Epigenetic analysis of Paget's disease of bone identifies differentially methylated loci that predict disease status

Impact

PDB associated differences in DNA methylation are reproducible and reflect key environmental modulators of bone homeostasis including viral processes, vitamin D metabolism as well as mechanical sheer load.

Author information

Ilhame Diboun[1], Sachin Wani[2], Stuart H Ralston[2], Omar M E Albagha[1,2]

1. Division of Genomic and Translational Biomedicine, College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

2. Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Correspondence to:
Omar Albagha, PhD
Centre for Genomic and Experimental Medicine,
MRC Institute of Genetics and Molecular Medicine,
University of Edinburgh,
Edinburgh EH4 2XU
UK
Tel: +44 131 6518727
Email: omar.albagha@ed.ac.uk

Competing interests

Prof S H Ralston has received research funding from Amgen, Eli Lilly, Novartis, and Pfizer unrelated to the submitted work. The other authors have no conflicts of interest to declare

Author Contributions

Study conception: OMEA; study supervision: OMEA; data collection: SHR; genotyping: SW; data analysis: ID and OMEA; drafting the manuscript: ID; revising the manuscript: ID, OMEA, and SHR; all authors contributed to critically review the manuscript and approved its final version.

Funding

This work was funded by a consolidator grant from the European Research Council to OMEA (311723-GENEPAD) and in part by an advanced investigator grant from the European Research Council to SHR (787270 - Paget-Advance). The PRISM trial was supported by grants from the Arthritis Research Campaign (13627) and the Paget's Association.

Datasets

Raw and processed methylation data generated in this study can be found at GEO under the accession GSE163970.

Ethics Statement

The study was approved by the UK Multicenter Research Ethics Committee for Scotland (MREC01/0/53) and NHS Lothian, Edinburgh (08/S1104/8) ethics review committees. All participants provided written informed consent.

Abstract

Paget's Disease of Bone (PDB) is characterized by focal increases in disorganized bone remodeling. This study aims to characterize PDB associated changes in DNA methylation profiles in patients' blood. Meta-analysis of data from the discovery and replication set, comprising of 116 PDB cases and 130 controls, revealed significant differences in DNA methylation at 14 CpG sites, 4 CpG islands, and 6 gene-body regions. These loci, including two characterized as functional through eQTM analysis, were associated with functions related to osteoclast differentiation, mechanical loading, immune function, and viral infection. A multivariate classifier based on discovery samples was found to discriminate PDB cases and controls from the replication with a sensitivity of 0.84, specificity of 0.81, and an area under curve of 92.8%. In conclusion, this study has shown for the first time that epigenetic factors contribute to the pathogenesis of PDB and may offer diagnostic markers for prediction of the disease.

Introduction

Paget's disease of bone (PDB) is characterized by increased but disorganized bone remodeling, which causes affected bones to enlarge, become weak and deform. The axial skeleton is predominantly involved, and commonly affected sites include the skull, spine, pelvis, femora and tibia. Paget's disease is clinically silent until it has reached an advanced stage at which point irreversible damage to the skeleton has occurred (1). Bisphosphonates are an effective treatment (2) and can often improve bone pain but have a limited impact on other clinical outcomes in patients with advanced disease (3, 4). On a cellular level, PDB is characterized by increased osteoclast activity and biopsies from affected bone lesions exhibit increase in the number and size of osteoclasts.

Genetic factors play an important role in classical PDB and in monogenic PDB-like syndromes (5, 6). Mutations in *SQSTM1* are the most common cause of PDB but other susceptibility genes and loci have been identified through genome wide association studies (7-9). These include genes that play an important role in osteoclast differentiation such as *CSF1*, *TNFRSF11A* and *DCSTAMP*. Additionally, an expression quantitative trait locus (eQTL) in *OPTN* is associated with increased susceptibility to PDB (10). Functional analysis using mouse models showed that OPTN is a negative regulator of osteoclast differentiation and mice with loss of *OPTN* function develop PDB-like bone lesions with increasing age (10, 11).

Environmental factors also play a role, as evidenced by the fact that the disease is focal in nature and its incidence and severity has diminished in recent years (12). Several environmental triggers have been suggested including persistent viral infection, repetitive mechanical loading of the skeleton, low dietary calcium intake, environment pollutants and vitamin D deficiency (6).

The possible role of persistent viral infection with measles and distemper has been studied experimentally. for example, expression of the measles virus nucleocapsid protein in osteoclasts was found to trigger PDB-like phenotype in mice (13, 14). However, clinical studies which have sought to detect evidence of viral proteins and nucleic acids in humans with PDB have yielded conflicting results (2).

Accumulating evidence suggests that environmental and lifestyle factors can influence gene expression and clinical phenotype in various diseases through epigenetic mechanisms such as changes in DNA methylation. To gain insights into the role of epigenetic DNA methylation in PDB, we have conducted genome-wide profiling of DNA methylation in a cohort of 253 PDB patients and 280 controls and evaluated the predictive role of epigenetic markers in differentiating patients with PDB from controls.

Results

Characteristics of study cohort

Table 1 shows descriptive statistics for the study cohort. PDB cases in the discovery set were slightly older and included more males compared to controls but no difference in age or gender distribution was found in the replication set. The number of patients with *SQSTM1* mutations was similar in the discovery and replication set and accounts for approximately 14% of PDB cases.

Differentially methylated Sites (DMS)

Figure 1 shows the study design and summary of differential methylation results. After adjusting for all confounders, differential methylation analysis of the discovery set revealed 419 DMS with FDR < 0.05, 57 of which reached statistical significance (FDR < 0.05) in the replication set (Table S1). Meta-analysis of the DMS from discovery and replication revealed 14 Bonferroni significant DMS ($P < 1.17 \times 10^{-7}$; Table 2). The direction of effect for all replicated DMS was identical in the discovery and replication set and shows hypermethylation in PDB cases compared to controls. A Manhattan plot of the results is shown in Figure 2-A.

Differentially methylated regions (DMR)

Besides analyzing individual sites, our region-based analysis was intended to uncover densely hyper/hypo-methylated regions across the genome in PDB as well as identifying instances where the effect from individual sites is moderate, yet accumulatively significant. We tested natural concentrations of sites within CpG islands but also gene bodies and promoter regions, justified by the fact that promoter methylation often suppresses transcription whilst that from the gene body often stimulates gene expression (Figure 1).

Evaluation of the 25,773 CpG islands on the array, revealed 978 DMR that were significantly differentially methylated (FDR < 0.05) in the discovery set, 111 of which replicated at the same significance level in the replication set (Table S2). Stringent Bonferroni multiple testing correction revealed 4 islands that remained significant following discovery and replication, and these were located near *LTB*, *SKIV2L*, *EBF3* and *CCND1* (Table 3).

Gene body analysis revealed 258 (FDR < 0.05) replicated DMR out of a total of 947 differentially methylated genes initially identified in the discovery set (Table S3). Six gene body DMR reached significance after Bonferroni correction in both the discovery and replication set (Table 3). In the context of promoter regions, evidence for FDR significant association with the disease was equally observed in the discovery and replication set for 27 promoters DMR (Table S4), but none reached significance after Bonferroni correction. Figure 2-B&C show a regional plot for DMR LTB and HSPA13 from island and gene body analysis respectively, highlighting the co-occurrence of multiple differentially methylated sites along each region.

Mapping common regulatory patterns of DNA methylation into functional networks

To gain further insight into the pathology of PDB, we explored common methylation patterns amongst functional keywords identified as significantly over-represented amongst the *Pooled sites.* Figure 3 shows a graphical representation of these functional networks. In addition to bone-related cells, there is a strong presence of immune cells linked to key biological processes including proliferation, differentiation, autophagy and cell death. Furthermore, virus, cytokines, and interferon-gamma were among the over-represented keywords. The process of ubiquitination lies at the center of the graph with the largest number of links in the network.

Diagnostic capacity of differentially methylated markers

In order to determine whether differentially methylated markers might be of diagnostic value, we performed OPLS-DA in the discovery and replication cohorts. The results are summarised in Figure 4. The OPLS-DA procedure was first performed using the combined set of significant DMS and DMR identified from the discovery set (*Pooled sites*; n=2847, refer to methods for further details) and when the classifier was tested on the replication set, it yielded an AUC of 92.8%. To identify sites with the highest predictive ability, we applied the Net Regularization Extension of the Generalized Linear Model approach on the *Pooled* sites which highlighted 95 sites (*Best subset* sites; Table S5), out of the 2847 initial *Pooled* sites, as best discriminatory of PDB cases and controls (Figure 1)*.* The OPLS-DA procedure performed on this *Best subset* resulted in an AUC of 82.5%. A rather superior performance in comparison to similarly trained classifiers based on the DMS (AUC=67%), islands DMR (AUC = 76%), or promoter DMR (AUC = 79%) analyses. On the other hand, the AUC from a classifier restricted to the DMR gene bodies was 92% which is similar to that obtained from the whole *Pooled sites* (AUC 92.8, Figure 3).

Functional enrichment analysis of the 95 *Best subset* was consistent between IPA and GO with many genes annotated to the following broad functional terms: *immune function*; *bone lesions and bone homeostasis,* and *viral processes*. Several identified genes fell into more than one category. Overlaying the IPA knowledge-based repository of molecular interactions identified a handful of functional links between the genes located in the *Best subset* sites, highlighting important functional subnetworks (Figure 5-A). Additionally, we found that the effect size (absolute difference in DNA methylation between controls and PDB cases) was significantly higher for sites from the *Best subset* (mean ± SD; 0.011 ± 0.019) compared to the rest of those in the *Pooled* *sites* (0.007 ± 0.01; P-value = 1.9 x 10^{-3}). The magnitude of effect from each site in the *Best subset,* as calculated by the Elastic-Net Regularization Extension of the Generalized Linear Model, is color-coded in Figure 5-B.

eQTM analysis

eQTM analysis showed that the the Bonferroni significant DMS cg10964367 was associated with the expression level of ARHGEF10 (P = 3.9 x10^{-9}). Additionally, cg26724726 from gene body analysis was associated with the expression of LTB (P= 1.10 x 10^{-5}) and 8 of the *Best subset* sites were associated with the expression of nearby genes (Table S6).

Discussion

The present study is the first to investigate DNA methylation profiles in Paget's disease of bone. DNA methylation profiles from PDB patients were compared to controls and Meta-analysis of discovery and replication revealed 14 genome-wide significant DMS. Many were located within or near genes with functional relevance to the pathogenesis of PDB including bone-related functions, such as osteoclast differentiation, or functions related to environmental triggers associated with PDB such as viral infection and mechanical loading. TNK1 is a tyrosine kinase that has a pivotal role in innate immune responses by regulating the Interferon-stimulated genes downstream of the JAK-STAT pathway (15). It has previously been associated with frontotemporal dementia (16) which can co-exist with Paget's disease (17). MOSC2 is a member of the membrane-bound E3 ubiquitin ligase family that regulates endosome trafficking (18) Less is known about the specific functions of transcription factors NKX6-2 and LBX1 in bone metabolism, but mutations in the latter are associated with Scoliosis HS6ST3 plays a key role in the synthesis of heparan sulfate that potentiates key growth factors including the bone morphogenic protein BMP and Wnt (19). PENK encodes for proenkephalin, the precursor of a range of effector molecules including pain-associated pentapeptide opioids as well as modulators of osteoblast differentiation (20). Interestingly, PENK knockout mice have abnormal bone structure and mineralization (21). MAF was found to promote osteoblast differentiation and heterozygous deletion of MAF in mice results in age-related bone loss associated with accelerated formation of fatty marrow (22). SPATA18 is expressed in a variety of cancers including osteosarcoma and its transcription is induced by p53 (23). TAL1 has been found to regulate osteoclast differentiation through suppression of their fusion mediator DCSTAMP (24). The Zinc finger protein ZIC1 has a role in shear flow mechanotransduction in osteocytes (25). Expression of ZIC1 in human was found to be increased in loaded compared to unloaded bone and the increased expression in loaded bone is associated with reduced methylation in several CpGs in ZIC1 (26). NFYB confers chromatin access to other transcriptional regulators and is known to be involved in transition through cell cycle (27). Finally, the centrosomal ARHGEF10 has a role in the formation of mitotic spindle during mitosis (28).

Our analysis was extended to identify regions with frequent methylation changes in PDB amongst adjacent sites. Genomic regions have traditionally been evaluated in epigenetics studies based on linear combinations of methylation data from residing sites or through meta-analysis of effects/p-values from an initial site-level differential methylation analysis. The novel approach presented in this study is advantegous in to ways: First, our method allows for sites to be hyper or hypo methylated along the same region unlike the linear combination approach where opposing effects could neutralize one another. Second, it draws strength from the collective effects of neighbooring sites whilst avoiding the limitations of the site-level analysis approach.

Four Bonferroni significant DMR were identified in islands which were located near the following genes: LTB, a cytokine shown to simulate osteoclast activity (29); SKIV2L, with an RNA helicase activity, thought to be involved in blocking translation of viral mRNA and has been implicated in regulating host responses to viral infections (30); EBF3 which is involved in bone development and B cell differentiation (31) and CCND1, a Wnt target that was reported to be upregulated in response to mechanical loading of bone (32).

Additionally, six Bonferroni significant DMR in gene bodies were identified. These were located within genes with functions related to mitosis and ciliogenesis (SDCCAG8) (33); TGFB1-mediated signaling (RBPMS) (34); calcium signaling (CACNA1B) (35); protein ubiquitination (HSPA13) (36); cytoskeletal organization (PARD3B) (37) and histone acetylation (BRD1) (38).

The *Pooled* sites identified from the discovery set were able to discriminate cases and controls with a considerable accuracy when tested on the replication set. The Best subset analysis allowed the identification of a smaller subset of sites trading off the classification accuracy with the number of explanatory sites. The AUC of 82.5%, based on the 95 discriminatory sites from the best subset analysis, is promising and future experiments are warranted to study its clinical applicability.

In terms of disease pathology, the DNA methylation data reflected many environmental triggers thought to be involved in PDB. Some of the genes amongst the DMS and the 95 *Best* subset were associated with immune antiviral responses (Figure 5 & Table S5). This is of interest since a previous study in the PRISM cohort showed that levels of antibodies to Mumps virus were significantly higher in PDB cases compared to controls (39). Although we and others have failed to detect evidence of ongoing virus infection in PDB, the above data is consistent with the hypothesis that host immune responses to infection may be altered in PDB.

Differential methylation of ZIC1 and CCND1 indicate possible differences between cases and controls in these genes which are involved in mechano-transduction, a process which has been implicated in localisation of bone lesions in PDB (39, 40). Our study also highlighted genes that regulate the cell cycle, vesicular transport and cytoskeletal reorganization as being potentially involved in PDB. Other genes were identified that play a role in immune cell function and these were strongly represented in the best subset of differentially methylated sites. This lends support to the hypothesis that PDB may be a disorder with an osteoimmunological basis (41) and should prompt further work to investigate host-environment interactions including studies of the microbiome in this complex but fascinating disease (42).

Apart from providing new insights into the potential links between genes and environment in regulating susceptibility to PDB, this study has revealed the potential role of methylation signals as a biomarker for disease susceptibility. Potent bisphosphonates such as zoledronic acid can return the abnormalities of bone remodeling to normal in a large proportion of patients with PDB (4, 43). Unfortunately PDB often remains clinically silent until it has reached an advanced stage by which point irreversible skeletal damage may already have occurred (5). This study raises the possibility that epigenetic markers, possibly when combined with genetic profiling would be worth exploring as means of assessing the risk of developing PDB in people with a family history of the disorder so that early intervention can be considered where clinically appropriate.

One limitation of the study is the fact that the identified methylation changes were not shown to occur in the osteoclasts which are the cells of main interest in Paget's pathogenesis. This is primarily justified by the difficulty to collect bone tissue from PDB patients in a similarly sized cohort. Moreover, showing an epigenetic signature to PDB in the blood adds to the increasing evidence in the literature pointing to the possibility of pathogenic immune processes lying at the heart of PDB. More importantly, a predictive epigenetic signature in a readily accessible tissue such as the blood has clinical implication, also considering the silent nature of PDB and the possibility of avoiding much of the adverse symptoms of the disease with early diagnosis. Finally, one needs to consider that blood also contains progenitors of bone cells and that white blood cells share a similar ancestry with osteoclasts.

Materials and methods

Study Subjects

The DNA samples were derived from UK-based PDB patients and controls who took part in the PRISM trial (Paget's Disease: Randomized Trial of Intensive versus Symptomatic Management (ISRCTN12989577) (44). The PRSIM trial is a multi-center study in which participants were recruited from 27 different clinical centers across the United Kingdom. The epigenetic analysis was conducted in 253 cases with clinical and radiological evidence of PDB and 280 controls who were spouses of PDB cases (n=135) or subjects who had been referred for investigation of osteoporosis but had normal bone density upon examination by dual energy X-ray absorptiometry (n=131). The cohort was randomly divided into a discovery and replication set comprising of comparable numbers of cases and controls (Figure 1). According to the study by Tsai and Bell, a 10% difference in the mean of CpG methylation level between cases and controls at genome-wide significance level of 10^{-6} requires 112 individuals in each group to achieve 80% EWAS power (45). On this basis, our discovery set comprising of 116 cases and 130 controls is adequately powered and the results are further validated in an equally sized replication set.

DNA methylation profiling

Genomic DNA was extracted from peripheral blood using standard protocols. Bisulfite conversion was performed on 500µg of DNA using Zymo EZ-96 DNA methylation Kit (Zymo Research, USA). DNA methylation profiling was performed using the Illumina Infinium HumanMethylation 450K array (Illumina, USA) by following the manufacturer's protocol. The R package *RnBeads* version 1.10.8 was used for quality control (8). Samples with low methylated or unmethylated median intensity (<11.0) were excluded (n=35) along with samples with sex mismatch between reported and predicted sex (n=0). Probes with the following criteria were excluded: detection P value >0.05, cross reactive probes, containing a SNP within 3 bp of nucleotide extension site, or those located on sex chromosomes. Additionally, 723 sites were further excluded from the dataset for previously established association with smoking (46). A total of 56,356 probes were excluded from the initial 485,512 leaving 429,156 CpGs for analysis (Figure 1). The final dataset used for analysis comprised of 232 PDB cases and 260 controls. The *Enmix* method (47) was used for background correction whilst *SWAN* was used to achieve between and within array normalization. For all downstream analysis, the M-values, derived using the formulae $\log_2$((methylated signal+1)/(unmethylated signal+1)), were used.

Statistics

An overview of the analysis performed in this study is shown in Figure 1, in what follows we provide details of each analysis step:

- Differential methylation analysis of sites

In order to account for the heterogenous cellular composition of the measured samples, the counts of the following cell types CD14 monocytes, CD19 B-cells, CD4 T-cells, CD56 NK cells, CD8 T-cells, eosinophils, granulocytes and neutrophils were estimated using the *Houseman* reference method (48), part of the *RnBeads* pipeline. The reference methylome was obtained from previously published methylation data measured from sorted blood cells comprising 47 samples (49). These reference samples were normalized together with our data to make sure that extrapolation of cell type information was unaffected by differences between the two datasets. we performed Surrogate variable analysis (SVA) which captures additional unknown sources of variation based on joint methylation patterns amongst the different sites that do not correlate with the disease. The top 10 significant SVA components were extracted from the data using the SVA functionality in *RnBeads.*

In all statistical models described below, the term *confounders* refer to the following covariates: age, sex, array, bisulfite conversion batch, array scan batch, blood cell composition from the *Houseman* method(48) and the top 10 surrogate variant analysis (SVA) components. The term *phenotype* denotes the control/PDB state of each sample. The term *region* is used to describe clusters of sites along the genome including CpG islands, gene bodies and promoters. CpG islands were delineated in the illumina array manifest file as well as RnBeads annotation libraries. Gene bodies and promoters were manually assigned. More specifically, sites mapping to the transcription start site (TSS) according to the manifest were attributed to a promoter region whilst those falling at the 5' untranslated region or gene body were assigned to a gene body region.

A general linear model based on the limma moderated standard error (50) was used to assess differentially methylated sites (DMS) between cases and controls using the model: *CpG site ~ phenotype + confounders*. The model was first run on all sites in the discovery set and all DMS with a significant FDR (< 0.05) in the discovery set were assessed in the replication set. Meta-analysis looking at the combined effect from both discovery and replication was performed on this subset using the R package *Metafor* (51). The Bonferroni adjusted genome wide significance threshold of $P=1.17 \times 10^{-7}$ (0.05/429,156) was used.

- Differential methylation analysis of regions

Differentially methylated regions (DMR) were analyzed using binomial regression, member of the family of the generalized linear models, in two steps:

First the parameters of the *null* model, excluding the sites, were estimated as follows:

phenotype ~ confounders……………………………………………………...[1]

Next, all *n* sites within a given region (island/gene body/promoter) were incorporated into the model as follows:

phenotype ~ confounders + CpG $site_1$ *+ CpG* $site_2$ *+…. + CpG* $site_n$ ……..[2]

The difference in the deviance (equivalent to the residuals in the linear model) between the null model [1] and the full model [2] follows a χ^2 distribution with n degrees of freedom. A P-value for the effect of the region given n sites was calculated accordingly. The analysis effectively tests for the significance of improvement in the model fit with the addition of the methylation data from the region of interest. The generalized linear model outlined above was run initially on the discovery set. The model was then repeated on the replication set on regions that were significant in the discovery set at FDR < 0.05. A similar approach was used to derive the Bonferroni significant regions. Visualization of the effect of individual sites from selected DMR was conducted using R package coMET (52).

- Consolidating the DMS and DMR

In the Generalized Linear Model for region effect outlined in model formulae [2], the beta values from the individual sites are indicative of the sites' level of association with the phenotype. This is effectively similar to the General Linear Model used for site-level analysis but with the important discrepancy that each site is being assessed while accounting for possible contributions of neighboring sites to the global effect of the region. We therefore extracted all the beta values form the full model in [2] from all the DMR. We then applied fdr based multiple testing correction on the pvalues corresponding to these beta values from fitting the model in [2] for each selected DMR separately. Sites with fdr < 0.05 were pooled with the DMS to create a unified list of significantly methylated sites or *Pooled sites* (Figure 1).

- Discriminant analysis

Discriminant analysis was performed to assess the ability of the *Pooled sites* to tell apart cases from controls. We also used the Elastic-Net Regularization Extension of the Generalized Linear Model, provided by the R package *Glmnet (53)*, to identify the best subset of discriminatory sites (designated *Best subset*) of the list of *Pooled sites*. We trained an Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) classifier (54), implemented in the software SIMCA ver. 15 (Umetrics, Sweden), on the discovery data from *Pooled* and *Best subset* sites separately. Each model was then tested on the replication set and its performance was further assessed based on the area under curve (AUC) value from receiver operating characteristic (ROC) curve analysis. The sensitivity and specificity measures of the test were estimated based on a classification threshold equal to the median of the predicted scores by the OPLS-DA classifier. The *Best subset* sites were analyzed further to reveal enrichment in biological functions. This was conducted using Ingenuity Pathway Analysis (IPA) (Qiagen, Germany) as well as the Gene Ontology (GO) R package *topGO* (55) based on the Fisher's exact test statistics.

- Partial correlation analysis of *Pooled sites*

Correlations in methylation patterns between CpG sites hold valuable information about how different biological functions are linked together in PDB. To this end, partial correlations between the *Pooled sites* were derived using the R package ggm (56). In parallel, the extensive GO functional annotations enriched amongst the genes associated with the *Pooled sites* were manually reduced to a manageable, yet representative, set of keywords: For instance, GO categories 'regulation of proliferation', 'positive regulation of proliferation' and 'negative regulation of proliferation' were all reduced to 'proliferation'. The fisher's exact test statistics was then used to assess whether the *Pooled sites* associated with a given keyword were correlated (based on the ggms) with their counterparts from another functional keyword more often than can be accounted for by chance alone. Fisher's test p-values < 0.05 after FDR multiple testing correction were used to create pairs of functionally related keywords. The software Cytoscape (57) was used to visualize these associations.

Expression quantitative trait methylation (eQTM) analysis

To assess the effect of DNA methylation at CpGs sites on the expression of nearby genes, we used data from the BIOS QTL browser (58).

Acknowledgements

We wish to thank the patients and controls from the different centers who agreed to participate in this study. We would like to thank members of the PRISM trial research group across all participating centers for making DNA samples and data available for this study. We thank the Wellcome Trust Clinical Research Facility at Edinburgh University for performing the DNA methylation profiling.

Competing interests

Prof S H Ralston has received research funding from Amgen, Eli Lilly, Novartis, and Pfizer unrelated to the submitted work. The other authors have no conflicts of interest to declare.

References

1. Tan A, and Ralston SH. Paget's disease of bone. *QJM.* 2014;107(11):865-9.
2. Ralston SH, Corral-Gudino L, Cooper C, Francis RM, Fraser WD, Gennari L, et al. Diagnosis and Management of Paget's Disease of Bone in Adults: A Clinical Guideline. *J Bone Miner Res.* 2019;34(4):579-604.
3. Langston AL, Campbell MK, Fraser WD, MacLennan GS, Selby PL, Ralston SH, et al. Randomized trial of intensive bisphosphonate treatment versus symptomatic management in Paget's disease of bone. *J Bone Miner Res.* 2010;25(1):20-31.
4. Reid IR, Lyles K, Su G, Brown JP, Walsh JP, del Pino-Montes J, et al. A single infusion of zoledronic acid produces sustained remissions in Paget disease: data to 6.5 years. *J Bone Miner Res.* 2011;26(9):2261-70.
5. Gennari L, Rendina D, Falchetti A, and Merlotti D. Paget's Disease of Bone. *Calcif Tissue Int.* 2019;104(5):483-500.
6. Ralston SH, and Albagha OM. Genetics of Paget's disease of bone. *Curr Osteoporos Rep.* 2014;12(3):263-71.
7. Vallet M, Soares DC, Wani S, Sophocleous A, Warner J, Salter DM, et al. Targeted sequencing of the Paget's disease associated 14q32 locus identifies several missense coding variants in RIN3 that predispose to Paget's disease of bone. *Hum Mol Genet.* 2015;24(11):3286-95.
8. Albagha OM, Wani SE, Visconti MR, Alonso N, Goodman K, Brandi ML, et al. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. *Nat Genet.* 2011;43(7):685-9.
9. Albagha OM, Visconti MR, Alonso N, Langston AL, Cundy T, Dargie R, et al. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. *Nat Genet.* 2010;42(6):520-4.
10. Obaid R, Wani SE, Azfer A, Hurd T, Jones R, Cohen P, et al. Optineurin Negatively Regulates Osteoclast Differentiation by Modulating NF-kappaB and Interferon Signaling: Implications for Paget's Disease. *Cell Rep.* 2015;13(6):1096-102.
11. Wong SW, Huang BW, Hu X, Ho Kim E, Kolb JP, Padilla RJ, et al. Global deletion of Optineurin results in altered type I IFN signaling and abnormal bone remodeling in a model of Paget's disease. *Cell Death Differ.* 2020;27(1):71-84.
12. Corral-Gudino L, Borao-Cengotita-Bengoa M, Del Pino-Montes J, and Ralston S. Epidemiology of Paget's disease of bone: a systematic review and meta-analysis of secular changes. *Bone.* 2013;55(2):347-52.
13. Kurihara N, Hiruma Y, Yamana K, Michou L, Rousseau C, Morissette J, et al. Contributions of the measles virus nucleocapsid gene and the SQSTM1/p62(P392L) mutation to Paget's disease. *Cell Metab.* 2011;13(1):23-34.
14. Kurihara N, Zhou H, Reddy SV, Garcia Palacios V, Subler MA, Dempster DW, et al. Expression of measles virus nucleocapsid protein in osteoclasts induces Paget's disease-like bone lesions in mice. *J Bone Miner Res.* 2006;21(3):446-55.
15. Ooi EL, Chan ST, Cho NE, Wilkins C, Woodward J, Li M, et al. Novel antiviral host factor, TNK1, regulates IFN signaling through serine phosphorylation of STAT1. *Proc Natl Acad Sci U S A.* 2014;111(5):1909-14.
16. Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Philtjens S, Heeman B, et al. Loss of TBK1 is a frequent cause of frontotemporal dementia in a Belgian cohort. *Neurology.* 2015;85(24):2116-25.
17. Watts GD, Wymer J, Kovach MJ, Mehta SG, Mumm S, Darvish D, et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet.* 2004;36(4):377-81.
18. Zhang Y, Lu J, and Liu X. MARCH2 is upregulated in HIV-1 infection and inhibits HIV-1 production through envelope protein translocation or degradation. *Virology.* 2018;518:293-300.

19. Kuo WJ, Digman MA, and Lander AD. Heparan sulfate acts as a bone morphogenetic protein coreceptor by facilitating ligand-induced receptor hetero-oligomerization. *Mol Biol Cell.* 2010;21(22):4028-41.
20. Seitz S, Barvencik F, Gebauer M, Albers J, Schulze J, Streichert T, et al. Preproenkephalin (Penk) is expressed in differentiated osteoblasts, and its deletion in Hyp mice partially rescues their bone mineralization defect. *Calcif Tissue Int.* 2010;86(4):282-93.
21. IMPC.
22. Nishikawa K, Nakashima T, Takeda S, Isogai M, Hamada M, Kimura A, et al. Maf promotes osteoblast differentiation in mice by mediating the age-related switch in mesenchymal cell differentiation. *J Clin Invest.* 2010;120(10):3455-65.
23. Bornstein C, Brosh R, Molchadsky A, Madar S, Kogan-Sakin I, Goldstein I, et al. SPATA18, a spermatogenesis-associated gene, is a novel transcriptional target of p53 and p63. *Mol Cell Biol.* 2011;31(8):1679-89.
24. Courtial N, Smink JJ, Kuvardina ON, Leutz A, Gothert JR, and Lausen J. Tal1 regulates osteoclast differentiation through suppression of the master regulator of cell fusion DC-STAMP. *FASEB J.* 2012;26(2):523-32.
25. Kalogeropoulos M, Varanasi SS, Olstad OK, Sanderson P, Gautvik VT, Reppe S, et al. Zic1 transcription factor in bone: neural developmental protein regulates mechanotransduction in osteocytes. *FASEB J.* 2010;24(8):2893-903.
26. Varanasi SS, Olstad OK, Swan DC, Sanderson P, Gautvik VT, Reppe S, et al. Skeletal site-related variation in human trabecular bone transcriptome and signaling. *PLoS One.* 2010;5(5):e10692.
27. Ly LL, Yoshida H, and Yamaguchi M. Nuclear transcription factor Y and its roles in cellular processes related to human disease. *Am J Cancer Res.* 2013;3(4):339-46.
28. Shibata S, Teshima Y, Niimi K, and Inagaki S. Involvement of ARHGEF10, GEF for RhoA, in Rab6/Rab8-mediating membrane traffic. *Small GTPases.* 2019;10(3):169-77.
29. Horowitz M C LJA. In: Direct S ed. *Principles of Bone Biology*. 2002:961-77.
30. Eckard SC, Rice GI, Fabre A, Badens C, Gray EE, Hartley JL, et al. The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nat Immunol.* 2014;15(9):839-45.
31. Seike M, Omatsu Y, Watanabe H, Kondoh G, and Nagasawa T. Stem cell niche-specific Ebf3 maintains the bone marrow cavity. *Genes Dev.* 2018;32(5-6):359-72.
32. Holguin N, Brodt MD, and Silva MJ. Activation of Wnt Signaling by Mechanical Loading Is Impaired in the Bone of Old Mice. *J Bone Miner Res.* 2016;31(12):2215-26.
33. Insolera R, Shao W, Airik R, Hildebrandt F, and Shi SH. SDCCAG8 regulates pericentriolar material recruitment and neuronal migration in the developing cortex. *Neuron.* 2014;83(4):805-22.
34. Shanmugaapriya S, van Caam A, de Kroon L, Vitters EL, Walgreen B, van Beuningen H, et al. Expression of TGF-beta Signaling Regulator RBPMS (RNA-Binding Protein With Multiple Splicing) Is Regulated by IL-1beta and TGF-beta Superfamily Members, and Decreased in Aged and Osteoarthritic Cartilage. *Cartilage.* 2016;7(4):333-45.
35. Blair HC, Schlesinger PH, Huang CL, and Zaidi M. Calcium signalling and calcium transport in bone disease. *Subcell Biochem.* 2007;45:539-62.
36. Kaye FJ, Modi S, Ivanovska I, Koonin EV, Thress K, Kubo A, et al. A family of ubiquitin-like proteins binds the ATPase domain of Hsp70-like Stch. *FEBS Lett.* 2000;467(2-3):348-55.
37. Kohjima M, Noda Y, Takeya R, Saito N, Takeuchi K, and Sumimoto H. PAR3beta, a novel homologue of the cell polarity protein PAR3, localizes to tight junctions. *Biochem Biophys Res Commun.* 2002;299(4):641-6.
38. Mishima Y, Wang C, Miyagi S, Saraya A, Hosokawa H, Mochizuki-Kashio M, et al. Histone acetylation mediated by Brd1 is crucial for Cd8 gene activation during early thymocyte development. *Nat Commun.* 2014;5:5872.
39. Visconti MR, Usategui-Martin R, and Ralston SH. Antibody Response to Paramyxoviruses in Paget's Disease of Bone. *Calcif Tissue Int.* 2017;101(2):141-7.
40. Gasper TM. Paget's disease in a treadle machine operator. *Br Med J.* 1979;1(6172):1217-8.

41. Numan MS, Amiable N, Brown JP, and Michou L. Paget's disease of bone: an osteoimmunological disorder? *Drug Des Devel Ther.* 2015;9:4695-707.
42. Ohlsson C, and Sjogren K. Osteomicrobiology: A New Cross-Disciplinary Research Field. *Calcif Tissue Int.* 2018;102(4):426-32.
43. Reid IR, Miller P, Lyles K, Fraser W, Brown JP, Saidi Y, et al. Comparison of a single infusion of zoledronic acid with risedronate for Paget's disease. *N Engl J Med.* 2005;353(9):898-908.
44. Tan A, Goodman K, Walker A, Hudson J, MacLennan GS, Selby PL, et al. Long-Term Randomized Trial of Intensive Versus Symptomatic Management in Paget's Disease of Bone: The PRISM-EZ Study. *J Bone Miner Res.* 2017;32(6):1165-73.
45. Tsai PC, and Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol.* 2015;44(4):1429-41.
46. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics.* 2016;8(5):599-618.
47. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, and Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013;14:293.
48. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86.
49. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012;7(7):e41361.
50. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.
51. W V. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software.* 2010;36(3):48.
52. Martin TC, Yet I, Tsai PC, and Bell JT. coMET: Visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics.* 2015.
53. Friedman J, Hastie T, and Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.
54. Boccard J, and Rutledge DN. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal Chim Acta.* 2013;769:30-9.
55. Alexa A RJ. topGO: enrichment analysis for Gene Ontology. 2020.
56. M GM. Independencies Induced from a Graphical Markov Model After Marginalization and Conditioning: The R Package ggm. *Journal of Statistical Software.* 2006;15(6).
57. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
58. BIOS QTL.

Figures

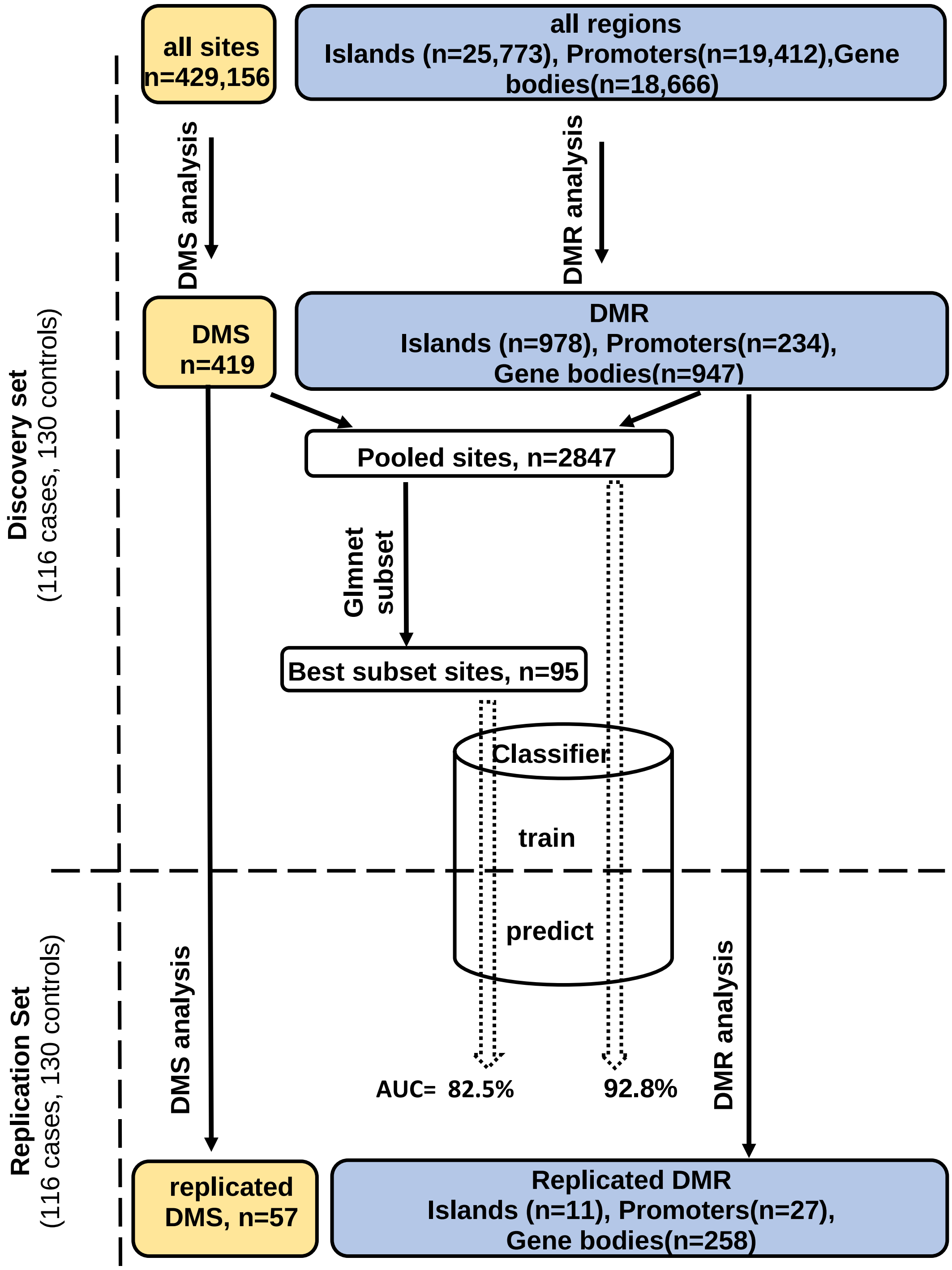

Figure 1. Study design and analysis workflow. Differentially methylated sites (*DMS*) and differentially methylated regions (*DMR*) were analyzed using, the General/Generalized linear model respectively, in the discovery set. Those reaching FDR < 0.05 were tested in the replication set to identify DMS/DMR that replicate at the same significance level. The DMS and the important sites within DMR were pooled together giving rise to the *Pooled sites* (refer to methods), of these a best PDB discriminatory subset was obtained using the Lasso and Elastic-Net regression. A multivariate classifier based on the discovery measurement of the Pooled/Best subset sites yielded an AUC value of 92.8% and 82.5% respectively when tested in the replication.

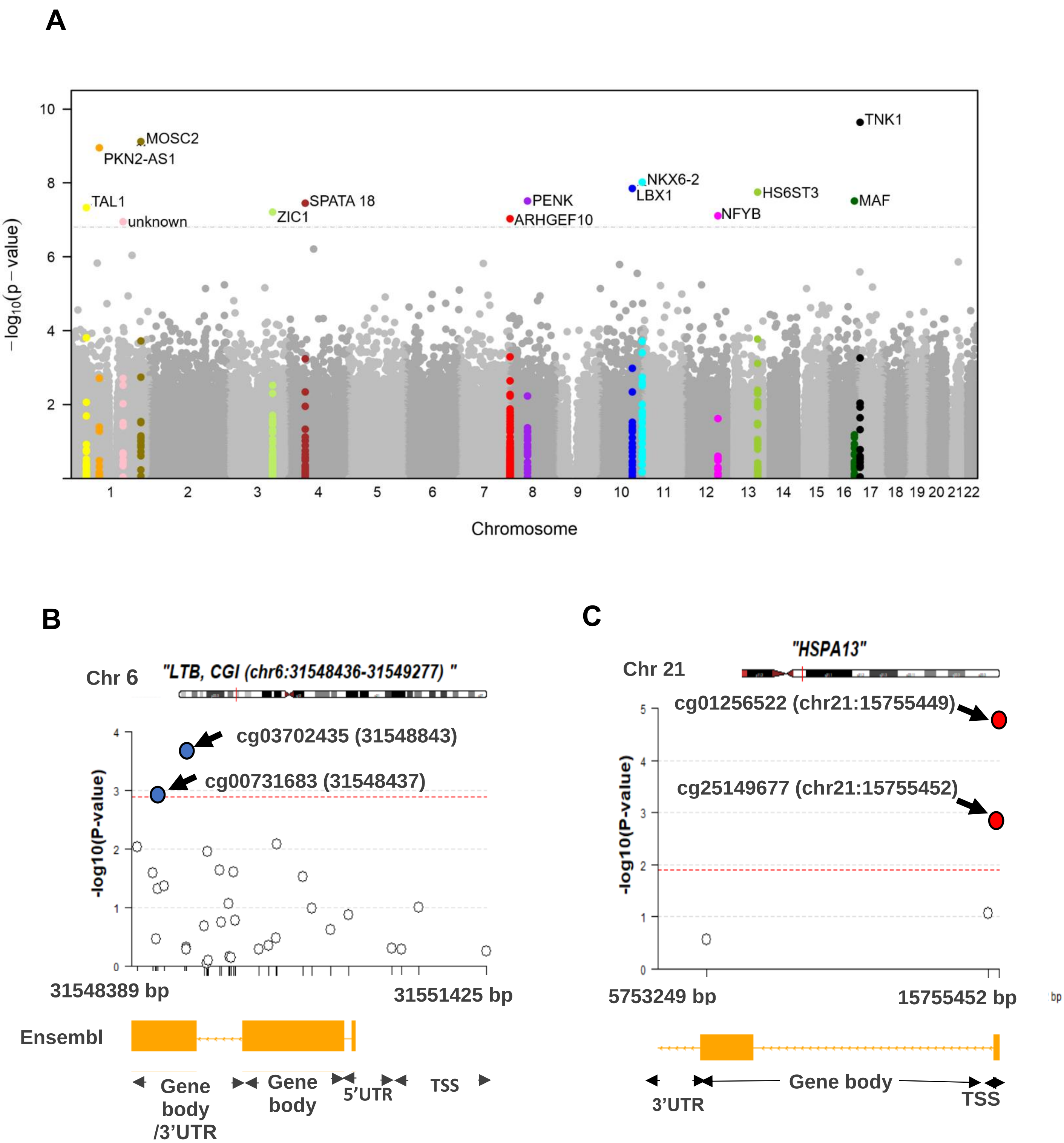

Figure 2. Differential methylation analysis comparing controls to PDB patients (n=246). A) Site analysis, a Manhattan plot showing the chromosomal positions (x-axis) versus the -log10 (P) of significant DMS and adjacent sites. For the Bonferroni significant sites however, the meta-analysis P-values are shown instead and highlighted in color. The horizontal dashed line indicates the Bonferroni corrected significance threshold ($P < 1.17 \times 10^{-7}$). B&C) Region analysis, showing the multitude of significantly hypermethylated (red) and hypomethylated (blue) sites from LTB (Bonferroni replicated from island analysis) and HSPA13 (Bonferroni replicated from gene body analysis). The dashed lines represent the fdr < 0.05 threshold for each region which depends on the number of sites within the region (refer to methods).

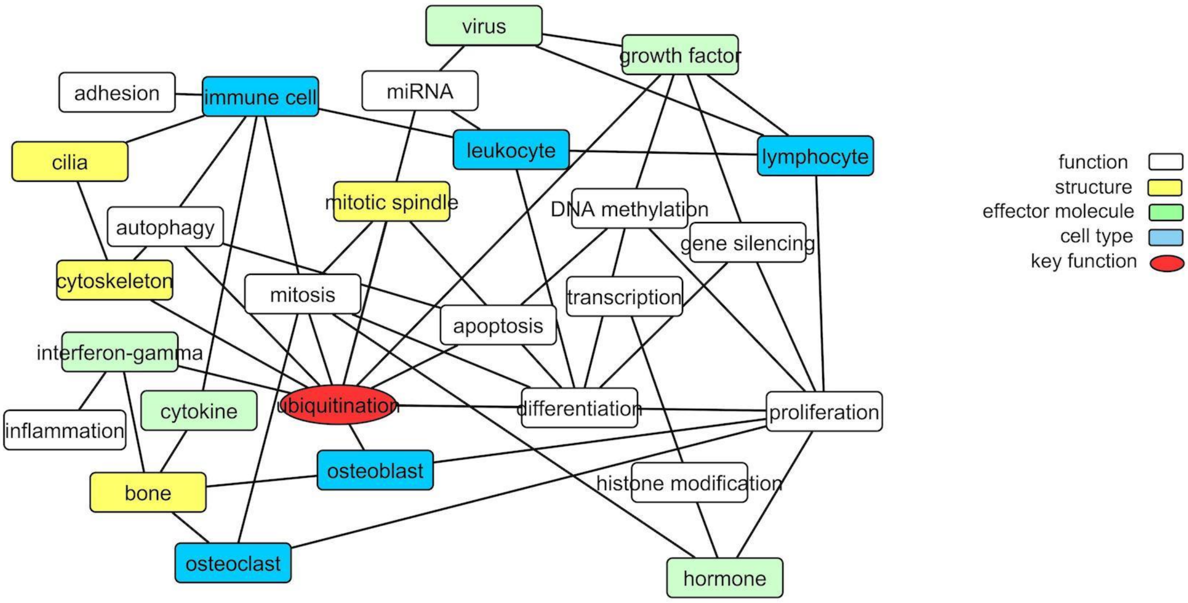

Figure 3: Translating the methylation data into functional network. Nodes are functional, cellular, molecular and sub-cellular keywords from GO annotations enriched amongst the *Pooled sites*. An edge between two nodes indicates that differentially methylated genes associated with the keyword in node 1 are significantly partially correlated with their counterparts from node 2 more often than can be accounted for by chance.

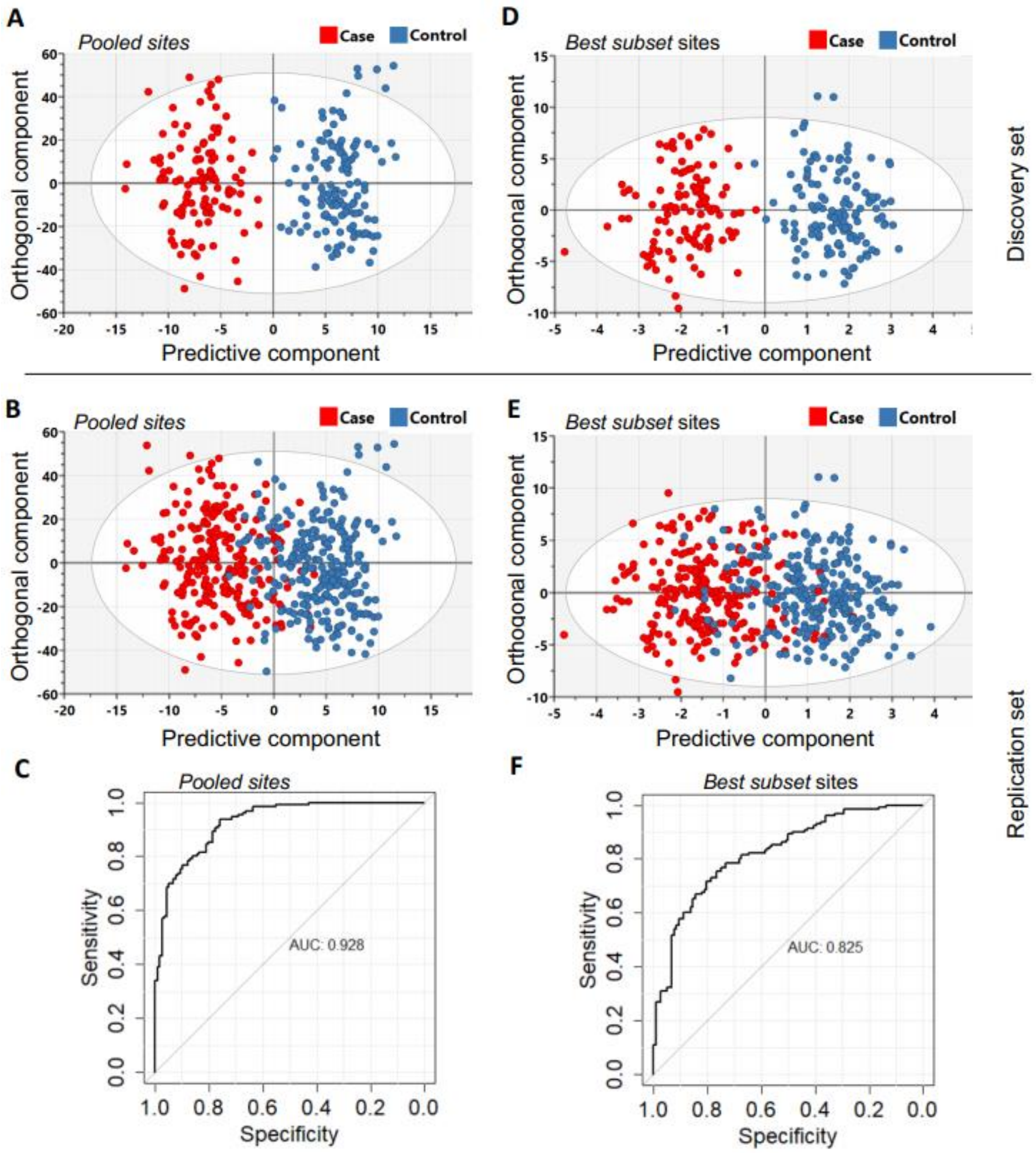

Figure 4. The orthogonal partial least squares discriminant analysis (OPLS-DA) was performed using the *Pooled sites* identified from the discovery set (n=246). (**A**) Classifier trained on all 2847 pooled sites with FDR < 0.05 (*Pooled sites*) from the discovery set. (**B**) Testing the classifier on the replication set. (**C**) ROC curve analysis yielded an overall sensitivity of 0.84, specificity of 0.81 and AUC=0.928. (**D**) Classifier trained on the *Best subset* sites from Glmnet analysis (n=95) using the discovery set. (**E**) Testing the classifier on the replication set. (**F**) ROC curve analysis showed an overall sensitivity of 0.77, specificity of 0.74 and AUC=0.825. The Scatter plots in A,B,C&D show the predictive component that discriminates PDB cases from controls (x-axis) versus the orthogonal component representing a multivariate confounding effect that is independent of PDB (y-axis).

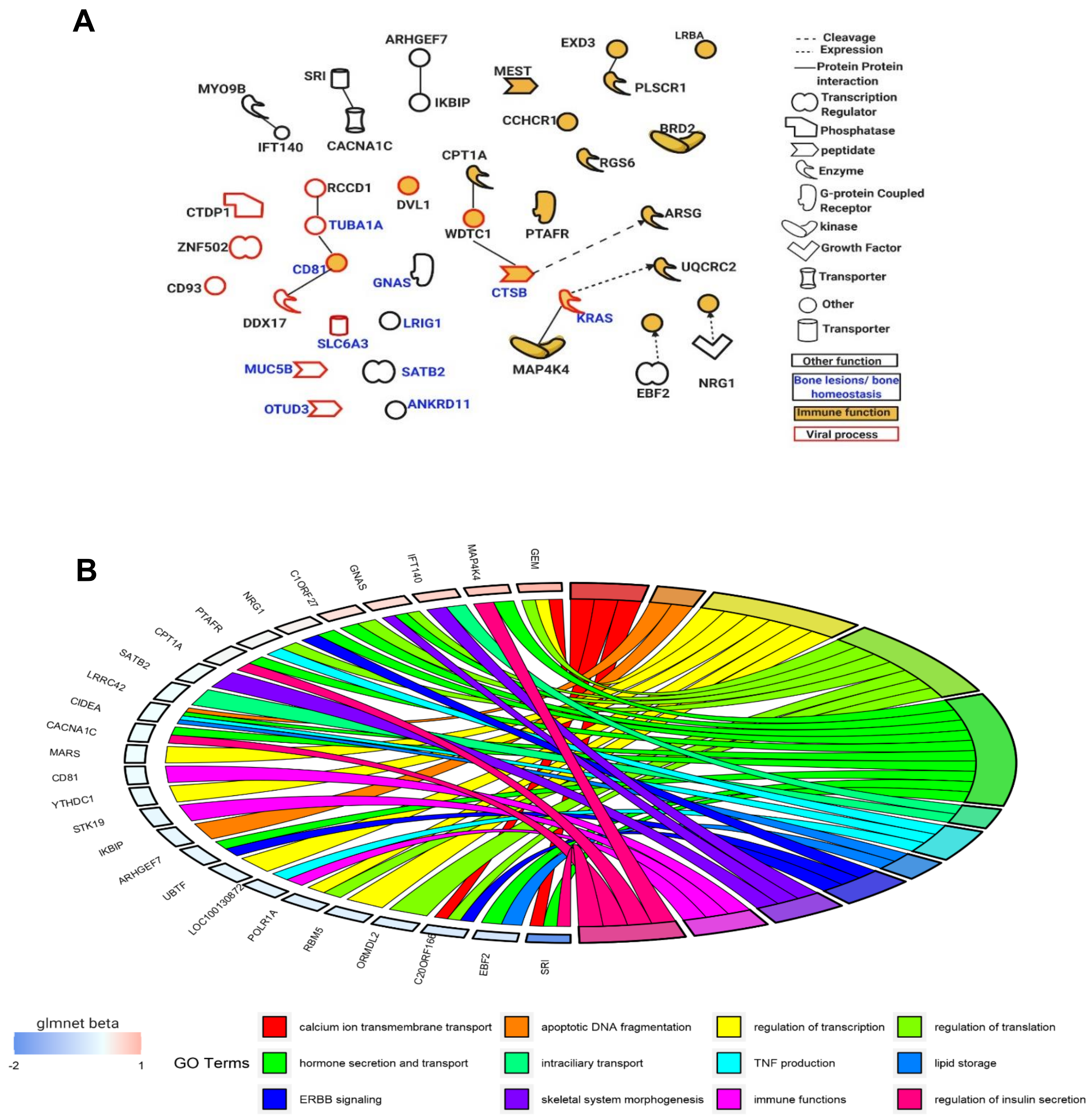

Figure 5: Functions of genes mapped near the *Best subset* of differentially methylated sites identified through the Elastic-Net Regularization Extension of the Generalized Linear Model. A) An IPA based network showning a subset of these genes with functional interactions (edges) or mapping to one of three functional classes: immune, viral and bone homeostasis. B) An overview of GO biological processes significantly enriched amongst the Best subset together with their beta values from the Glmnet R package implementing the extended Generalized Linear Model in question.

Tables

Table 1. Descriptive statistics of the study cohort

	Discovery		Replication	
	PDB Case	Control	PDB Case	Control
Number	116	130	116	130
Age (years), mean ± SD	72.1 ± 7.5*	70.0 ±7.4*	72.5 ± 8.7	72.3 ± 8.2
Males, n (%)	65 (56.0)*	48 (36.9)*	59 (50.9)	53 (40.8)
Females, n (%)	51 (44.0)*	82 (63.1)*	57 (49.1)	77 (59.2)
SQSTM1 Mutation n (%)	16 (13.8)	0 (0)	17 (14.6)	0 (0)

*P<0.05 comparing Paget's disease (PDB) cases to controls.

Table 2. Differentially methylated CpG sites (DMS) in Paget's disease of bone

CpG Site			Discovery		Replication		Meta-analysis		Annotations
Probe ID	**Chr**	**Position**	**Δ Beta***	**P Value**	**Δ Beta***	**P Value**	**Δ Beta***	**P Value**	**Nearest gene**
cg10290814	17	7284330	-0.018	$1.2X10^{-6}$	-0.015	$1.4x10^{-4}$	-0.017	$2.3x10^{-10}$	*TNK1*
cg19361865	1	220922163	-0.014	$5.4x10^{-6}$	-0.012	$9.7x10^{-5}$	-0.013	$7.6x10^{-10}$	*MOSC2*
cg09152582	1	88928362	-0.021	$2.1x10^{-5}$	-0.018	$3.5x10^{-5}$	-0.019	$1.1x10^{-9}$	*PKN2-AS1*
cg09260089	10	134599860	-0.024	$4.6x10^{-5}$	-0.024	$1.2x10^{-4}$	-0.024	$9.5x10^{-9}$	*NKX6-2*
cg24879273	10	102989645	-0.026	$4.9x10^{-5}$	-0.016	$1.7x10^{-4}$	-0.021	$1.4x10^{-8}$	*LBX1*
cg03839709	13	96743492	-0.014	$2.7x10^{-4}$	-0.014	$3.4x10^{-5}$	-0.014	$1.8x10^{-8}$	*HS6ST3*
cg16419235	8	57360613	-0.036	$1.9x10^{-4}$	-0.029	$8.3x10^{-5}$	-0.032	$3.1x10^{-8}$	*PENK*
cg04317962	16	79623625	-0.017	$1.4x10^{-6}$	-0.019	$2.9x10^{-3}$	-0.018	$3.1x10^{-8}$	*MAF*
cg01429039	4	52918065	-0.023	$1.8x10^{-4}$	-0.020	$1.1x10^{-4}$	-0.021	$3.5x10^{-8}$	*SPATA18*
cg03885399	1	47691550	-0.020	$4.4x10^{-6}$	-0.014	$3.6x10^{-3}$	-0.017	$4.7x10^{-8}$	*TAL1*
cg04738965	3	147127662	-0.037	$4.0x10^{-5}$	-0.028	$7.1x10^{-4}$	-0.033	$6.2x10^{-8}$	*ZIC1*
cg10954182	12	104532377	-0.016	$1.9x10^{-4}$	-0.009	$2.1x10^{-4}$	-0.013	$7.8x10^{-8}$	*NFYB*
cg10964367	8	1771973	-0.025	$1.3x10^{-4}$	-0.019	$3.8x10^{-4}$	-0.022	$9.4x10^{-8}$	*ARHGEF10*
cg12739454	1	164290833	-0.018	$2.4x10^{-4}$	-0.012	$2.4x10^{-4}$	-0.015	$1.1x10^{-7}$	-

*Δ Beta represents the difference in DNA methylation in cases as compared to controls (Beta Control-Beta PDB). Position in base pairs in reference to human genome build 37 (GRCh37). Chr, chromosome; CpG, cytosine-phosphate-guanine. All Pvalues are genome-wide significant based on Bonferroni corrected pvalue < 0.05.

Table 3. Differentially methylated regions (DMR) in Paget's disease of bone

Region	Chr	Number of sites	Discovery P-Value*	Replication P-value*	Gene
Island	6	53	1.40×10^{-2}	3.25×10^{-4}	*LTB*
Island	6	59	4.11×10^{-3}	2.47×10^{-3}	*SKIV2L;RDBP*
Island	10	49	2.65×10^{-3}	4.72×10^{-3}	*EBF3*
Island	11	49	3.57×10^{-3}	9.52×10^{-3}	*CCND1*
Gene Body	1	52	2.01×10^{-5}	3.14×10^{-5}	*SDCCAG8*
Gene Body	9	36	6.09×10^{-3}	1.20×10^{-2}	*CACNA1B*
Gene Body	8	51	2.49×10^{-2}	4.39×10^{-3}	*RBPMS*
Gene Body	21	5	3.19×10^{-2}	2.88×10^{-3}	*HSPA13*
Gene Body	2	52	3.80×10^{-2}	2.39×10^{-3}	*PARD3B*
Gene Body	22	34	4.49×10^{-2}	7.10×10^{-3}	*BRD1*

*P-values are adjusted for multiple testing using the Bonferroni method