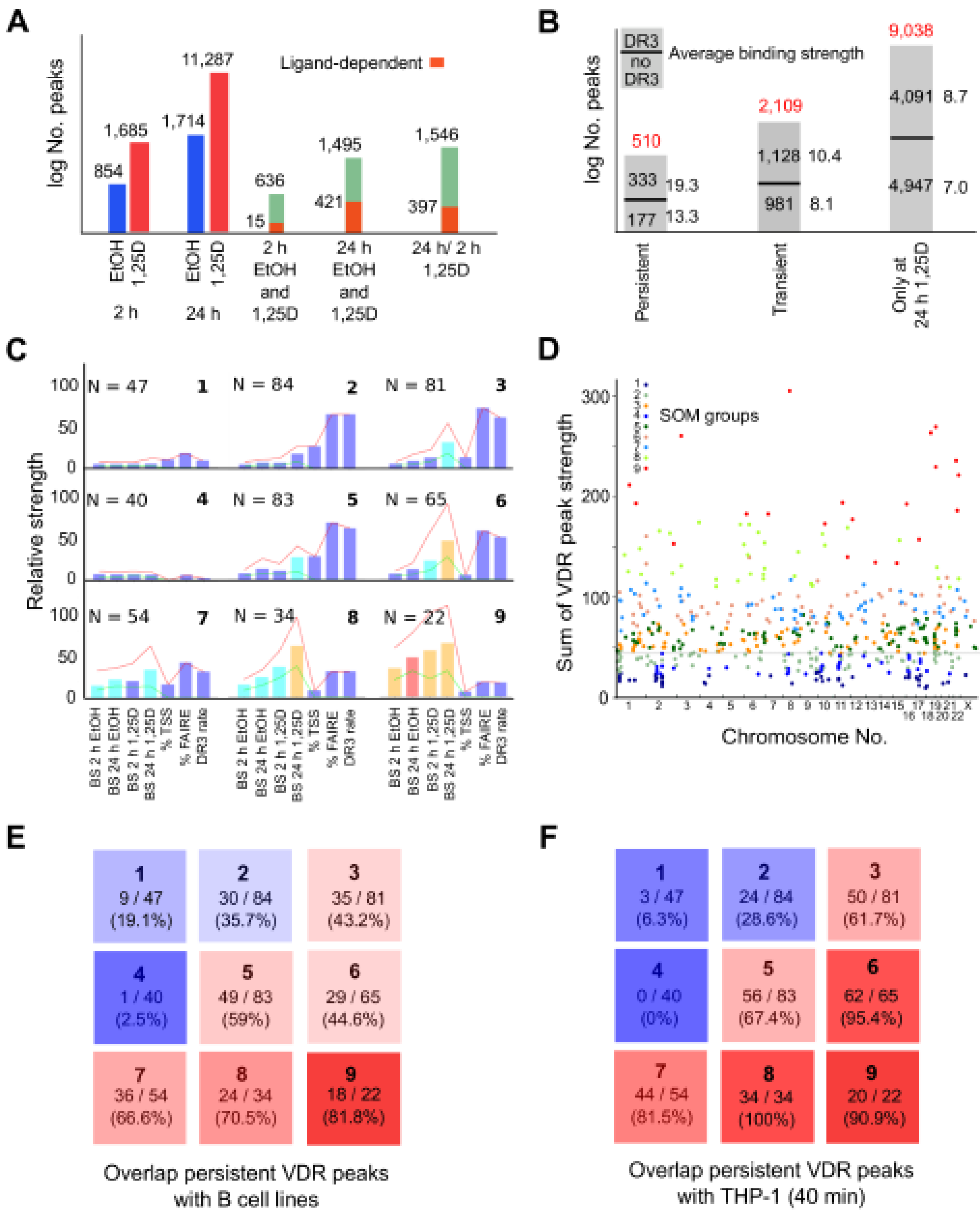
displayed over time. Blue shades indicate the variation in gene expression and the green rectangles FC ≤ 2.

Figure 3: Relating TAD classes to biological processes. A Voronoi tessellation was used to display for each of the five TAD classes the three most prominently enriched biological processes, their p-values and the number of genes representing the process (for more details see **Table S5**). The most relevant attributes, number of persistent and transient VDR sites, were chosen for the x and y axis, respectively. Black dots indicate the center of the respective TAD class area.

Figure 4: Genome-wide gene regulatory scenarios of vitamin D in monocytes. A Circos plot displays genome-wide the location of the 339 conserved persistent VDR binding sites, the 311 primary and 276 secondary vitamin D target genes and four different types of vitamin D-triggered TADs. The height and orientation of the signals indicate binding, strength, induction and up- or down-regulation. This allows differentiating several gene regulatory scenarios, of which a TAD carrying only primary vitamin D target genes (red) and at least one persistent VDR site is the most common. In contrast, secondary vitamin D target genes (blue) are significantly less likely associated with persistent VDR loci.

Figure 5: Example TADs containing primary vitamin D target genes under the control of transient VDR sites. THP-1 cells were treated for 2 and 24 h with vehicle (EtOH) or 1,25(OH)$_2$D$_3$ (1,25D) and ChIP-seq, FAIRE-seq and RNA-seq were performed in three biological repeats. The IGV browser [26] was used to visualize the TAD around the primary vitamin D target genes *KCNK5* (**A**) as well as *CCR1* and *CCRL2* (**B**). The peak tracks display data from ChIP-seq for VDR (2 h, orange; 24 h, red), FAIRE-seq (light blue), and CTCF (blue). Gene structures are shown in blue. The change in expression of the genes within the TADs by the ligand treatment was measured by RNA-seq.

Fig. 1


A
log No. peaks
Ligand-dependent
11,287
1,685
1,714
854
636
15
1,495
421
1,546
397
EtOH
1,25D
2 h
24 h
2 h EtOH and 1,25D
24 h EtOH and 1,25D
24 h/ 2 h 1,25D
B
log No. peaks
DR3
no DR3
Average binding strength
9,038
4,091 8.7
4,947 7.0
2,109
1,128 10.4
981 8.1
510
333 19.3
177 13.3
Persistent
Transient
Only at 24 h 1,25D
C
Relative strength
N = 47
N = 84
N = 81
N = 40
N = 83
N = 65
N = 54
N = 34
N = 22
BS 2 h EtOH
BS 24 h EtOH
BS 2 h 1,25D
BS 24 h 1,25D
% TSS
% FAIRE
DR3 rate
D
Sum of VDR peak strength
SOM groups
Chromosome No.
E
1 9 / 47 (19.1%)
2 30 / 84 (35.7%)
3 35 / 81 (43.2%)
4 1 / 40 (2.5%)
5 49 / 83 (59%)
6 29 / 65 (44.6%)
7 36 / 54 (66.6%)
8 24 / 34 (70.5%)
9 18 / 22 (81.8%)
Overlap persistent VDR peaks with B cell lines
F
1 3 / 47 (6.3%)
2 24 / 84 (28.6%)
3 50 / 81 (61.7%)
4 0 / 40 (0%)
5 56 / 83 (67.4%)
6 62 / 65 (95.4%)
7 44 / 54 (81.5%)
8 34 / 34 (100%)
9 20 / 22 (90.9%)
Overlap persistent VDR peaks with THP-1 (40 min)

ACCEPTED MANUSCRIPT

Fig. 2

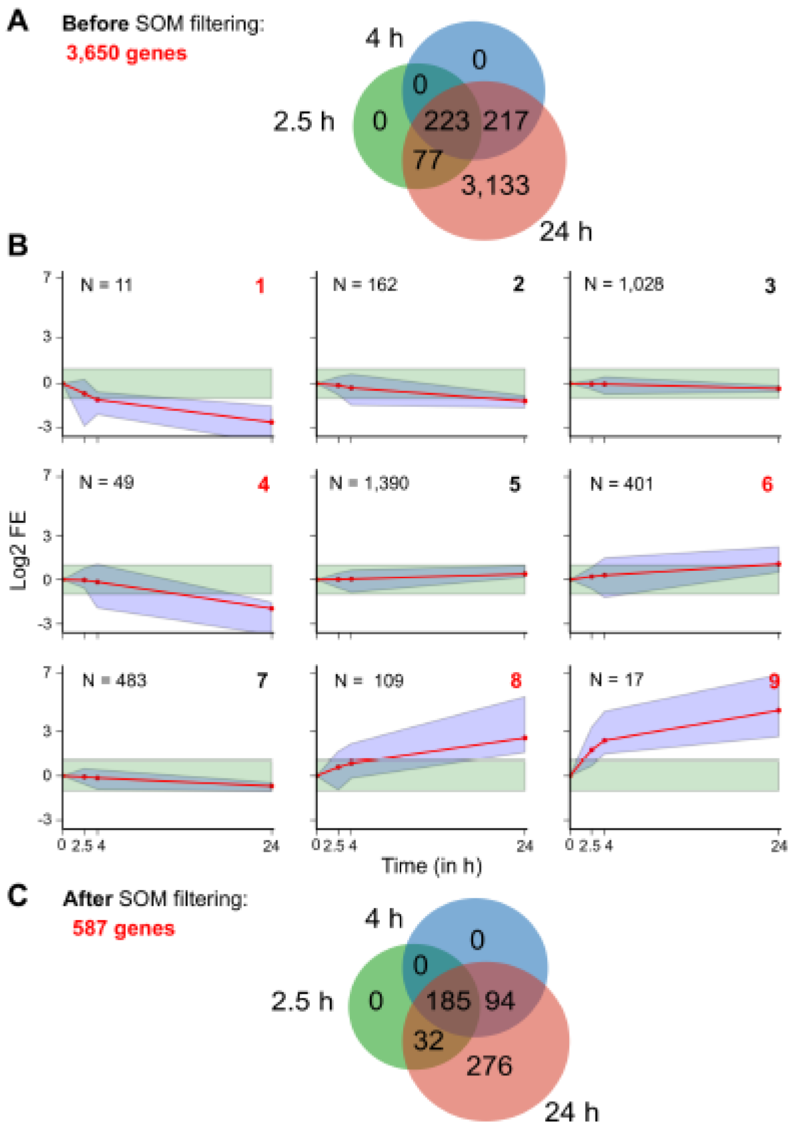
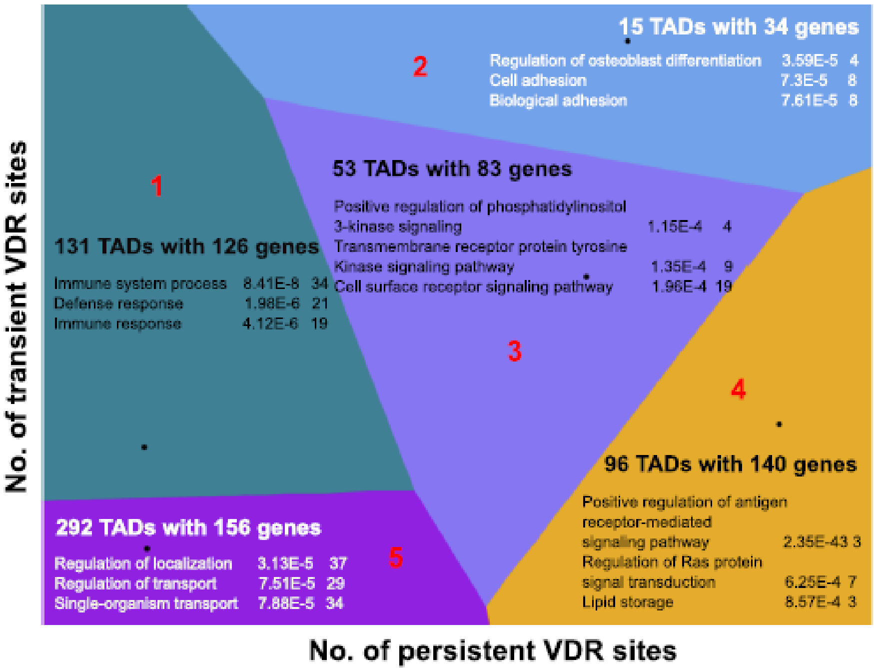
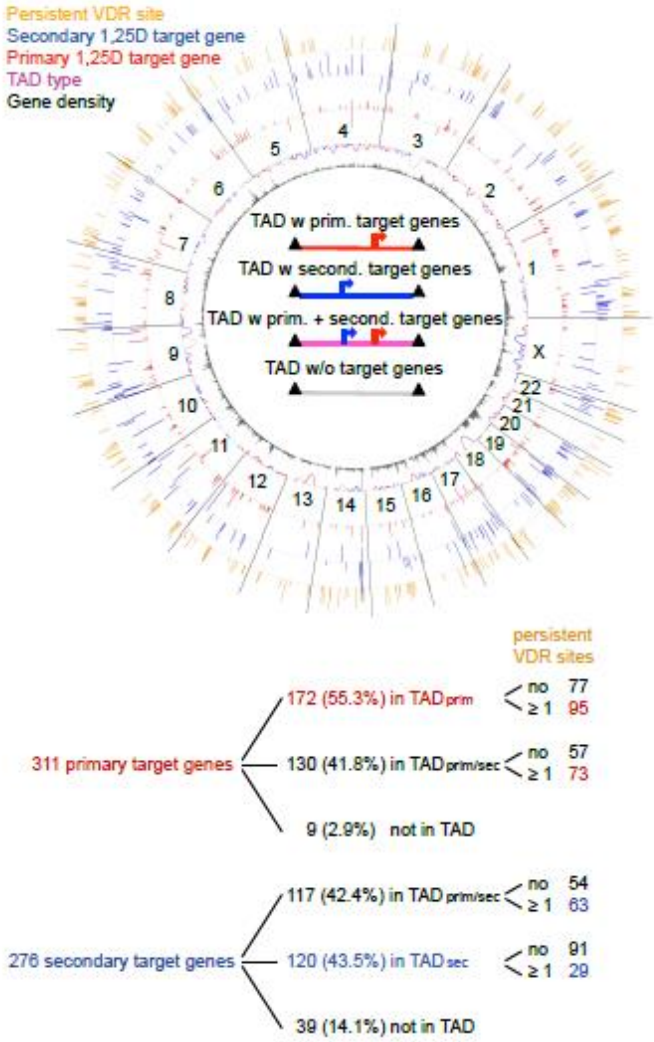


Fig. 3


No. of transient VDR sites
No. of persistent VDR sites
1
131 TADs with 126 genes
Immune system process 8.41E-8 34
Defense response 1.98E-6 21
Immune response 4.12E-6 19
2
15 TADs with 34 genes
Regulation of osteoblast differentiation 3.59E-5 4
Cell adhesion 7.3E-5 8
Biological adhesion 7.61E-5 8
3
53 TADs with 83 genes
Positive regulation of phosphatidylinositol 3-kinase signaling 1.15E-4 4
Transmembrane receptor protein tyrosine Kinase signaling pathway 1.35E-4 9
Cell surface receptor signaling pathway 1.96E-4 19
4
96 TADs with 140 genes
Positive regulation of antigen receptor-mediated signaling pathway 2.35E-43 3
Regulation of Ras protein signal transduction 6.25E-4 7
Lipid storage 8.57E-4 3
5
292 TADs with 156 genes
Regulation of localization 3.13E-5 37
Regulation of transport 7.51E-5 29
Single-organism transport 7.88E-5 34

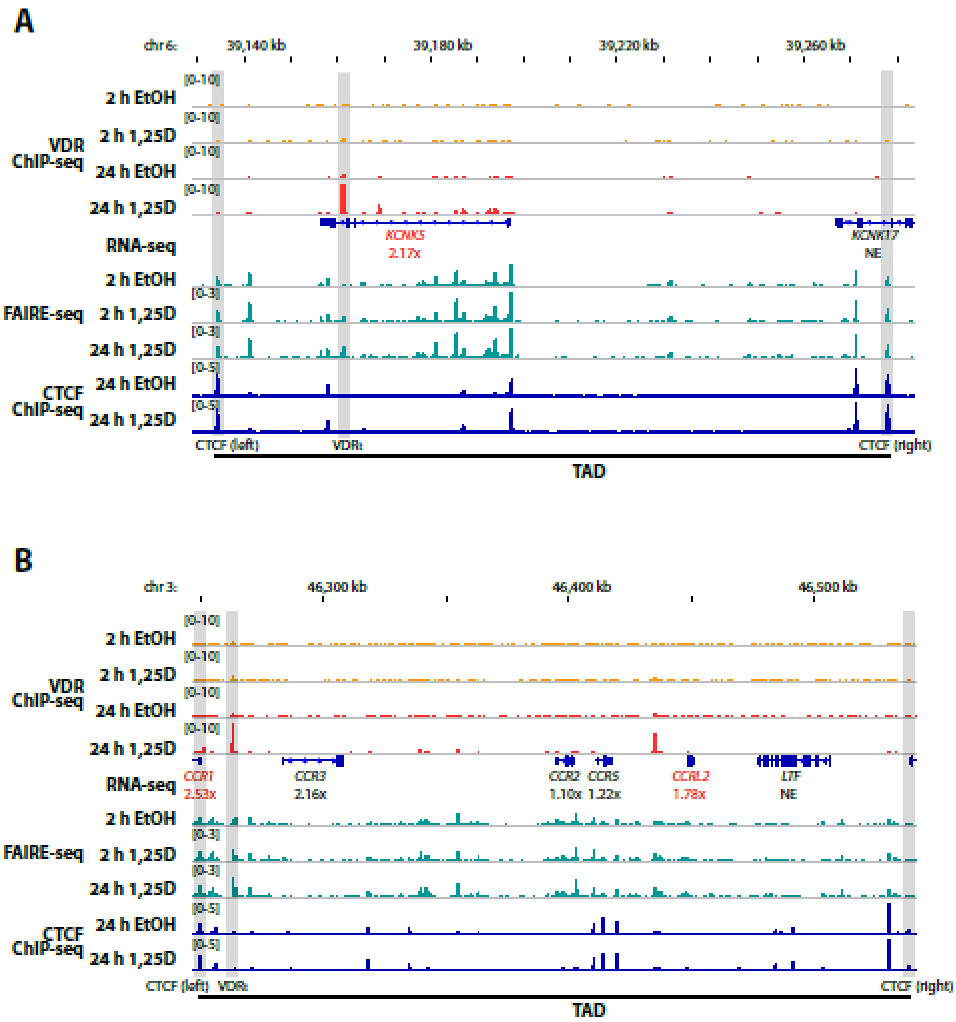
ACCEPTED MANUSCRIPT

Fig. 4



ACCEPTED MANUSCRIPT

Fig. 5



ACCEPTED MANUSCRIPT

Conflict of Interest

There is no conflict of interests between the authors and all authors disclose any financial interest that might be constructed to influence the results or interpretation.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

There are 510 persistent VDR sites in 1,25(OH)$_2$D$_3$-stimulated THP-1 cells.

311 of 587 vitamin D target genes are primary.

97% of the primary vitamin D target genes are located within 1,25(OH)$_2$D$_3$-modulated TADs.

There are TAD classes carrying vitamin D target genes involved in distinct biological processes.

ACCEPTED MANUSCRIPT