Published in final edited form as:

Hum Immunol. 2011 March ; 72(3): 262–268. doi:10.1016/j.humimm.2010.12.010.

Vitamin D receptor gene methylation is associated with ethnicity, tuberculosis and *TaqI* polymorphism

Charlene Andraos^a, Gerrit Koorsen^a, Julian C Knight^b, and Liza Bornman^{a,*}

^aDepartment of Biochemistry, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa

^bWellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
julian@well.ox.ac.uk

Abstract

The Vitamin D Receptor (*VDR*) gene encodes a transcription factor which, on activation by vitamin D, modulates diverse biological processes including calcium homeostasis and immune function. Genetic variation involving *VDR* shows striking differences in allele frequency between populations and has been associated with disease susceptibility including tuberculosis and autoimmunity, although results have often been conflicting. We hypothesized that methylation of *VDR* may be population specific and that the combination of differential methylation and genetic variation may characterise TB predisposition. We use bisulphite conversion and/or pyrosequencing to analyse the methylation status of 17 CpGs of *VDR* and to genotype 7 SNPs in the 3' CpG Island (CGI 1060), including the commonly studied SNPs *ApaI* (rs7975232) and *TaqI* (rs731236). We show that for lymphoblastoid cell lines from two ethnically diverse populations (Yoruba from HapMap, n=30 and Caucasians, n=30) together with TB cases (n=32) and controls (n=29) from the Venda population of South Africa there are methylation variable positions (MVPs) in the 3' end that significantly distinguish ethnicity (9/17 CpGs) and TB status (3/17 CpGs). Moreover methylation status shows complex association with *TaqI* genotype highlighting the need to consider both genetic and epigenetic variants in genetic studies of *VDR* association with disease.

Keywords

VDR (vitamin D (1,25- dihydroxyvitamin D3) receptor); gene polymorphism; CpG methylation; TB (tuberculosis); ethnic differences

1. Introduction

The Vitamin D Receptor (*VDR*) gene encodes a ligand-activated transcription factor mediating multiple actions of vitamin D including calcium homeostasis, cell growth, cell differentiation, modulation of the immune response and activation of monocyte-macrophages [1-3]. Single nucleotide polymorphisms (SNPs) of *VDR*, specifically in the regulatory region (Cdx2, A1012G), exon 2 (*FokI*) and 3' end (*BsmI*, *ApaI* and *TaqI*; Fig. 1A), have been associated with a number of diseases including tuberculosis (TB), end stage renal disease, cancer, osteoporosis and autoimmune diseases such as systemic lupus erythematosus, cirrhosis and multiple sclerosis [4]. Differential susceptibility to *VDR*-related

*Corresponding author (Liza Bornman): Tel.: +27 11 559 2406; fax: +27 11 559 2605; lizab@uj.ac.za; zacharlene.andraos@gmail.com; gkoorsen@uj.ac.za.

disease between ethnically diverse populations is widely studied and explained from a genetic perspective. For example, the distribution of *VDR* alleles differs between European and African populations [5]. The effects of genetic variation in disease predisposition among populations may be influenced by differential exposure to environmental factors [6], which may be mediated through epigenetic mechanisms [7]. Epigenetic mechanisms may contribute to differential disease susceptibility due to aberrant gene regulation [7]. DNA methylation, an epigenetic mechanism, occurs at Cytosine-phosphate-Guanine dinucleotides (CpGs) in CpG-rich regions called CpG islands (CGIs) and alters transcription factor binding and gene expression either through site-specific or regional mechanisms [8]. The extent of epigenetic diversity between populations regarding *VDR* and its possible role in predisposition to disease is unknown.

TB is a complex, multi-factorial disease influenced by both environmental and genetic factors. Vitamin D restricts *Mycobacterium tuberculosis* (*M. tuberculosis*) growth in macrophages through the production of the anti-microbial peptide, cathelicidin [9]. Genetic variation in *VDR* has been associated with TB [10-14] but the role of epigenetic variation involving *VDR*, in differential TB susceptibility has, to our knowledge, not yet been reported. The consideration of *VDR* methylation as a possible factor contributing to disease predisposition, particularly TB, is motivated by previous studies showing that the *VDR* is indeed epigenetically regulated in certain diseases or conditions. For example, aberrant *VDR* methylation in breast cancer may be involved in chemotherapeutic drug resistance [15] and *VDR* methylation in rat colonic mucosa is reduced by estrogen, leading to an increase in *VDR* production [16]. Furthermore, variably methylated CpGs, known as methylation variable positions (MVPs), are concentrated in genes that regulate transcription, growth, metabolism, differentiation and oncogenesis [17], all of which are functions of the *VDR*. MVPs have the statistical power to discriminate between biological states and diseases [18,19].

Identifying *VDR* methylation as an additional factor contributing to disease susceptibility may also minimise inconsistent *VDR*-disease associations often observed among populations and for traits ranging from bone mass density and diabetes [4] to TB [20] where the association with *TaqI* (T/t or nucleotides T/C) in particular has generated controversy [10,11]. Rather than studying genetic variants in isolation, the common disease genetic epigenetic (CDGE) hypothesis encourages the simultaneous analysis of both genetic and epigenetic variants (MVPs) to construct 'hepitypes' in order to increase the power of associations in epidemiological studies and minimise inconsistent associations [18,21]. Analysing both the genetic and epigenetic characteristics of a gene and their interaction in disease has recently been the focus of a number of studies [22-24]. *Cis* interactions between genetic and epigenetic variants involve the direct modulation of CpG methylation by SNPs [18] and the correlation of methylation with SNPs within 1 Mb of a CpG [24] while *trans* interactions involve regulation effects between CpGs and SNPs from more distant genomic regions, including regions from different chromosomes [24].

As a step towards determining whether *VDR* methylation may contribute to differential disease susceptibility, we first sought to identify variation in DNA methylation between individuals of African ancestry from Ibadan Nigeria (Yoruba, YRI, part of the International HapMap project) and individuals of European ancestry (Caucasian, CAU). Given the connection between epigenetic mechanisms, genetic variations and environmental factors, YRI and CAU were chosen for analysis based on their diverse *VDR* genetic characteristics and differential exposure to environmental factors (geographic location, lifestyle and nutritional status), which may have the potential to influence methylation. Secondly, we investigated variation in methylation between pulmonary TB cases and controls in the Venda population, Limpopo Province, South Africa. Lastly, we aimed to identify interaction

of population or TB with genetic variants at key CpGs. Following *in silico* identification of *VDR* CGIs and a pilot screen for variable methylation, SNPs and differences in site-specific and regional methylation levels in a CGI at the 3' end of the *VDR* were identified with pyrosequencing. Results indicated a large degree of site-specific methylation variation between populations and some between TB cases and controls while the *TaqI* SNP showed an interaction with methylation, providing further support for the consideration of both *VDR* genetics and epigenetics in epidemiological studies.

2. Materials and Methods

2.1. Bioinformatics

CGIs of the *VDR* (chromosome 12q13.11, negative strand) were identified *in silico* according to epigenome criteria [25] using coordinates chr12:46,511,579 – 46,633,637 (UCSC Golden Path Genome Browser Database, <http://genome.ucsc.edu/>, Mar. 2006 assembly: hg18 / NCBI36). The transcription start site (TSS) of transcript GenBank accession no. **AK024830** at exon 9 was verified with Genomatix ELDorado/Gene2Promoter Release 4.9 (January 2010). SNPs were located using dbSNP, Build 130 (<http://www.ncbi.nlm.nih.gov/SNP/> [26]).

2.2. Subjects

Methylation status was studied considering ethnicity and TB status. For the ethnicity study, DNA from Epstein-Barr Virus (EBV) transformed B-lymphocytes of unrelated individuals (30 CAU from North America, from the HD50CAU panel and 30 YRI from the International HapMap Project) was purchased from Coriell Cell Repositories (Camden, NJ, USA). Considering TB status, DNA samples from 32 pulmonary TB cases and 29 healthy TB contacts (controls) were randomly selected from a study population previously collected in the Venda, Limpopo Province, South Africa [12]. Briefly, pulmonary TB was confirmed by sputum smears, positive for acid-fast bacilli. After informed consent was obtained from patients or their parents or guardians, venous blood was collected. DNA was extracted by use of the Nucleon BACC2 DNA extraction kit (Nucleon Bioscience) and standard phenol-chloroform procedures. Ethical approval was provided by the Medical Research Council (MRC) of South Africa, the Northern Province Department of Health and Welfare and the Ethics Committee of the Faculty of Science, University of Johannesburg. Equal numbers of males and females were used for each population, except for Venda controls (12 males and 17 females). The median age for CAU was 32 (range: 22-44), for Venda controls 34 (range: 21-62) and for Venda TB cases 38 (range: 18-62). The ages for YRI individuals were unknown but they were parents in trios collected for the International HapMap Project. It is well known that methylation is tissue/cell type-specific. Therefore, DNA extracted from EBV-transformed B-lymphocytes (CAU and YRI DNA) was analysed separately from whole blood DNA (Venda DNA) in subsequent methylation analyses to avoid detecting significant differences due to cell type differences.

2.3. Methylation analysis

An estimate of methylation across CGI 1065, CGI 1062 and CGI 1060 in the *VDR* was obtained using the modified Enzymatic Regional Methylation Assay (ERMA; [27]). The methylation percentage at each of 17 CpGs in CGI 1060 (288 bp) was quantified using bisulphite pyrosequencing (EpigenDx Inc. Worcester, MA). Coordinates for the region covered at CGI 1060 in this analysis was chr12:46,524,819 – 46,525,106 (UCSC release hg18, Mar. 2006).

2.4. Genotype analysis

Seven SNPs in CGI 1060 (rs7975232 [*ApaI*], rs4987032, rs731236 [*TaqI*], rs11574115, rs34189316, rs35435255 and rs2229829) were genotyped in the CAU and Venda population through pyrosequencing at the Division of Human Genetics, Medical Faculty, University of the Witwatersrand, National Health Laboratory Services (NHLS). Except for *ApaI* and *TaqI*, five SNPs were monomorphic and were excluded from further analysis. The *ApaI* and *TaqI* genotyping data for the YRI population was obtained from the International HapMap Project website, HapMap Data Rel 24/phaseII Nov08 built upon NCBI B36 assembly, dbSNP b126.

2.5. Statistical analysis

Statistical analysis was performed using PASW Statistics 18, Release 18.0.0 (Jul 30, 2009; SPSS Inc., Chicago, Illinois). The Mann-Whitney U test was used to test the significance of methylation differences between ethnic groups and between controls and TB cases, without Bonferroni correction, since sites compared were unique and there were only two groups per site. To study interaction of ethnicity or TB status with *TaqI* genotype regarding methylation, Kruskal-Wallis was used followed by Mann-Whitney U to test significance of pair-wise differences, applying Bonferroni correction. *ApaI* showed no significant effect on methylation (results not shown) and was therefore excluded from similar statistical analyses.

3. Results

3.1. A polymorphic CpG island at the 3' end of *VDR* is variably methylated

We first defined CGIs at *VDR* and their relationship with underlying genetic diversity. *In silico* analysis facilitated the mapping of six CGIs across *VDR*. Three were located at experimentally verified *VDR* promoters: the tissue-specific promoter at exon 1f (CGI 1065); the primary promoter at exons 1a-1d (CGI 1062); and a promoter at exon 9 (CGI 1060; Fig. 1A). To identify the most variably methylated CGI located at promoters, methylation densities of CGIs were analysed using ERMA for seven DNA samples from primary whole blood cells (a male and female from the Venda population, comprising a control and a TB case for each), EBV-transformed B lymphocytes (a CAU male and female) and the monocytic cell line, U937. CGI 1060 was the most variably methylated (mean methylation density 55.3%, CV = 52%), compared to CGI 1065 (mean methylation density 35.3%, CV = 48%) and 1062 (which was largely unmethylated).

Bioinformatic annotation of CGI 1060 (Fig. 1B) revealed 17 CpGs and 7 SNPs. The latter comprised of rs7975232 (*ApaI*), rs4987032, rs731236 (*TaqI*), rs11574115, rs34189316, rs35435255 and rs2229829. The monomorphic status of 5 SNPs excluding *ApaI* and *TaqI*, is probably a result of the small, unrepresentative sample size. Yet, analysis of dbSNP data showed either monomorphism (rs4987032, rs2229829 and rs11574115) or slight variation (rs34189316 and rs35435255) in larger Caucasian and/or African American groups. The existence of CpGs 5 and 6 was found to depend on the specific allele present for rs4987032 (A/G) and rs731236 (*TaqI*, C/T) respectively. CGI 1060 was positioned 454 bp upstream of the TSS for the non-coding transcript [AK024830](#), of which the promoter, GXP_168256, traverses CpGs 16 and 17. CGI 1060 was selected for further analysis based on its methylation variability and the presence of TB-associated SNPs, *ApaI* and *TaqI*.

3.2. Methylation varies based on ethnicity and disease status

We compared methylation levels of CGI 1060 between populations of European (CAU) and African (YRI) ancestry to study ethnicity-related methylation differences (Fig. 2). Site-specific and regional methylation differences were considered as both may alter gene expression. Mean, site-specific methylation was significantly higher in YRI at CpGs 2, 5, 7,

8, 10, 13 and 14 and significantly lower at CpGs 4 and 6 as compared to CAU. Suggestive differences in regional methylation between YRI and CAU were observed but not found to be statistically significant ($P = 0.07$). CpG 6 coincided with *TaqI* and showed the least methylation of all CpGs in CAU and YRI. The *TaqI* 'C' allele, when present, was always methylated, causing CpG 6 to be methylated distinctly according to genotype: 0% (TT), 50% (TC) or 100% (CC). The interquartile ranges (IQRs) for both populations at CpG 6 was similar but the median and mean were higher in CAU (mean difference, $P = 0.05$). The IQR for all other CpGs were narrower for YRI than for CAU (Fig. 2). Methylation levels of CGI 1060 were then compared between TB cases and healthy TB contacts from a Venda population to study TB-related methylation variability (Fig. 3). CpG 3 showed higher methylation in controls than TB cases ($P = 0.04$), while CpGs 7 and 8 showed higher methylation in TB cases than controls ($P = 0.01$). No difference in regional methylation between TB cases and controls was observed (Fig. 3). Distribution ranges for site-specific and regional methylation were narrower for Venda DNA, extracted from whole blood, than for DNA extracted from EBV-transformed B lymphocytes in the case of YRI and CAU (for example compare boxplot ranges between Fig. 2 and 3).

3.3. Methylation is associated with *TaqI* genotype at specific sites and globally

To investigate whether methylation level was influenced by *TaqI* or by factors independent of *TaqI*, the Kruskal-Wallis test together with Mann-Whitney U and Bonferroni correction was used to screen for CpGs showing interaction (or lack thereof) of *TaqI* with ethnicity (Fig. 4) and *TaqI* with TB status (Fig. 5). Interaction of *TaqI* with ethnicity is evident when methylation in TT individuals differs significantly from TC/CC individuals in either CAU or YRI but not in both ethnic groups. Similarly, interaction of *TaqI* with TB status is evident when methylation in TT individuals differs significantly from TC/CC individuals in either TB cases or controls but not in both TB status groups. Interaction of *TaqI* with ethnicity or TB status is also evident when methylation is significantly different between TT individuals of both ethnic or TB status groups or between TC/CC individuals of both ethnic or TB status groups but not in both genotype groups. Lastly, an influence of ethnicity or TB status on methylation, independent of *TaqI* genotype, is evident when methylation of one ethnic or TB status group is significantly different from the other in both *TaqI* genotype groups (i.e. irrespective of *TaqI* genotype).

Fig. 4 shows interaction of *TaqI* genotype with ethnicity at CpGs 4, 14, and at regional methylation. CpG 8 showed an independent ethnic effect without interaction with *TaqI* genotype. CpGs 10 and 13, previously shown to differ between CAU and YRI (Fig. 2), showed no interaction of *TaqI* genotype with ethnicity, implicating an ethnic effect independent of *TaqI* genotype.

Fig. 5 shows interaction of *TaqI* genotype with TB status at CpGs 16 and 17. None of the three sites shown to differ between TB cases and controls (Fig. 3) showed significant interaction of *TaqI* genotype with TB status, implicating a TB effect independent of *TaqI* genotype for these three sites. At regional level, both TB cases and controls showed significantly higher methylation in TC/CC individuals compared to TT individuals (Fig. 5), a similar observation noted in YRI but not in CAU (Fig. 4).

In Fig. 4 and 5, CpG 6 showed significantly higher methylation of TC/CC individuals compared to TT individuals in both ethnic and TB status groups indicating an independent *TaqI* genotype effect without interaction with ethnicity (Fig. 4) or TB status (Fig. 5).

4. Discussion

This study identified several site-specific methylation differences at the 3' end of *VDR* in CGI 1060 between a Caucasian (CAU) and an African (YRI) population. Moreover, three site-specific methylation differences were observed between pulmonary TB cases and healthy contacts in a Venda population from South Africa. The disease associated *TaqI* SNP (rs731236) was found to be located in a CpG imposing a direct *cis* effect on site-specific and regional methylation, particularly in African populations (YRI and Venda). These observed methylation differences in the *VDR* with respect to ethnicity and TB status, and the modulation by *TaqI*, emphasizes the importance to consider both genetic and epigenetic variation in disease-association. Besides their potential as epigenetic markers for disease, MVPs studied in conjunction with SNPs may shed light on inconsistent disease association reports for *VDR* SNPs.

CGI 1060 was the most variably methylated CGI identified. Furthermore, it harboured the SNPs *Apal* in intron 8 and *TaqI* in exon 9 (Fig. 1B), both previously implicated in TB susceptibility [10-14]. *TaqI* and rs4987032 (immediately upstream of *TaqI* in exon 9) could directly affect methylation due to their respective location in CpGs 5 and 6 of CGI 1060. The lack of methylation of CGI 1062 was in line with its location at the primary promoter of the *VDR* at exons 1a-1d (Fig. 1A). The variable methylation of CGI 1060 may influence gene expression through post-transcriptional gene regulation by **AK024830**, a transcript with characteristics of an intronic long non-coding RNA [28].

Considering ethnicity, YRI showed significantly higher methylation than CAU at several CpG sites in CGI 1060, while our data is suggestive of regional methylation differences. Selective site-specific methylation differences suggest that factors depending on DNA sequence contribute to the observed methylation variations [29]. At CpG 6, the significantly higher methylation in CAU compared to YRI reflects the higher *TaqI* 'C' allele frequency in CAU (0.383) compared to YRI (0.214), that was significantly different in the current study population ($\chi^2 = 3.929$, $df = 1$, $P = 0.047$). Thus *TaqI* genotype imposed a direct effect on site-specific methylation, which may also influence regional methylation differences. Not only may methylation be influenced by genotype but also by environmental factors. For example, epigenetic variations among populations have been shown to be influenced by differences in geographic location, lifestyles and nutritional status [7]. Epigenetic mechanisms are influenced by heavy metal exposure [30], the latter often being a consequence of geographic location. Individuals from Nigeria, the country of origin for YRI, are more exposed to heavy metals than Europeans and Americans [31,32]. Cells, from which DNA was extracted for the CAU population, was immortalised 6-32 years before YRI cells (personal communication with Coriell). Though cryopreservation of animal cells is not thought to influence methylation, repeated growth and serial passaging of cells may [33] and the role of this phenomenon in observed differences can not be excluded. Our results suggest that ethnically diverse populations, at least when comparing YRI and CAU lymphoblastoid cell lines, may not only differ genetically in the *VDR* but also at an epigenetic level. The effect of methylation variation at CGI 1060 is unknown but it may contribute to *VDR*-related disease predisposition. For example, Das et al [7] proposed an epigenetic predisposition to prostate cancer in black men, after observing a higher *TMS1* methylation prevalence in healthy black men than healthy white European men.

Considering TB status, three site-specific methylation differences (CpGs 3, 7 and 8) were observed (Fig. 3). Whether the methylation differences at CpGs 3, 7 and 8 contribute to TB predisposition or are merely the result of infection with *M. tuberculosis* and treatment of TB is unknown. Chan et al [34] showed that *E-cad* promoter methylation is initiated by *Helicobacter pylori* infection, leading to gastric carcinogenesis. Kovalenko et al [35]

showed that treatment with the anti-tuberculosis drug pyrazinamide leads to methylation alterations of LINE-1 (long interspersed nucleotide elements), *GSTP* (glutathione-S-transferase) and *p16(INK4A)* genes in rat liver. If methylation is a cause rather than an effect of TB, CpGs 3, 7 and 8 may serve as epigenetic markers (MVPs) with the potential to statistically discriminate predisposed individuals from those resistant to TB. Site-specific methylation differences between TB and controls were detected despite the small number of individuals used in the current study, suggesting that epigenetic variants may be stronger markers of disease than genetic variants that require large cohorts.

The wider distribution range of methylation levels observed in CAU compared to YRI individuals may reflect differences in the times of immortalisation and subsequent culturing, as mentioned previously or it may be due to variation in exposure to methylation agents or inherent factors affecting methylation. The higher methylation levels and narrower distribution ranges observed in the Venda compared to CAU and YRI may simply be explained by differences in DNA source as DNA for CAU and YRI was extracted from cultured EBV-transformed B lymphocytes and that for Venda from whole blood.

Variable methylation may be regulated by genetic factors through *cis* or *trans* mechanisms. Further study of the interaction of *TaqI* as a *cis* genetic factor with either ethnicity or TB status revealed *TaqI*-independent factors involved in methylation variability at: (i) the three CpGs differentially methylated between CAU and YRI (CpGs 8, 10 and 13; Fig. 2 and 4) and at (ii) all three CpGs differentially methylated between TB cases and controls (CpGs 3, 7 and 8; Fig. 3 and 5). Besides its direct effect on CpG 6, *TaqI* genotype showed an interaction with ethnicity at CpG 4 and at regional level (Fig. 4). The ethnicity-based effect of *TaqI* on regional methylation (Fig. 4) was also seen in the Venda (TB and controls; Fig. 5) and cannot be explained by the differences in frequencies of TC/CC individuals in YRI and Venda compared to CAU: CAU TT = 0.367, TC/CC = 0.633; YRI TT = 0.607, TC/CC = 0.393; Venda TT = 0.617, TC/CC = 0.383. The trend of TC/CC individuals in both Caucasians and Africans having higher regional methylation at CGI 1060 of the *VDR* is significant only in Africans, likely due to environmental enhancement of TC/CC methylation.

The interaction between *TaqI* and TB status at CpGs 16 and 17 (Fig. 5) supports the analysis of both *VDR* genotype and *VDR* epigenotype in epidemiological studies addressing TB susceptibility. From a genetic perspective, it is still uncertain whether *TaqI* or a linked marker is associated with TB. For example, the tt genotype, which is often associated with lower bone mineral density, was found less frequently in patients with pulmonary TB in The Gambia [10] but was not confirmed in a subsequent West African study that supported *VDR* haplotype association [11], similar to a South African study [12]. A study in India showed the tt genotype to be associated with TB [14]. Wilkinson et al [36] suggested gene-environment interaction, associating TB with TT/Tt genotype in combination with 25-hydroxycholecalciferol deficiency. Yet others found no association [37]. The interaction observed at CpGs 16 and 17 suggests that *TaqI* does not function alone but together with site-specific *VDR* methylation to influence TB susceptibility. Olesen et al [38] found a significant interaction between *ApaI* and ethnicity ($P = 0.003$), analysing TB risk in West Africans. This involvement of ethnicity in the association of disease risk with *ApaI*, located on the 'shore' of CGI 1060, justifies further study of methylation and 3' haplotypes of the *VDR* in TB susceptibility. Using *TaqI*-dependent MVPs in addition to SNPs in genetic association studies may provide a more comprehensive understanding of the molecular basis of disease association and could shed light on dissimilar *VDR* SNP associations found between populations.

In conclusion, our results show differential methylation of a 3' *VDR* CGI between a population of African and European ancestry; individuals with African ancestry showing higher methylation at most sites. In the Venda from South Africa additional methylation differences were related to TB status. Interaction between the disease-associated SNP *TaqI*, population and TB with respect to *VDR* methylation, motivates the consideration of both genotype and epigenotype in association studies among different populations. The basis of variation in methylation and its impact on *VDR* gene expression and disease susceptibility are yet to be determined.

Acknowledgments

This work was supported by the Medical Research Council (MRC), National Research Foundation (NRF) and the South African Coal, Oil and Gas Corporation (SASOL). We would like to thank the Wellcome Trust Centre for Human Genetics (WTCHG) Genomics group (Oxford, UK) for valuable advice and Jaclyn Smith (Statistical Consultation Services, University of Johannesburg) for assistance with statistical analyses.

Abbreviations

CAU	Caucasian
CGI	CpG island
EBV	Epstein-Barr virus
ERMA	enzymatic regional methylation assay
IQR	interquartile range
MVPs	methylation variable positions
TSS	transcription start site
TB	tuberculosis
VDR	Vitamin D receptor
YRI	Yoruba

References

- [1]. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution. *Genome Res.* 2010 10.1101/gr.107920.110.
- [2]. Kato S. The function of vitamin D receptor in vitamin D action. *J Biochem.* 2000; 127(5):717–22. [PubMed: 10788778]
- [3]. Hayes CE, Nashold FE, Spach KM, Pedersen LB. The immunological functions of the vitamin D endocrine system. *Cell Mol Biol (Noisy-legrand).* 2003; 49:277–300. [PubMed: 12887108]
- [4]. Valdivielso JM, Fernandez E. Vitamin D receptor polymorphisms and diseases. *Clin Chim Acta.* 2006; 371:1–12. [PubMed: 16563362]
- [5]. Zmuda JM, Cauley JA, Ferrell RE. Molecular epidemiology of vitamin D receptor gene variants. *Epidemiol Rev.* 2000; 22:203–17. [PubMed: 11218372]
- [6]. Ramos RG, Olden K. Gene-Environment Interactions in the Development of Complex Disease Phenotypes. *Int J Environ Res Public Health.* 2008; 5(1):4–11. [PubMed: 18441400]
- [7]. Das PM, Ramachandran K, VanWert J, Ferdinand L, Gopisetty G, Reis IM, et al. Methylation mediated silencing of TMS1/ASC gene in prostate cancer. *Mol Cancer.* 2006; 5:28. [PubMed: 16848908]
- [8]. Singal R, Ginder GD. DNA Methylation. *Blood.* 1999; 93(12):4059–70. [PubMed: 10361102]
- [9]. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, et al. Toll-like receptor triggering of a Vitamin-D Human Antimicrobial Response. *Science.* 2006; 311:1770–3. [PubMed: 16497887]

- [10]. Bellamy R, Ruwende C, Corrah T, McAdam KPWJ, Thursz M, Whittle HC, Hill AVS. Tuberculosis and Chronic Hepatitis B Virus Infection in Africans and Variation in the Vitamin D Receptor Gene. *J Infect Dis.* 1999; 179:721–4. [PubMed: 9952386]
- [11]. Bornman L, Campbell SJ, Fielding K, Bah B, Sillah J, Gustafson P, et al. Vitamin D Receptor Polymorphisms and Susceptibility to Tuberculosis in West Africa: A Case-Control and Family Study. *J Infect Dis.* 2004; 190:1631–41. [PubMed: 15478069]
- [12]. Lombard Z, Dalton D-L, Venter PA, Williams RC, Bornman L. Association of HLA-DR, -DQ, and Vitamin D Receptor Alleles and Haplotypes with Tuberculosis in the Venda of South Africa. *Hum Immunol.* 2006; 67:643–54. [PubMed: 16916662]
- [13]. Selvaraj P, Alagarasu K, Harishankar M, Vidyarani M, Narayanan PR. Regulatory region polymorphisms of vitamin D receptor gene in pulmonary tuberculosis patients and normal healthy subjects of south India. *Int J Immunogenet.* 2008b; 35:251–54. [PubMed: 18397302]
- [14]. Selvaraj P, Narayanan PR, Reetha AM. Association of vitamin D receptor genotypes with the susceptibility to pulmonary tuberculosis in female patients & resistance in female contacts. *Indian J Med Res.* 2000; 111:172–9. [PubMed: 10943070]
- [15]. Marik R, Fackler M, Gabrielson E, Zeiger MA, Sukumar S, Stearns V, et al. DNA methylation-related vitamin D receptor insensitivity in breast cancer. *Cancer Biol Ther.* 2010; 10(1)
- [16]. Smirnoff P, Liel Y, Gnainsky J, Shany S, Schwartz B. The protective effect of estrogen against chemically induced murine colon carcinogenesis is associated with decreased CpG island methylation and increased mRNA and protein expression of the colonic vitamin D receptor. *Oncol Res.* 1999; 11(6):255–64. [PubMed: 10691027]
- [17]. Barros SP, Offenbacher S. Epigenetics: Connecting Environment and Genotype to Phenotype and Disease. *J Dent Res.* 2009; 88(5):400–8. [PubMed: 19493882]
- [18]. Murrell A, Rakyan VK, Beck S. From genome to epigenome. *Hum Mol Genet.* 2005; 14(1):R3–R10. [PubMed: 15809270]
- [19]. Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, et al. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2004; 2(12):e405. [PubMed: 15550986]
- [20]. Gao L, Tao Y, Zhang L, Jin Q. Vitamin D receptor genetic polymorphisms and tuberculosis: updated systematic review and meta-analysis. *Int J Tuberc Lung Dis.* 2010; 14(1):15–23. [PubMed: 20003690]
- [21]. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* 2004; 20(8):350–8. [PubMed: 15262407]
- [22]. Murrell A, Heeson S, Cooper WN, Douglas E, Apostolidou S, Moore GE, et al. An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum Mol Genet.* 2004; 13(2):247–55. [PubMed: 14645199]
- [23]. Ronneberg JA, Tost J, Solvang HK, Alnaes GI, Johansen FE, Brendeford EM, et al. GSTP1 promoter haplotypes affect DNA methylation levels and promoter activity in breast carcinomas. *Cancer Res.* 2008; 68(14):5562–71. [PubMed: 18632608]
- [24]. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet.* 2010; 86(3):411–9. [PubMed: 20215007]
- [25]. Bock C, Walter J, Paulsen M, Lengauer T. CpG island mapping by epigenome prediction. *PLoS Comput Biol.* 2007; 3(6):e110. doi:10.1371/journal.pcbi.0030110. [PubMed: 17559301]
- [26]. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001; 29(1):308–11. [PubMed: 11125122]
- [27]. Zhang Z, Chen C, Manev H. Enzymatic Regional Methylation Assay for Determination of CpG Methylation Density. *Anal Chem.* 2004; 76:6829–32. [PubMed: 15538812]
- [28]. Louro R, Smirnova AS, Verjovski-Almeida S. Long intronic noncoding RNA transcription: Expression noise or expression choice? *Genomics.* 2009; 93:291–8. [PubMed: 19071207]
- [29]. Handa V, Jeltsch A. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J Mol Biol.* 2005; 348:1103–12. [PubMed: 15854647]

- [30]. Arita A, Costa M. Epigenetics in metal carcinogenesis: Nickel, Arsenic, Chromium and Cadmium. *Metallomics*. 2009; 1:222–28. [PubMed: 20461219]
- [31]. Maduabuchi JM, Adigba EO, Nzegwu CN, Oragwu CI, Okonkwo IP, Orisakwe OE. Arsenic and chromium in canned and non-canned beverages in Nigeria: a potential public health concern. *Int J Environ Res Public Health*. 2007; 4(1):28–33. [PubMed: 17431312]
- [32]. Maduabuchi JM, Nzegwu CN, Adigba EO, Oragwu CI, Agbo FN, Agbata CA, et al. Iron, Manganese and Nickel Exposure from Beverages in Nigeria: A Public Health Concern? *J Health Sci*. 2008; 54(2):335–8. [PubMed: 18397302]
- [33]. Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, et al. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell*. 2010; 9(1):54–63. [PubMed: 19895632]
- [34]. Chan AO-O, Lam S-K, Wong BC-Y, Wong W-M, Yuen M-F, Yeung Y-H, et al. Promoter methylation of E-cadherin gene in gastric mucosa associated with *Helicobacter pylori* infection and in gastric cancer. *Gut*. 2003; 52:502–6. [PubMed: 12631658]
- [35]. Kovalenko VM, Bagnyukova TV, Sergienko OV, Bondarenko LB, Shayakhmetova GM, Matvienko AV, et al. Epigenetic changes in the rat livers induced by pyrazinamide treatment. *Toxicol Appl Pharmacol*. 2007; 225(3):293–9. [PubMed: 17920094]
- [36]. Wilkinson RJ, Llewelyn M, Toossi Z, Patel P, Pasvol G, Lalvani A, et al. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London : a case-control study. *Lancet*. 2000; 355:618–21. [PubMed: 10696983]
- [37]. Babb C, van der Merwe L, Beyers N, Pfeiffer C, Walzl G, Duncan K, van Helden P, Hoal EG. Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. *Tuberculosis*. 2007; 87(4):295–302. [PubMed: 17449323]
- [38]. Olesen R, Wejse C, Velez DR, Bisseye C, Sodemann M, Aaby P, et al. DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. *Genes Immun*. 2007; 8(6):456–67. [PubMed: 17611589]
- [39]. Crofts LA, Hancock MS, Morrison NA, Eisman JA. Multiple promoters direct the tissue-specific expression of novel N-terminal variant human vitamin D Receptor gene transcripts. *Proc Natl Acad Sci USA*. 1998; 95:10529–34. [PubMed: 9724737]
- [40]. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet*. 2004; 36(1):40–5. [PubMed: 14702039]

Web References

- 41. [date last accessed: 06/2010] <http://genome.ucsc.edu/http://genome.ucsc.edu/>
- 42. [date last accessed: 11/2009] <http://www.ncbi.nlm.nih.gov/SNP/http://www.ncbi.nlm.nih.gov/SNP/>

Fig. 1. A diagram of the human vitamin D receptor gene (*VDR*) showing putative CpG Islands (CGI) and SNPs previously associated with TB (A) and annotated sequence of CGI 1060 at the 3' end of the *VDR* (B)

A) The *VDR* covers at least 105 kb and is located on the negative strand of chromosome 12 (q13.11). Open bars show exons 1f-1c in the regulatory region; largely non-coding. Black bars show main coding exons 2-9 with the 3' UTR. Curved arrows indicate transcription start sites (TSSs). The position of CGIs are shown above the gene diagram and that of SNPs previously implicated in TB susceptibility below the gene diagram: Cdx2 [13], A1012G [13], *FokI*, *BsmI*, *ApaI*, *TaqI* [10-12,14]. *ApaI* and *TaqI* are located in CGI 1060. The diagram is not to scale and was adapted from Crofts et al [39]. **B)** CGI 1060 (bold font) overlaps with exon 9 (underlined) and promoter GXP_168256 (dashed line) of a non-coding transcript **AK024830**. The curved arrow indicates the transcription start site (TSS), 'T', for transcript **AK024830** (1687 bp), first reported by Ota et al [40]. CGI 1060 harbours seventeen CpGs (1-17) and seven SNPs indicated with IUB redundancy codes (R: A/G, Y:

C/T, M: A/C, K: G/T and S: G/C) and reference SNP numbers below. The region shown is chr12:46,524,238 – 46,525,237, negative strand (UCSC release hg18, Mar. 2006).

Fig. 2. Site-specific methylation of CpG Island (CGI) 1060 differed between a Caucasian and Yoruba population

Boxplots show location and dispersion of Caucasian (CAU, $n = 30$, white) and Yoruba (YRI, $n = 28$, grey) methylation levels. The line in the box shows the median and the height of the box the interquartile range (IQR, 50% of the cases; from the 25th to 75th percentile). The whiskers extend to 1.5 IQR's or, if no case has a value in that range, to the minimum or maximum values. Outliers ($\circ/\bullet$) are values between 1.5 IQR's and 3 IQR's from the end of a box, while values more than 3 IQR's from the end of a box are defined as extreme (asterisks below boxplots). Approximately 95% of the data are expected to lie between the whiskers if the data are distributed normally. Significant differences in methylation between CAU and YRI are shown as brackets with asterisks above: * P -value 0.05 but > 0.01 , ** P -value 0.01 but > 0.001 , *** P -value 0.001. Methylation was higher in YRI compared to CAU at CpGs 2, 5, 7, 8, 10, 13 and 14, while higher in CAU compared to YRI at CpGs 4 and 6. The *TaqI* (C/T) SNP occupied the first nucleotide of CpG 6. Regional methylation (far right), the average methylation across all 17 CpGs in CGI 1060, did not differ between CAU and YRI ($P = 0.07$). EBV-transformed B lymphocytes served as DNA source. Methylation levels were determined with bisulphite pyrosequencing and significance of differences between populations, with Mann-Whitney U.

Fig. 3. Methylation differed between TB cases and controls at CpGs 3, 7 and 8 in CGI 1060

Boxplots show location and dispersion of Venda control (n = 28, white) and Venda TB (n = 32, grey) methylation levels. Boxplot detail is outlined in the legend of Fig. 2. Significant differences in methylation between TB cases and controls are shown as brackets with asterisks above: * P -value 0.05 but > 0.01 , ** P -value 0.01 but > 0.001 . Methylation was higher in controls compared to TB cases at CpG 3 and higher in TB cases compared to controls at CpGs 7 and 8. Regional methylation did not differ between TB cases and controls ($P = 1.00$). Whole blood served as DNA source. Methylation levels were determined with bisulphite pyrosequencing and significance of differences between TB cases and controls, using Mann-Whitney U.

Fig. 4. Interaction between ethnicity and *TaqI* genotype was observed in methylation levels at CpGs 4, 14 and at regional level

Methylation levels were determined with bisulphite pyrosequencing and interaction between factors analysed using Kruskal-Wallis followed by Mann-Whitney U tests, with Bonferroni correction. Bars illustrate CAU TT (n = 11, white), CAU TC/CC (n = 19, white striped), Yoruba (YRI) TT (n = 17, grey) and YRI TC/CC (n = 11, grey striped). Error bars show $\pm$ SD. CpGs showing significant differences, withstanding Bonferroni correction, are indicated; * $P = 0.008$ and ** $P = 0.001$. Note that only significant differences indicating interaction of *TaqI* genotype with ethnicity and independent effects of ethnicity and *TaqI* genotype are shown. Methylation at CpGs 4, 14 and at regional level showed differences reflecting interaction of *TaqI* genotype with ethnicity. CpG 6 showed significant differences due to *TaqI* genotype, independent of ethnicity. CpG 8 showed significant differences due to ethnicity, independent of *TaqI* genotype.

Fig. 5. Interaction between TB status and *TaqI* genotype was observed in methylation levels at CpGs 16 and 17

Methylation levels were determined with bisulphite pyrosequencing and interaction between factors analysed using Kruskal-Wallis followed by Mann-Whitney U tests, with Bonferroni correction. Bars illustrate control TT (n = 17, white), control TC/CC (n = 11, white striped), TB TT (n = 20, grey) and TB TC/CC (n = 12, grey striped). Error bars show $\pm$ SD. CpGs showing significant differences, withstanding Bonferroni correction, are indicated; ** P 0.001. Note that only significant differences indicating interaction of *TaqI* genotype with TB status and independent effects of TB status and *TaqI* genotype are shown. Methylation at CpGs 16 and 17 showed differences reflecting interaction of *TaqI* genotype with TB status. Significant effects of *TaqI* genotype, independent of TB status, were evident at CpG 6 and at regional level.