

# VITAMIN D RECEPTOR BINDING WITH DNA IN DUODENAL CRYPT, DUODENAL VILLI, & COLONIC EPITHELIAL CELLS OF MICE

DENNIS A. ALDEA, ROHIT AITA, SOHAIB HASSAN,  
EVAN S. COHEN, JOSEPH HUR,  
OSCAR PELLÓN-CÁRDENAS, LEI CHEN,  
MICHAEL P. VERZI (FACULTY ADVISOR)

## \* ABSTRACT

Vitamin D receptor (VDR) is a transcription factor that mediates calcium absorption by intestinal epithelial cells. Although calcium absorption is canonically thought to occur only in the small intestine, recent studies have shown that VDR activity in the colon alone is sufficient to prevent calcium deficiency in mice. Here, we further investigate VDR activity in the colon. We assess VDR-DNA binding in mouse duodenal crypt, duodenal villi, and colonic epithelial cells using Chromatin Immunoprecipitation sequencing (ChIP-seq). We find that most VDR-responsive elements are common to all intestinal epithelial cells, though some VDR-responsive elements are regionally-enriched and exhibit greater VDR-binding affinity in either duodenal epithelial cells or colonic epithelial cells. We also assess chromatin accessibility in the same three cell types using Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq). By integrating the VDR ChIP-seq and ATAC-seq data, we find that regionally-enriched VDR-responsive elements exhibit greater chromatin acces-

sibility in the region of their enrichment. Finally, we assess the transcription factor motifs present in VDR-responsive elements. We find that duodenum- and colon-enriched VDR-responsive elements exhibit different sets of transcription factor motifs other than VDR, suggesting that VDR may act together with different partner transcription factors in the two regions. Our work is the first investigation of VDR-DNA binding in the colon and provides a basis for further investigations of VDR activity in the colon.

## 1 INTRODUCTION

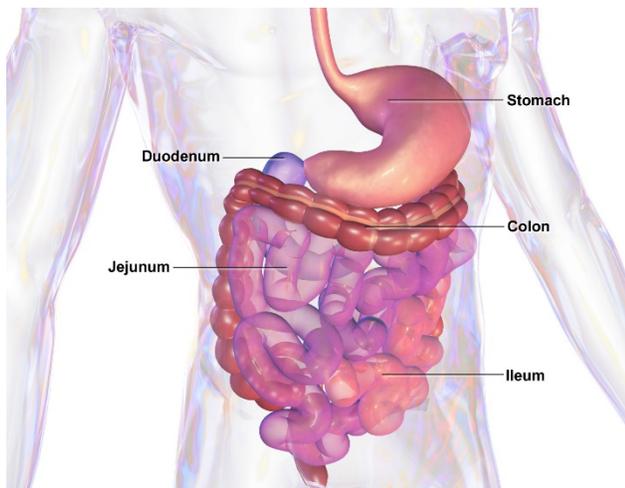
### INTESTINAL ANATOMY

The small and large intestines are a pair of digestive organs that form the lower half of the vertebrate gastrointestinal tract following the stomach. The small intestine is divided into the duodenum, the jejunum, and the ileum (FIGURE 1). The large intestine follows the small intestine and is divided into the cecum, the colon, and the rectum. Both the small and large intestines are lined by the intestinal epithelium—a layer of cells separating the contents of the intestines from the rest of the body. In the small intestine, the intestinal epithelium is a tightly folded structure consisting of villi—finger-like protrusions, and crypts—thin invaginations between villi.<sup>[6]</sup> The folded structure of the intestinal epithelium increases the surface area available for nutrient absorption in the small intestine. Villi are composed of differentiated epithelial cells, and crypts are composed of undifferentiated intestinal stem cells. Intestinal stem cells in the crypts differentiate and migrate to the villi, maintaining the intestinal epithelium even as epithelial cells are constantly shed from the villi.<sup>[5]</sup> Villi are not present in the colon; instead, the colonic epithelium is a heterogeneous mixture of differentiated epithelial cells and undifferentiated stem cells.<sup>[6]</sup>

### VITAMIN D RECEPTOR

Vitamin D receptor (VDR) is a transcription factor present in intestinal epithelial cells. Transcription factors are proteins that regulate the transcription of genes from DNA to RNA. Transcription factors bind to particular genomic regions near their target genes, inducing conformational changes in the DNA





**FIGURE 1: THE GASTROINTESTINAL TRACT**<sup>[2]</sup>

*The small and large intestines are the two organs of the lower gastrointestinal tract. The small intestine follows the stomach and is divided into the duodenum, the jejunum, and the ileum. The large intestine follows the small intestine and is divided into the cecum (not shown), the colon, and the rectum (not shown).*

that either promote or inhibit transcription of the target genes. Most transcription factors can only bind to genomic regions located in open chromatin—DNA that is not condensed by histone proteins. A transcription factor's responsive elements are the open genomic regions to which the transcription factor binds. A transcription factor's binding specificity results from the presence of particular nucleotide sequences, called transcription factor motifs, in the transcription factor's responsive elements.

VDR binds calcitriol (1,25-dihydroxyvitamin D<sub>3</sub>), which is the activated form of vitamin D. Vitamin D, whether synthesized by skin cells as vitamin D<sub>3</sub> or obtained from the diet as vitamin D<sub>2</sub>, is inactive and must be metabolized in the kidneys to calcitriol before binding with VDR. From the kidneys, calcitriol enters the bloodstream and ultimately ligates with and activates VDR in intestinal epithelial cells. Once activated by calcitriol, VDR induces the transcription of genes required for calcium absorption by the intestinal epithelial cells.

VDR dysfunction is implicated in inflammatory bowel disease (IBD) and colorectal cancer (CRC). Certain atypical variants of the *VDR* gene are associated with increased prevalence of IBD.<sup>[8]</sup> IBD patients also exhibit reduced levels of VDR in the intestinal epithelium. These findings suggest a link between VDR dysfunction and IBD. Clinical studies of CRC patients have shown that different variants of *VDR* correlate with differences in patient survival.<sup>[21]</sup> Metastatic CRC tumor cells exhibit reduced VDR expression compared to typical intestinal cells.<sup>[16]</sup> Reduced levels of VDR expression enhance the abnormal Wnt/ $\beta$ -catenin pathway that drives the growth of most CRC tumors.<sup>[11]</sup> Although the precise role of VDR deficiency in CRC tumor formation is still unknown, these findings suggest that VDR inhibition drives the metastasis of CRC tumors.

One of the best appreciated roles of VDR is its mediation of calcium absorption in the small intestine. Knockout studies investigate the function of a protein by inactivating the gene encoding the protein and observing the effect. When *Vdr* is knocked out in both the small and large intestines, the *Vdr*-knockout mice develop calcium deficiencies and ultimately rickets, a disease characterized by severe reduction in bone density.<sup>[20]</sup> However, when *Vdr* is knocked out in the small intestine but not in the colon, the mice are still able to absorb enough calcium to avoid developing rickets.<sup>[7]</sup> Though these studies show that VDR is active in the colon, the differences in VDR-DNA binding between the duodenum and colon have not yet been investigated.

#### RESEARCH OBJECTIVES

Here, we investigate VDR-DNA binding in the duodenum and colon using laboratory mice (*Mus musculus*) as model organisms. Mice and humans both express VDR, and *Vdr*-knockout mice exhibit a phenotype analogous to severe vitamin D deficiency in humans.<sup>[20]</sup> These findings suggest that the regulatory role of VDR is similar in mice and humans.

We assess VDR-DNA binding in mouse duodenal villi, duodenal crypt, and colonic epithelial cells using Chromatin Immunoprecipitation sequencing (ChIP-seq). By comparing VDR-DNA bind-

ing across cell types, we identify VDR-responsive elements that are regionally-enriched—exhibiting greater VDR-binding affinity in either the duodenum or in the colon. These regionally-enriched VDR-responsive elements cannot result from differences in VDR motif presence, since genomic sequences do not differ between cell types. We hypothesize that differences in VDR-DNA binding result from differences in open chromatin regions between cell types.

We examine this hypothesis by assessing chromatin accessibility in each cell type using Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq). By integrating the VDR ChIP-seq and the ATAC-seq data, we find that regionally-enriched VDR-responsive elements exhibit greater chromatin accessibility in the region of their enrichment. Thus, we conclude that differential chromatin accessibility causes differential VDR-DNA binding between intestinal regions.

Finally, we assess the transcription factor binding site motifs present in VDR-responsive elements. We find that duodenum- and colon-enriched VDR-responsive elements exhibit different sets of binding site motifs for transcription factors other than VDR, suggesting that VDR may act in conjunction with different partner transcription factors in the two regions.

Our work is the first survey of VDR-responsive elements in the colon. We find that tissue-enriched VDR-responsive elements in duodenal and colonic epithelium cells differ in chromatin accessibility and secondary transcription factor binding motifs. These findings provide a basis for further investigations of differences in VDR-mediated gene expression between the small and large intestines. Understanding the action of VDR specific to the colon may explain the role of VDR in inflammatory bowel disease and colorectal cancer.

## 2 METHODOLOGY

### EXPERIMENTAL MICE

Four wild-type mice (C57BL/6J strain) were used as experimental animals. All animal protocols were approved by the Rutgers Institutional Animal Care and Use Committee. The mice were given food and water *ad libitum* and exposed to a 12 h light, 12 h dark

cycle. The mice were euthanized when they were 4–6 weeks old. One hour prior to euthanasia, the mice were treated with 10 ng/g body mass 1,25-dihydroxyvitamin D3 (Cayman Chemical: #11792), injected intraperitoneally.

### SAMPLE PREPARATION

Tissue samples were harvested from duodenal villi, duodenal crypts, and colonic epithelium of the mice immediately following euthanasia. Duodenal villi, duodenal crypts, and colonic epithelial cells were extracted from the tissue samples and pelleted by centrifugation.<sup>[4]</sup>

### CHIP-SEQ PROTOCOL & ANALYSIS

VDR-targeted ChIP-seq was conducted on the duodenal villi, duodenal crypts, and colonic epithelial cells using mouse monoclonal IgG2a VDR antibody D-6 (Santa Cruz Biotechnology: #sc-13133) and rabbit polyclonal IgG VDR antibody C-20 (Santa Cruz Biotechnology: #sc-1008). The VDR binding site library was purified and amplified using QIAquick PCR purification kit #50 (Qiagen: #28104) before being sequenced.

Sequencing adapters were removed from the VDR ChIP-seq read FASTQ files using NGmerge.<sup>[9]</sup> Each pair of forward and reverse adapter-trimmed read FASTQ files was aligned to mouse genome assembly mm9 using Bowtie2.<sup>[10,15]</sup> Each alignment SAM file was converted to an alignment BAM file using the SAMtools suite.<sup>[13]</sup>

A composite VDR ChIP-seq alignment BAM file was generated for each cell type by combining each set of replicate VDR ChIP-seq alignment BAM files using the merge utility in the SAMtools suite. An alignment track BigWig file was generated from each composite and replicate alignment BAM file using the bamCoverage utility in the deepTools suite.<sup>[18]</sup>

VDR-binding peaks were identified from each VDR ChIP-seq alignment BAM file using the callpeak utility in MACS.<sup>[22]</sup> Peaks overlapping ENCODE mm9 blacklisted regions were removed from each peak set BED file using the subtract utility in the BEDtools suite.<sup>[17]</sup> The ENCODE blacklists list genomic regions known to yield false ChIP-seq signals

due to the inaccuracies of a particular genome assembly.<sup>[11]</sup> Each peak set BED file was then shifted so that each peak would be centered on its summit—the nucleotide with the greatest ChIP-seq signal within the peak region, as determined by MACS. The summit of each peak represents the most probable transcription factor binding site; therefore, summit-centering ensures that each peak is centered around the binding site.

For each cell type composite, the VDR ChIP-seq signal was plotted versus distance from the nearest VDR ChIP-seq peak using the SiteproBW program included in the Cistrome suite.<sup>[14]</sup>

#### ATAC-SEQ PROTOCOL & ANALYSIS

ATAC-seq was conducted on the duodenal villi, duodenal crypts, and colonic epithelial cells using Nextera Tn5 transposase (Illumina: #FC-121-1030). The transposed chromatin was purified and amplified using QIAquick PCR purification kit #50 (Qiagen: #28104) before being sequenced.

Sequencing adapters were removed from the ATAC-seq read FASTQ files using NGmerge.<sup>[9]</sup> Each pair of forward and reverse adapter-trimmed read FASTQ files was aligned to mouse genome assembly mm9 using Bowtie2.<sup>[10,15]</sup> Each alignment SAM file was converted to an alignment BAM file using the SAMtools suite.<sup>[13]</sup>

A composite ATAC-seq alignment BAM file was generated for each cell type by combining each set of replicate ATAC-seq alignment BAM files using the merge utility in the SAMtools suite. An alignment track BigWig file was generated from each composite and replicate alignment BAM file using the bamCoverage utility in the deepTools suite.<sup>[18]</sup>

The median alignment size of each ATAC-seq alignment BAM file was determined using the CollectInsertSizeMetrics utility included in the Picard suite.<sup>[3]</sup> Each alignment BAM file was converted to a BED file using the bamtoBED utility included in the BEDtools suite.<sup>[17]</sup> Peak region BED files were called from each alignment BED file using MACS.<sup>[22]</sup> It was necessary to convert the alignment BAM files to BED files so that MACS would properly interpret the non-overlapping forward and reverse peaks typical of ATAC-seq but not of ChIP-seq. MACS was run with a

shift distance of negative one-half the median alignment size and an extsize distance equal to the median alignment size for each alignment BED file.

For each cell type composite, the ATAC-seq signal was plotted versus distance from the nearest ATAC-seq peak using the SiteproBW program included in the Cistrome suite.<sup>[14]</sup>

#### DIFFERENTIAL BINDING ANALYSIS

Peaks exhibiting differential VDR-binding affinities between cell types were identified using DiffBind.<sup>[19]</sup> The following contrasts were examined: colonic epithelium versus duodenal crypts, colonic epithelium versus duodenal villi, and duodenal crypts versus duodenal villi. Each set of differentially bound peaks was filtered to only include peaks assigned a significance value less than 0.001 by DiffBind. Each filtered peak set was exported to a BED file.

For each contrast, the VDR ChIP-seq signals of each contrasted cell type were plotted versus the enriched VDR ChIP-seq peak sets for each contrasted cell type using the SiteproBW program included in the Cistrome suite.<sup>[14]</sup>

#### MOTIF ANALYSIS

Transcription factor motifs were identified from the composite VDR ChIP-seq peak set BED files of each cell type using HOMER. Due to limited computational resources, the size of each composite peak set was reduced by random sampling. Sample peak set BED files were generated by randomly selecting one-fifth of the peaks in each composite peak set BED file. HOMER was used to identify motifs enriched in each sample peak set BED file.

Transcription factor motifs were also identified from each differential VDR ChIP-seq peak set BED file using HOMER. However, the smaller size of the differential peak sets meant that sampling was not necessary; the entirety of each differential peak set BED file was analyzed using HOMER.

## 3 RESULTS

#### GENOME ALIGNMENT

All of the VDR ChIP-seq read FASTQ files were successfully aligned to mouse genome assembly

SAMPLE <sup>[A]</sup>	READS (A)	ALIGNMENTS <sup>[B]</sup> (B)	ALIGNMENT RATE (B/A × 100%)	UNIQUE ALIGNMENTS <sup>[C]</sup> (C)	UNIQUE ALIGNMENT RATE (C/B × 100%)
COLON 1	13,100,994	12,125,236	92.55%	9,127,982	75.28%
COLON 2	13,333,617	12,883,875	96.63%	10,186,317	79.06%
COLON 3	18,445,285	17,731,640	96.13%	14,716,273	82.99%
COLON 4	16,672,121	15,935,857	95.58%	12,561,570	78.83%
<i>COLON COMPOSITE</i>	—	<i>58,676,608</i>	—	<i>46,166,439</i>	<i>78.68%</i>
CRYPT 1	12,750,045	10,859,762	85.17%	3,100,376	28.55%
CRYPT 2	13,937,899	12,806,269	91.88%	9,021,688	70.45%
CRYPT 3	12,336,135	11,724,422	95.04%	5,819,723	49.64%
CRYPT 4	15,028,118	14,458,601	96.21%	7,679,487	53.11%
<i>CRYPT COMPOSITE</i>	—	<i>49,849,074</i>	—	<i>25,325,174</i>	<i>50.80%</i>
VILLI 1	12,503,339	11,740,518	93.90%	6,495,870	55.33%
VILLI 2	13,991,816	13,096,292	93.60%	11,134,576	85.02%
VILLI 3	18,795,355	17,970,894	95.61%	11,094,953	61.74%
VILLI 4	11,094,953	10,588,783	95.44%	4,858,968	45.89%
<i>VILLI COMPOSITE</i>	—	<i>53,396,487</i>	—	<i>34,583,701</i>	<i>64.77%</i>

TABLE 1: VDR CHIP-SEQ GENOME ALIGNMENT METRICS

[A] Composite alignment BAM files were constructed using SAMtools merge.<sup>[14]</sup>

[B] Genome alignment was conducted using mouse genome assembly mm9 and Bowtie2.<sup>[10,15]</sup>

[C] Duplicate alignments were removed using MACS filterdup.<sup>[22]</sup>

mm9.<sup>[15]</sup> The successful genomic alignments are indicated by the high alignment rates; all samples exhibited alignment rates greater than 85%, and all but one sample exhibited alignment rates greater than 90% (TABLE 1).

#### PEAK CALLING

The accuracy of VDR ChIP-seq peak calling was assessed by comparing each cell type's composite peak set BED file to its composite track BigWig file. The resulting signal plots indicate that all of the com-

posite peak sets exhibit a majority of reads near the centers of peak regions (FIGURE 2). Thus, we confirm that the VDR-responsive elements identified by MACS exhibit elevated VDR binding, as expected of responsive elements.

#### DIFFERENTIAL BINDING ANALYSIS

The correlation between VDR ChIP-seq peak sets was assessed using DiffBind.<sup>[19]</sup> The resulting correlation matrix indicates that the colonic epithelium peak sets are all more closely correlated to each

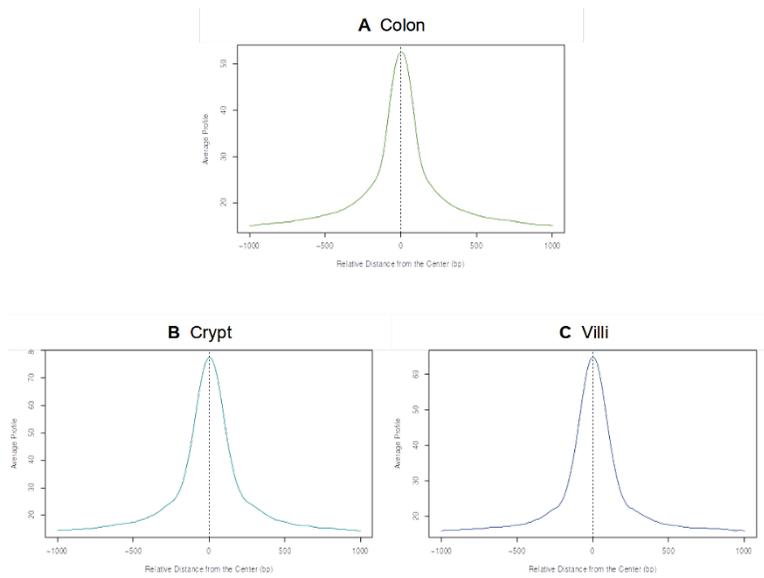


FIGURE 2: VDR CHIP-SEQ SIGNAL VERSUS DISTANCE FROM VDR CHIP-SEQ PEAKS

Most VDR Chip-seq reads are located near VDR Chip-seq peaks. VDR Chip-seq signal versus distance from nearest VDR Chip-seq peak. Figure generated using SiteproBW.<sup>[14]</sup>

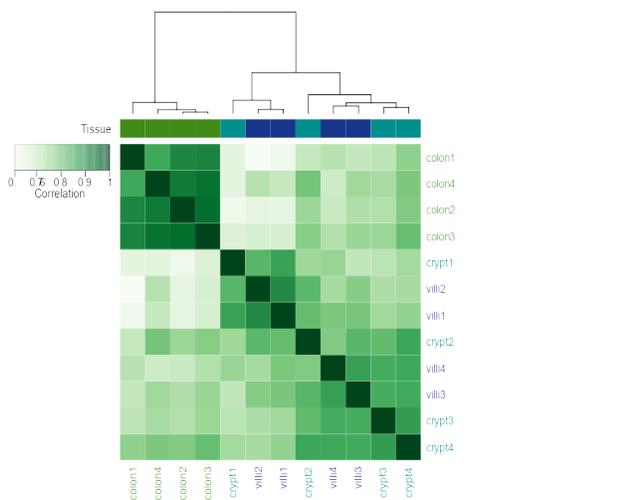


FIGURE 3: VDR CHIP-SEQ DIFFERENTIAL BINDING HEATMAP

The difference between colonic and duodenal VDR Chip-seq peak sets is greater than the difference between duodenal crypt and duodenal villi VDR Chip-seq peak sets. Darker shading in the heatmap represents greater similarity between peak sets; greater vertical separation in the dendrogram represents greater difference between peak sets. Figure generated by DiffBind.<sup>[19]</sup>

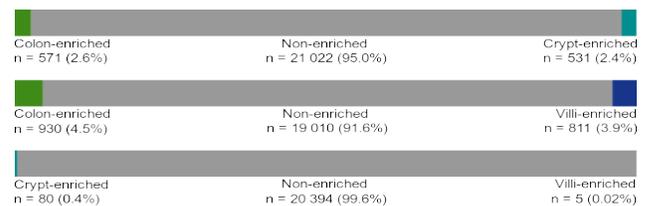
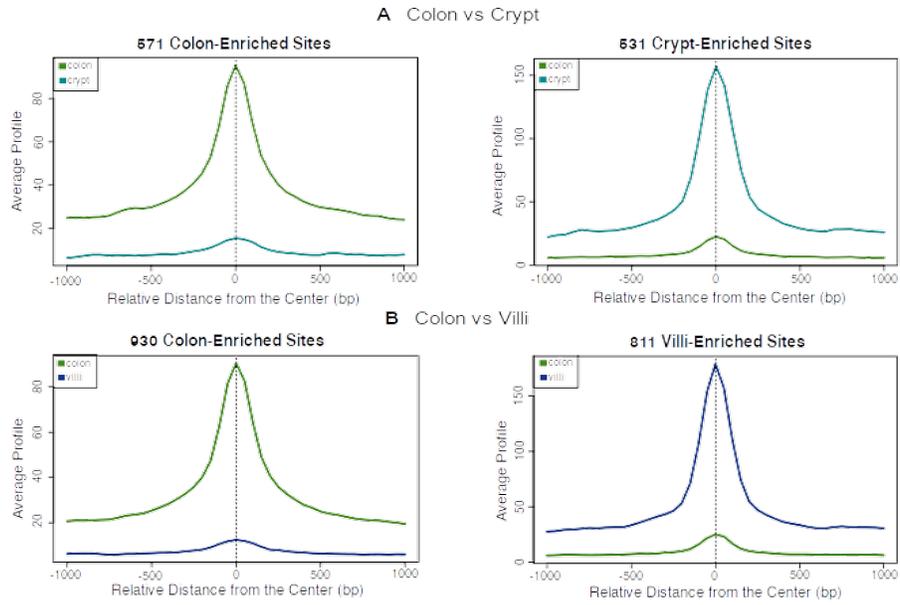


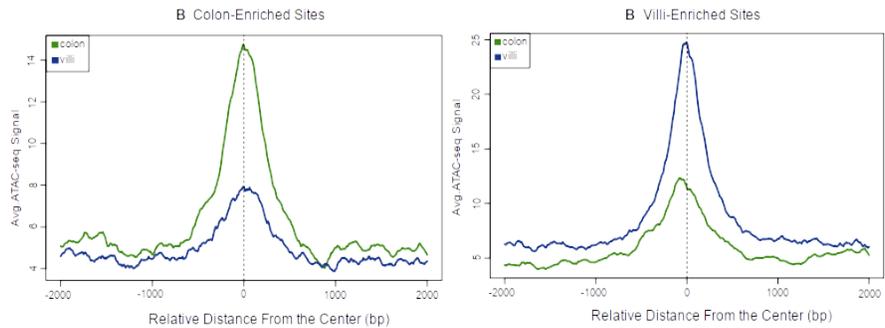
FIGURE 4: DIFFERENTIAL VDR CHIP-SEQ PEAK SET SIZES

The majority of VDR Chip-seq peaks are common to duodenal villi, duodenal crypt, and colonic epithelial cells. A small minority of VDR Chip-seq peaks differ between duodenal epithelial and colonic epithelial cells.



**FIGURE 5: VDR CHIP-SEQ SIGNAL VERSUS DISTANCE FROM DIFFERENTIAL VDR CHIP-SEQ PEAKS**

*Differential VDR ChIP-seq peaks exhibit greater VDR binding in the tissue of their enrichment. VDR ChIP-seq signal versus distance from nearest differential VDR ChIP-seq peak. Figure generated by SiteproBW.<sup>[14]</sup>*



**FIGURE 6: ATAC-SEQ SIGNAL VERSUS DISTANCE FROM DIFFERENTIAL VDR CHIP-SEQ PEAKS**

*Differential VDR ChIP-seq peaks exhibit greater chromatin accessibility in the tissue of their enrichment. ATAC-seq signal versus distance from nearest differential VDR ChIP-seq peak. Figure generated by SiteproBW.<sup>[14]</sup>*

other than to any of the duodenal crypt or duodenal villi peak sets (FIGURE 3). The correlation matrix also indicates that the duodenal crypt and duodenal villi peak sets do not differ significantly.

The majority of VDR ChIP-seq peaks are common to all three cell types. Out of the 23,381 VDR-binding sites exhibited in either colonic epithelial or duodenal villi cells, only 1,741 sites differ be-

SAMPLE <sup>[A]</sup>	PEAKS <sup>[B]</sup> (A)	NON-BLACKLISTED PEAKS <sup>[C]</sup> (B)	BLACKLIST PASS RATE (B/A × 100%)
COLON 1	1,762	1,484	84.22%
COLON 2	5,013	4,617	92.10%
COLON 3	6,143	5,701	92.80%
COLON 4	10,108	9,608	95.05%
<i>COLON COMPOSITE</i>	<i>23,881</i>	<i>22,888</i>	<i>95.84%</i>
CRYPT 1	13,382	12,838	95.93%
CRYPT 2	2,192	1,810	82.57%
CRYPT 3	1,138	954	83.83%
CRYPT 4	2,481	2,212	89.16%
<i>CRYPT COMPOSITE</i>	<i>13,488</i>	<i>12,719</i>	<i>94.30%</i>
VILLI 1	10,387	9,720	93.58%
VILLI 2	6,002	5,383	89.69%
VILLI 3	1,539	1,225	79.60%
VILLI 4	3,780	3,585	94.84%
<i>VILLI COMPOSITE</i>	<i>19,044</i>	<i>18,083</i>	<i>94.95%</i>

TABLE 2: VDR CHIP-SEQ PEAK CALLING METRICS

[A] Composite alignment BAM files were constructed using SAMtools merge.<sup>[14]</sup>

[B] Peak calling was conducted using MACS.<sup>[22]</sup>

[C] Peaks included in the ENCODE mm9 blacklist were removed.<sup>[1]</sup>

CONTRAST	ENRICHMENT	ENRICHED PEAKS <sup>[A]</sup> (A)	SIGNIFICANT PEAKS <sup>[B]</sup> (B)	SIGNIFICANCE RATE (B/A × 100%)
COLON VS. CRYPT	COLON	1,243	571	45.94%
	CRYPT	1,073	531	49.49%
COLON VS. VILLI	COLON	1,823	930	51.01%
	VILLI	1,637	811	49.54%
CRYPT VS. VILLI	CRYPT	80	80	100.00%
	VILLI	5	5	100.00%

TABLE 3: DIFFERENTIAL VDR CHIP-SEQ PEAK SET METRICS

[A] Enriched peaks were determined using DiffBind.<sup>[19]</sup>

[B]  $p < 0.001$

tween the two cell types (TABLE 3). There is a comparable difference between colonic epithelial and duodenal crypt cells, which differ in only 1,102 VDR-binding sites out of 20,751 sites total. Nonetheless, these small differences are far greater than the minuscule difference between duodenal crypt and duodenal villi cells, which differ by only 85 VDR-binding sites out of 20,479 sites total (FIGURE 4).

The accuracy of the differential binding analysis performed by DiffBind was assessed by comparing each cell type's enriched peak set BED file against all of the composite track BigWig files. The resulting signal plots indicate that all of the tissue-enriched peak sets exhibit greater overlap with VDR-responsive elements in the tissue of their enrichment than with VDR-binding sites in other tissues (FIGURE 5). Thus, we confirm that the tissue-enriched VDR-responsive elements identified by DiffBind exhibit elevated VDR binding in the tissue of their enrichment, as expected of tissue-enriched responsive elements.

COMPARISON OF VDR CHIP-SEQ & ATAC-SEQ

To investigate our hypothesis that regionally-enriched VDR binding results from differential chromatin accessibility between tissues, we compared duodenum- and colon-enriched VDR-binding peaks BED files against all of the composite ATAC-seq BigWig files. The resulting signal plots indicate that all of the regionally-enriched VDR ChIP-seq peak sets exhibit greater overlap with open chromatin sites in the region of their enrichment than with open chromatin sites in the other region (FIGURE 6). Thus, we conclude that differences in VDR binding result from differences in chromatin accessibility between regions.

MOTIF FINDING

Due to limited computational resources, motif finding was conducted on samples generated by randomly selecting one-fifth of the peaks in each composite VDR ChIP-seq peak set (TABLE 2).

VDR is the most significant motif present in any of the composite VDR ChIP-seq peak set samples (FIGURE 7), indicating that the VDR ChIP-seq was performed correctly. Besides VDR motifs, the colon-enriched peak set samples also exhibit HOXB13,

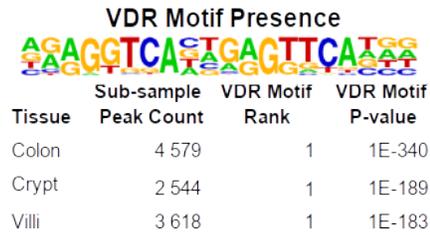


FIGURE 7: VDR MOTIF PRESENCE IN VDR CHIP-SEQ PEAK SET SAMPLES

*VDR motif was the top transcription factor motif identified in all of the composite VDR ChIP-seq peak sets. Motifs were identified using HOMER on a random sample of one-fifth of each composite peak set*

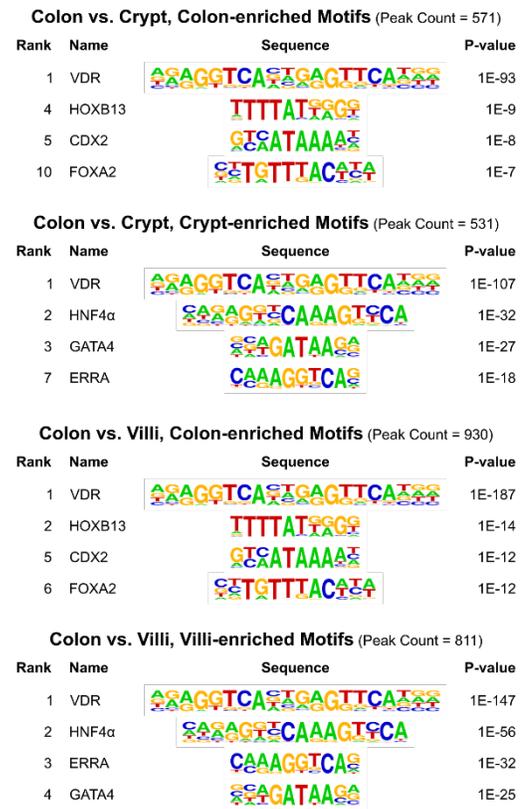


FIGURE 8: VDR AND SECONDARY MOTIF PRESENCES IN DIFFERENTIAL VDR CHIP-SEQ PEAK SETS

*Colon- and duodenum-enriched VDR ChIP-seq peaks exhibit different secondary transcription factor motifs other than VDR. Motifs were identified using HOMER on each differential peak set.*

CDX2, and FOXA2 motifs, whereas the crypt- and villi-enriched peak set samples exhibit HNF4 $\alpha$ , ERRA, and GATA4 (FIGURE 8). These results suggest that differences in VDR binding between the duodenal epithelium and colonic epithelium may result from differences in the helper transcription factors that facilitate VDR-DNA binding.

## 4 DISCUSSION

We found that the VDR-binding profile of colonic epithelial cells is largely similar to that of duodenal villi and duodenal crypt cells. Out of about twenty thousand VDR binding sites total, only a minority of several hundred binding sites differ between colonic epithelial cells and duodenal epithelial cells. Nonetheless, this difference is greater than that between duodenal villi and duodenal crypt cells, which differ in less than one hundred binding sites.

By comparing VDR ChIP-seq and ATAC-seq data, we found that colon- and duodenum-enriched VDR-binding sites exhibited greater chromatin accessibility in the tissue of their enrichment. We also determined that colon- and duodenum-enriched VDR-binding sites exhibit distinct sets of secondary (i.e. non-VDR) transcription factor motifs. Colon-enriched VDR-binding sites exhibit HOXB13, CDX2, and FOXA2 motifs; duodenum-enriched VDR-binding sites exhibit HNF4 $\alpha$ , GATA4, and ERRA motifs. These findings concur with those of a previous investigation which found that VDR-binding sites in the duodenum exhibited HNF4 $\alpha$  and GATA4 motifs in addition to VDR motifs.<sup>[12]</sup>

Tissue-specific secondary transcription factors may cause differential VDR binding, either by causing the differences in open chromatin observed at differential binding sites, or by directly binding to VDR and affecting VDR-DNA binding. Possible binding interactions between VDR and secondary transcription factors could be examined using protein immunoprecipitation assays. Additional investigations can be conducted using intestinal organoid models. By knocking out secondary transcription factors in intestinal organoids, the role of these factors in VDR-mediated regulation of gene expression could be determined. Such investigations would ad-

vance our understanding of VDR's role in intestinal health and diseases, including colorectal cancers, and possibly offer new treatments for those affected by these conditions.

## 5 DATA & SOURCE CODE ACCESS

The complete data collected in this investigation are available upon request. This investigation did not involve human subjects, and these data do not include HIPAA-protected health information.

The source code for the analysis pipeline used in this investigation is available upon request. The following programs were used in the analysis pipeline: NGmerge v0.3,<sup>[9]</sup> Bowtie2 v2.2.6,<sup>[10]</sup> SAMtools v0.1.19,<sup>[13]</sup> deepTools v3.3.0,<sup>[18]</sup> MACS v2.1.0,<sup>[22]</sup> BEDtools v2.17.0,<sup>[17]</sup> Cistrome v0.6.7,<sup>[14]</sup> Picard v2.18.27,<sup>[3]</sup> DiffBind v1.16.3,<sup>[19]</sup> HOMER v4.8.3 ■

## 6 REFERENCES

- [1] Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE blacklist: Identification of problematic regions of the genome. *Scientific Reports*, 9(1), 9354.
- [2] Blausen Medical Communications. (2014). Blausen 0432 GastroIntestinalSystem.png. [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/FILE:BLAUSEN\\_0432\\_GASTROINTESTINALSYSTEM.PNG](https://commons.wikimedia.org/wiki/File:Blausen_0432_GastroIntestinalSystem.png)
- [3] Broad Institute. Picard. [HTTPS://BROADINSTITUTE.GITHUB.IO/PICARD](https://broadinstitute.github.io/picard)
- [4] Chen, L., Toke, N. H., Luo, S., Vasoya, R. P., Fullem, R. L., Parthasarathy, A., Perekatt, A. O., & Verzi, M. P. (2019). A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. *Nature Genetics*, 51, 777-785.
- [5] Crosnier, C., Stamatakis, S., & Lewis, J. (2006). Organizing cell renewal in the intestine: Stem cells, signals and combinatorial control. *Nature Reviews Genetics*, 7(5), 349-359.
- [6] de Santa Barbara, P., van den Brink, G. R., & Roberts, D. J. (2003). Development and differentiation of the intestinal epithelium. *Cellular and Molecular Life Sciences*, 60(7), 1322-1332.
- [7] Dhawan, P., Veldurthy, V., Yehia, G., Hsaio, C., Porta, A., Kim, K., Patel, N., Lieben, L., Verlinden, L., Carmeliet, G., & Christakos, S. (2017). Transgenic expression of the vitamin D receptor restricted to the ileum, cecum, and colon of vitamin D receptor knockout mice rescues vitamin D receptor-dependent rickets. *Endocrinology*, 158(11), 3792-3804.
- [8] Eloranta, J. J., Wenger, C., Mwinyi, J., Hiller, C., Gubler, C., Vavricka, S. R., Fried, M., Kullak-Ublick, G. A., & Swiss IBD Cohort Study Group. (2011). Association of a common vitamin D-binding protein polymorphism with inflammatory bowel disease. *Pharmacogenetics and Genomics*, 21(9), 559-564.
- [9] Gaspar, J. M. (2018). NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*, 19, 536.

- [10] Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359.
- [11] Larriba, M. J., Ordóñez-Morán, P., Chicote, I., Martín-Fernández, G., Puig, I., Muñoz, A., & Pálmer, H. G. (2011). Vitamin D receptor deficiency enhances Wnt/ $\beta$ -catenin signaling and tumor burden in colon cancer. *PLoS One*, 6(8), e23524.
- [12] Lee, S. M., Riley, E. M., Meyer, M. B., Benkusky, N. A., Plum, L. A., DeLuca, H. F., & Pike, J. W. (2015). 1,25-Dihydroxyvitamin D3 controls a cohort of vitamin D receptor target genes in the proximal intestine that is enriched for calcium-regulating components. *Journal of Biological Chemistry*, 290(29), 18199-18215.
- [13] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- [14] Liu, T., Ortiz, J. A., Taing, L., Meyer, C. A., Lee, B., Zhang, Y., Shin, H., Wong, S. S., Ma, J., Lei, Y., Pape, U. J., Poldinger, M., Chen, Y., Yeung, K., Brown, M., Turpaz, Y., & Liu, X. S. (2011). Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biology*, 12(8), R83.
- [15] Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), 520-562.
- [16] Pálmer, H. G., Larriba, M. J., García, J. M., Ordóñez-Morán, P., Peña, C., Peiró, S., Puig, I., Rodríguez, R., de la Fuente, R., Bernad, A., Pollán, M., Bonilla, F., Gamallo, C., de Herreros, A. G., & Muñoz, A. (2004). The transcription factor SNAIL represses vitamin D receptor expression and responsiveness in human colon cancer. *Nature Medicine*, 10(9), 917-919.
- [17] Quinlan, A. R., & Hall, I. M. (2010). BEDtools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842.
- [18] Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2014). DeepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160-W165.
- [19] Stark, R., & Brown, D. (2011). DiffBind: Differential binding analysis of ChIP-seq peak data. [HTTPS://BIOCONDUCTOR.ORG/PACKAGES/RELEASE/BIOC/VIGNETTES/DIFFBIND/INST/DOC/DIFFBIND.PDF](https://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf)
- [20] Suda, T., Masuyama, R., Bouillon, R., & Carmeliet, G. (2015). Physiological functions of vitamin D: What we have learned from global and conditional VDR knockout mouse studies. *Current Opinion in Pharmacology*, 22, 87-99.
- [21] Vaughan-Shaw, P. G., O'Sullivan, F., Farrington, S. M., Theodoratou, E., Campbell, H., Duntop, M. G., & Zgaga, L. (2017). The impact of vitamin D pathway genetic variation and circulating 25-hydroxyvitamin D on cancer outcome: Systematic review and meta-analysis. *British Journal of Cancer*, 116(8), 1092-1110.
- [22] Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biology*, 9(9), R137.



Dennis Aldea is a senior undergraduate student in the genetics program at Rutgers New Brunswick. His interest in scientific research was sparked by his experience in the Rutgers Waksman Student Scholars Program, a biotechnology outreach program for high school students. Dennis currently works with Dr. Michael Verzi at the Human Genetics Institute of New Jersey, where he conducts bioinformatics analyses to investigate the genetic basis of intestinal diseases and cancers.