Edinburgh Research Explorer

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

OPEN ACCESS

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

Ilhame Diboun[1], Sachin Wani[2], Stuart H Ralston[2], Omar ME 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, United Kingdom

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 cross-validation set, each comprising 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 expression quantitative trait-methylation 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 cross-validation 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, and pelvis. Paget's disease is clinically silent until it has reached an advanced stage at which point irreversible damage to the skeleton has occurred (*Tan and Ralston, 2014*). Bisphosphonates are an effective treatment (*Ralston et al., 2019*) and can often improve bone pain but have a limited impact on other clinical outcomes in patients with advanced disease (*Langston et al., 2010*; *Reid et al., 2011*). 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 (*Gennari et al., 2019*; *Ralston and Albagha, 2014*). Mutations in *SQSTM1* are the most common cause of PDB, but other susceptibility genes and loci have been identified through genome-wide association studies (*Vallet et al., 2015*; *Albagha et al., 2011*; *Albagha et al., 2010*). 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 (*Obaid et al., 2015*). 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 (*Obaid et al., 2015*; *Wong et al., 2020*).

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 (*Corral-Gudino et al., 2013*). Several

eLife digest Our skeleton stays healthy through an endless regeneration process, with specialized cells constantly absorbing and creating new bone tissue. Illnesses emerge when this breaking down and rebuilding cycle becomes imbalanced. For instance, in Paget's disease of bone (PDB for short) the skeleton becomes misshapen and fragile, with complications including pain, fractures, neurological problems, hearing loss and even cancer. For most patients however, symptoms are only present at an advanced stage, when irreversible damage to the skeleton has already occurred.

Certain inherited genetic changes play a role in the development of PDB, but lifestyle and environmental factors are also thought to contribute. Indeed, accumulating evidence suggests that diet, pollution and infection may influence how genes involved in bone metabolism are activated. In this process, the environment may trigger chemical marks to be added onto DNA sequences, which ultimately switches specific genes on and off.

To investigate whether the pattern of chemical marks in individuals with PDB may be characteristic, Diboun et al. scanned the genetic information of over 200 PDB patients, and compared it to healthy counterparts. Combining genomic analysis and machine learning revealed several chemical signatures that were remarkably different in the DNA of PDB individuals. These signatures affected sites close to genes involved in bone development, as well as response to mechanical loading and infection. This provides strong evidence that PDB could be, in part, triggered by the environment, as the placement of these marks is highly influenced by external factors.

This research sheds light onto the underlying changes that trigger PDB. Future experiments should explore whether it may be possible to use these genetic changes to identify patients before the onset of irreversible and debilitating damage.

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 (*Ralston and Albagha, 2014*).

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 (*Kurihara et al., 2011*; *Kurihara et al., 2006*). However, clinical studies that have sought to detect evidence of viral proteins and nucleic acids in humans with PDB have yielded conflicting results (*Ralston et al., 2019*).

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 cross-validation set. The number of patients with *SQSTM1* mutations was similar in the discovery and cross-validation set and accounts for approximately 14% of PDB cases. All controls were negative for *SQSTM1* mutations as shown in *Table 1*.

Differentially methylated sites

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 differentially methylated sites (DMS) with false discovery rate (FDR) < 0.05, of which 57 reached statistical

Table 1. Descriptive statistics of the study cohort.

	Discovery		Cross-validation	
	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
Male, n (%)	65 (56.0)*	48 (36.9)*	59 (50.9)	53 (40.8)
Female, 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.

significance (FDR < 0.05) in the cross-validation set (*Supplementary file 1*). Meta-analysis of the DMS from discovery and cross-validation revealed 14 Bonferroni significant DMS out of a total of 429,156 tested CpG sites (p<1.17×10^{-7}; *Table 2*). The direction of effect for all replicated DMS was identical in the discovery and cross-validation set and shows hypermethylation in PDB cases compared to controls. A Manhattan plot of the results is shown in *Figure 2A*, and a quantile–quantile (Q–Q) plot is presented in *Figure 2—figure supplement 1*.
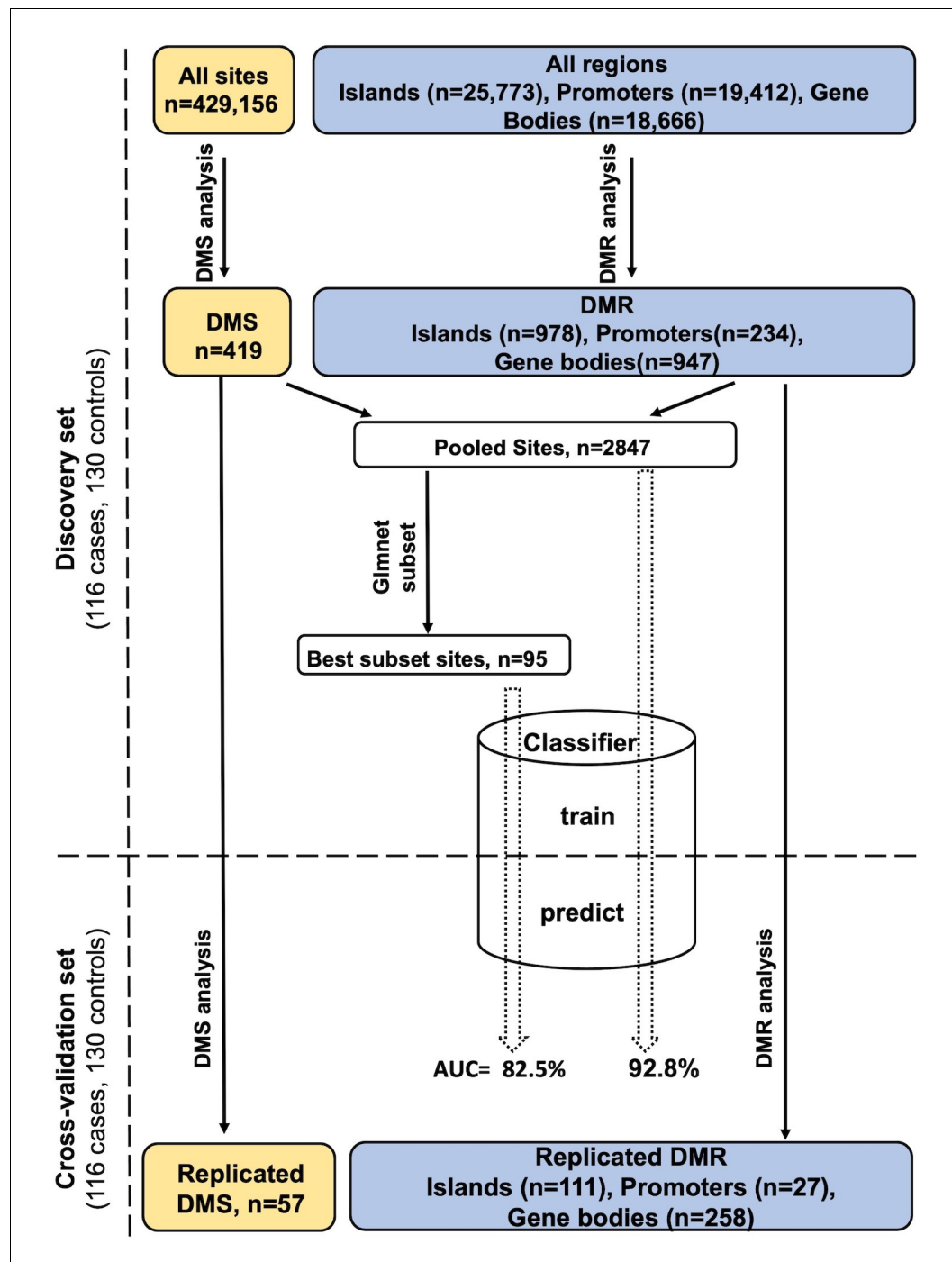
Differentially methylated regions

Besides analyzing individual sites, our region-based analysis was intended to uncover densely hyper/ hypo-methylated regions with unique effects 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 with independent effects 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 differentially methylated regions (DMR) that were significantly differentially methylated (FDR < 0.05) in the discovery set, of which 111 replicated at the same significance level in the cross-validation set (*Supplementary file 2*). Stringent Bonferroni multiple testing correction revealed four islands that remained significant in the discovery and cross-validation, 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 (*Supplementary file 3*). Six gene body DMR reached significance after Bonferroni correction in both the discovery and cross-validation set (*Table 3*). In the context of promoter regions, evidence for FDR significant association with the disease was equally observed in the discovery and cross-validation set for 27 promoters DMR (*Supplementary file 4*), but none reached significance after Bonferroni correction. *Figure 2B and C* show a regional plot for DMR within LTB and HSPA13 from island and gene body analysis respectively, highlighting the co-occurrence of multiple, yet independent, 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 (a unified list of 2847 candidate CpGs identified from the DMS and DMR analysis, refer to Materials and methods). *Figure 3* shows a graphical representation of these functional keywords. 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.

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 cross-validation 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 Materials and 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 cross-validation.

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

CpG Site			Discovery		Cross-validation		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	1.2×10^{-6}	−0.015	1.4×10^{-4}	−0.017	2.3×10^{-10}	TNK1
cg19361865	1	220922163	−0.014	5.4×10^{-6}	−0.012	9.7×10^{-5}	−0.013	7.6×10^{-10}	MOSC2
cg09152582	1	88928362	−0.021	2.1×10^{-5}	−0.018	3.5×10^{-5}	−0.019	1.1×10^{-9}	PKN2-AS1
cg09260089	10	134599860	−0.024	4.6×10^{-5}	−0.024	1.2×10^{-4}	−0.024	9.5×10^{-9}	NKX6-2
cg24879273	10	102989645	−0.026	4.9×10^{-5}	−0.016	1.7×10^{-4}	−0.021	1.4×10^{-8}	LBX1
cg03839709	13	96743492	−0.014	2.7×10^{-4}	−0.014	3.4×10^{-5}	−0.014	1.8×10^{-8}	HS6ST3
cg16419235	8	57360613	−0.036	1.9×10^{-4}	−0.029	8.3×10^{-5}	−0.032	3.1×10^{-8}	PENK
cg04317962	16	79623625	−0.017	1.4×10^{-6}	−0.019	2.9×10^{-3}	−0.018	3.1×10^{-8}	MAF
cg01429039	4	52918065	−0.023	1.8×10^{-4}	−0.020	1.1×10^{-4}	−0.021	3.5×10^{-8}	SPATA18
cg03885399	1	47691550	−0.020	4.4×10^{-6}	−0.014	3.6×10^{-3}	−0.017	4.7×10^{-8}	TAL1
cg04738965	3	147127662	−0.037	4.0×10^{-5}	−0.028	7.1×10^{-4}	−0.033	6.2×10^{-8}	ZIC1
cg10954182	12	104532377	−0.016	1.9×10^{-4}	−0.009	2.1×10^{-4}	−0.013	7.8×10^{-8}	NFYB
cg10964367	8	1771973	−0.025	1.3×10^{-4}	−0.019	3.8×10^{-4}	−0.022	9.4×10^{-8}	ARHGEF10
cg12739454	1	164290833	−0.018	2.4×10^{-4}	−0.012	2.4×10^{-4}	−0.015	1.1×10^{-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 p-values are genome-wide significant based on Bonferroni corrected p-value < 0.05.

Diagnostic capacity of differentially methylated markers

In order to determine whether differentially methylated markers might be of diagnostic value, we performed orthogonal partial least squares-discriminant analysis (OPLS-DA) in the discovery and cross-validation cohorts (refer to Materials and methods). The results are summarized 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 Materials and methods for further details) and when the classifier was tested on the cross-validation set, it yielded an area under curve (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 resulted in the identification of 95 sites (which we also refer to as '*Best subset*' sites; *Supplementary file 5*), 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* sites was consistent between Ingenuity Pathway Analysis (IPA) and Gene Ontology (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 5A*). 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=1.9×10^{-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 5B*.
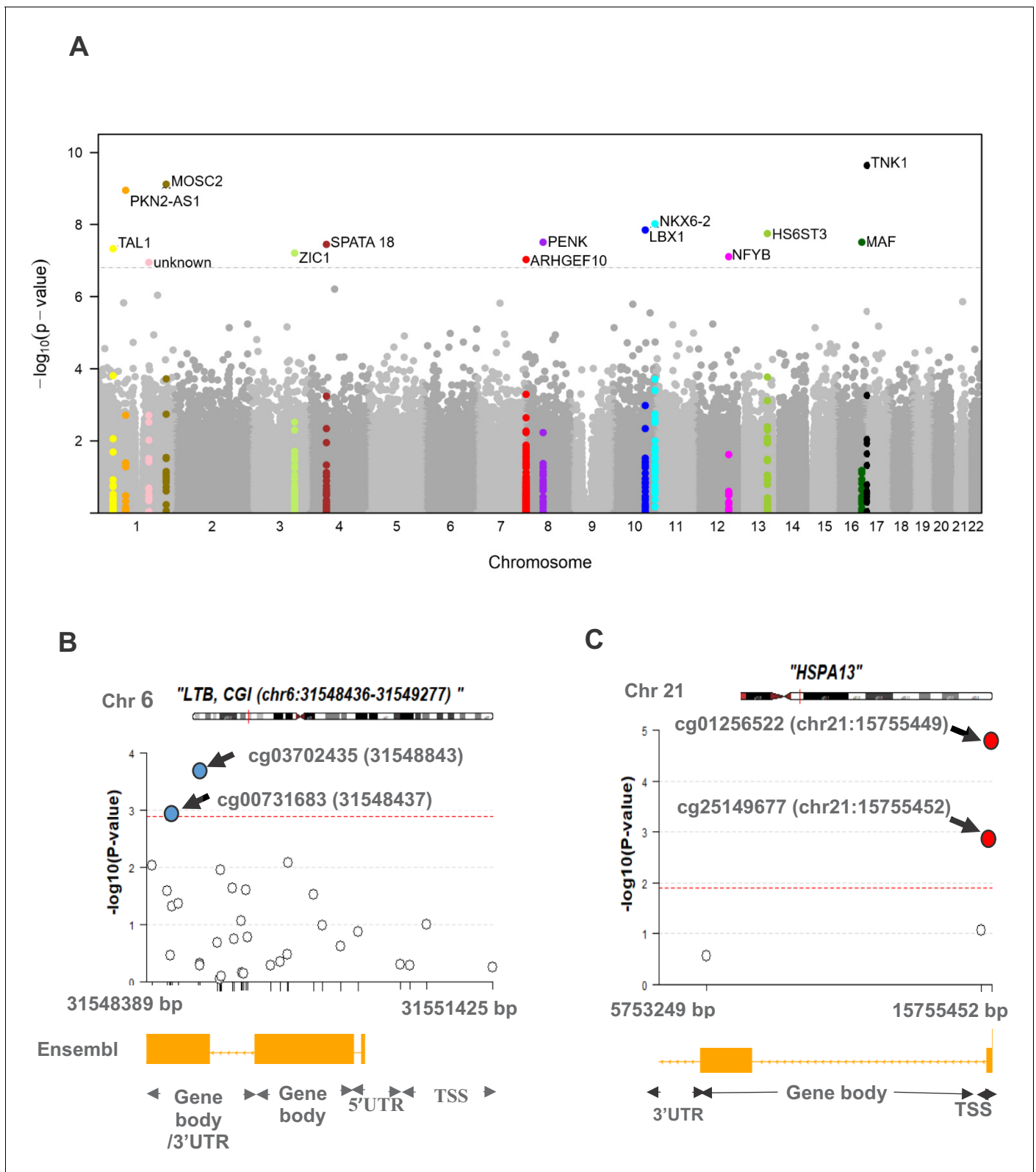
Figure 2. Differential methylation analysis comparing controls to PDB patients (n = 246). (**A**) Site analysis, a Manhattan plot showing the chromosomal positions (x-axis) versus the −log10 (p) of significant DMS and adjacent sites. For the Bonferroni significant sites however, the meta-analysis p-values are shown instead and highlighted in color. The horizontal dashed line indicates the Bonferroni corrected significance threshold (p<1.17×10^{-7}). (**B, C**) Region analysis, showing the multitude of significantly hyper-methylated (red) and hypo-methylated (blue) sites from LTB (Bonferroni replicated from

Figure 2 continued on next page

Correlation of methylation profiles between blood and bone tissue

DNA methylation profiles are known to be tissue specific, and our DMS and DMR analyses were performed on blood, but the primary relevant tissue in PDB is bone. Therefore, we assessed if the methylation profiles for the DMS and DMR identified from this study are correlated between blood and bone tissue using previously published data by *Ebrahimi et al., 2021*. In their study, Ebrahimi et al. focused the correlation analysis on 64,349 CpG probes that fit their analysis criteria to define the most highly correlated positions, of which 28,549 CpG sites showed significant (FDR < 0.05) high correlation ($r^2 > 0.74$) between bone and blood. We assessed if CpG sites annotated to genes identified from our DMS and DMR analyses (*Tables 2* and *3*) showed high correlation between bone and blood as reported by *Ebrahimi et al., 2021*. Results showed that CpGs annotated to 8 of the 14 genes from our DMS analysis were among the highly correlated sites between blood and bone ($r^2 > 0.74$; FDR < 0.05); *Supplementary file 6*. For DMRs, of the 10 genes reported in our study (*Table 3*), 6 had at least one CpG with high correlation between blood and bone (*Supplementary file 6*).

Expression quantitative trait-methylation (eQTM) analysis

eQTM analysis, based on the BIOS QTL (*Bonder et al., 2017*; *Bios QTL, 2021*), showed that the Bonferroni significant DMS cg10964367 was associated with the expression level of ARHGEF10 ($p=3.9\times10^{-9}$). Additionally, cg26724726 from gene body analysis was associated with the expression of LTB ($p=1.10\times10^{-5}$), and eight of the *Best subset* sites were associated with the expression of nearby genes (*Supplementary file 7*).

Discussion

The present study is the first to investigate DNA methylation profiles in PDB. DNA methylation profiles from PDB patients were compared to controls, and meta-analysis of discovery and cross-validation 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

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

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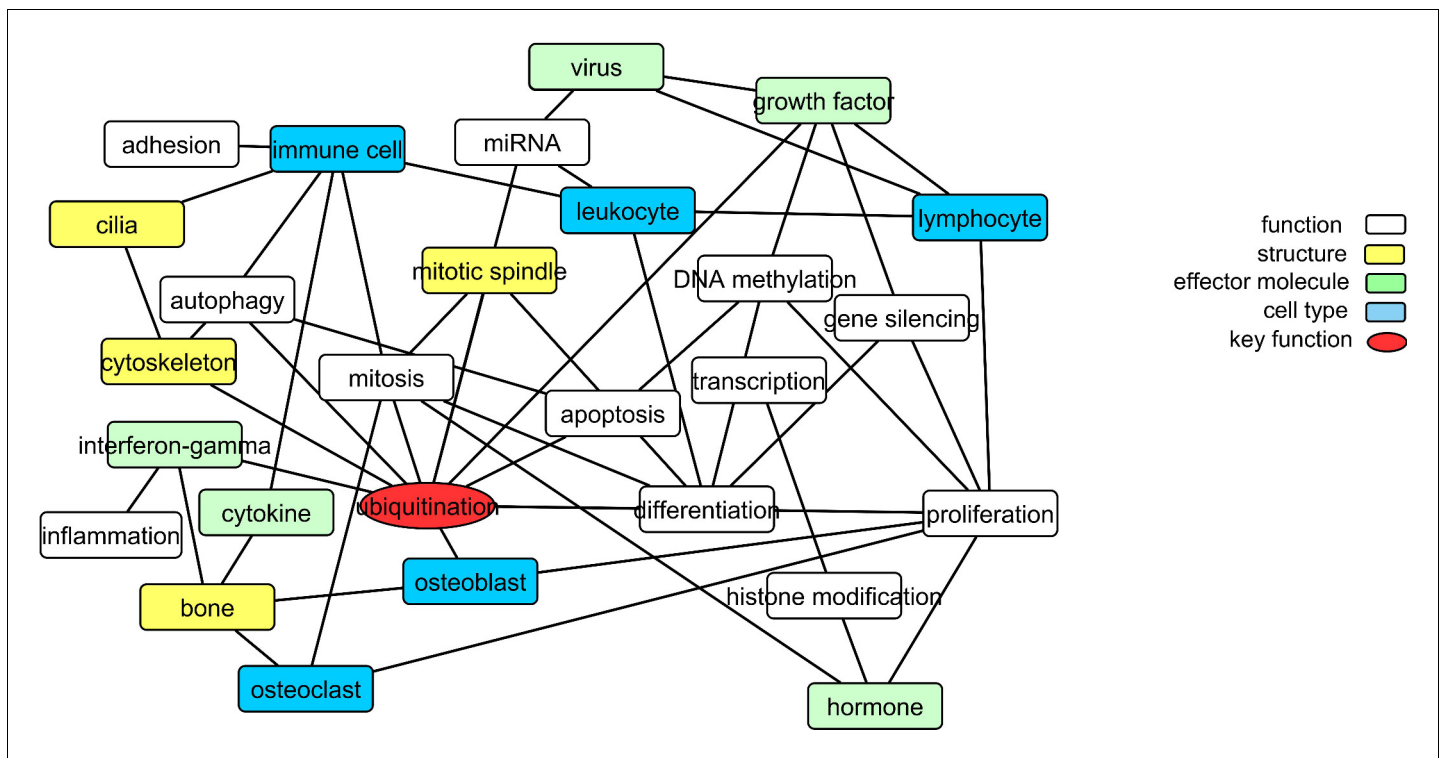
Figure 3. Translating the methylation data into functional networks. Nodes are functional, cellular, molecular, and sub-cellular keywords from GO annotations enriched amongst the *Pooled sites*. An edge between two nodes indicates that differentially methylated genes associated with the keyword in node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

(*Ooi et al., 2014*). It has previously been associated with frontotemporal dementia (*Gijselinck et al., 2015*), which can co-exist with Paget's disease (*Watts et al., 2004*). MOSC2 is a member of the membrane-bound E3 ubiquitin ligase family that regulates endosome trafficking (*Zhang et al., 2018*). 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 (*Kuo et al., 2010*). PENK encodes for proenkephalin, the precursor of a range of effector molecules including pain-associated pentapeptide opioids as well as modulators of osteoblast differentiation (*Seitz et al., 2010*). Interestingly, PENK knockout mice have abnormal bone structure and mineralization (*Dickinson et al., 2016*). 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 (*Nishikawa et al., 2010*). SPATA18 is expressed in a variety of cancers including osteosarcoma, and its transcription is induced by p53 (*Bornstein et al., 2011*). TAL1 has been found to regulate osteoclast differentiation through suppression of their fusion mediator DCSTAMP (*Courtial et al., 2012*). The zinc finger protein ZIC1 has a role in shear flow mechanotransduction in osteocytes (*Kalogeropoulos et al., 2010*). 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 (*Varanasi et al., 2010*). NFYB confers chromatin access to other transcriptional regulators and is known to be involved in transition through cell cycle (*Ly et al., 2013*). Finally, the centrosomal ARHGEF10 has a role in the formation of mitotic spindle during mitosis (*Shibata et al., 2019*).

Our analysis was extended to identify regions with frequent but independent methylation changes in PDB amongst sites that are adjacent to each other. 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 differs from the traditional methods in that enrichment of a region does not stem from frequent occurrences of correlated DMS within the
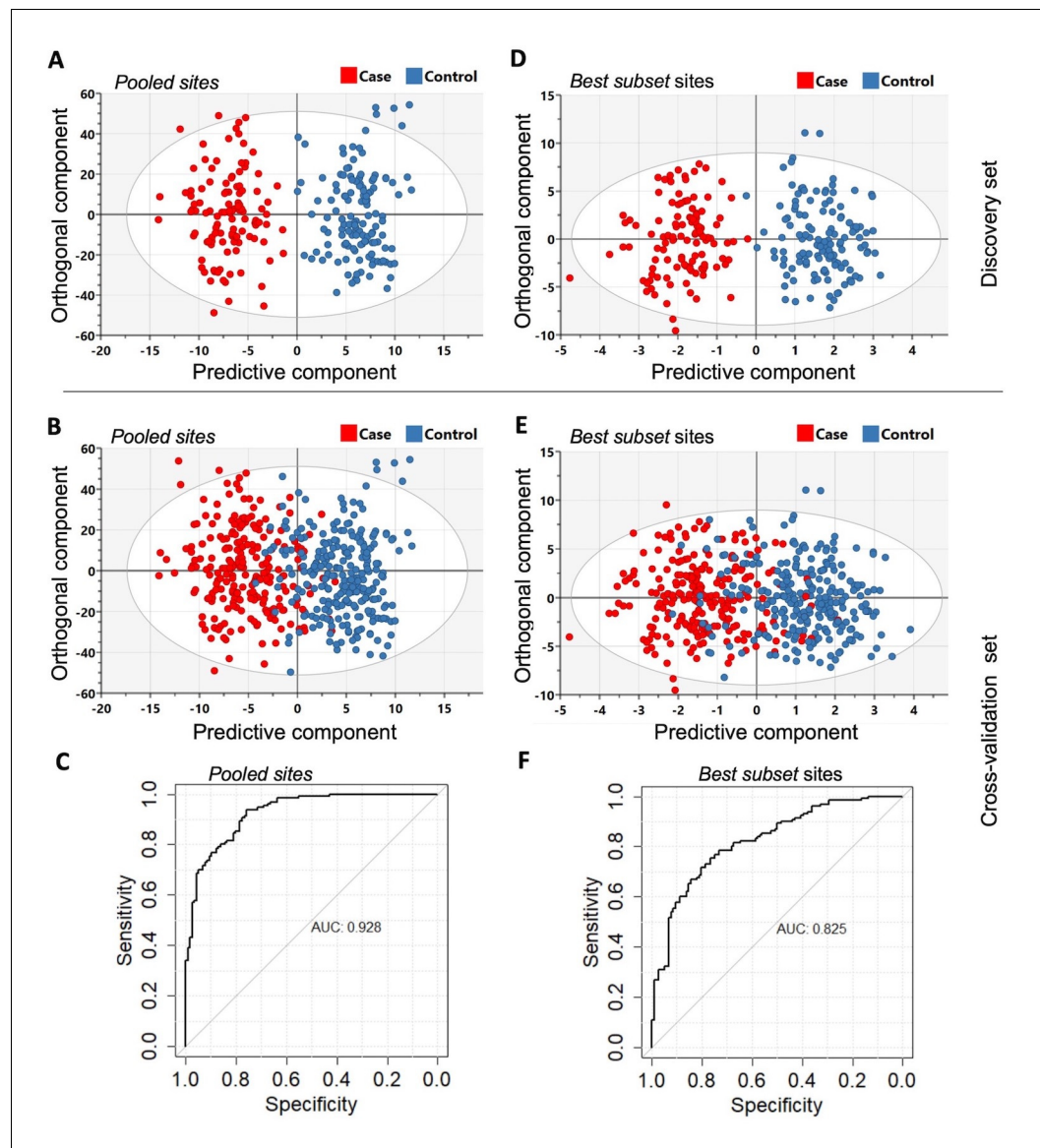
Figure 4. The orthogonal partial least squares-discriminant analysis (OPLS-DA) was performed using the *Pooled sites* identified from the discovery set (n = 246). (**A**) Classifier trained on all 2847 pooled sites with FDR < 0.05 (*Pooled sites*) from the discovery set. (**B**) Testing the classifier on the replication (or cross-validation) set. (**C**) ROC curve analysis yielded an overall sensitivity of 0.84, specificity of 0.81, and AUC of 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 (or cross-validation) set. (**F**) ROC curve analysis showed an overall sensitivity of 0.77, specificity of 0.74, and AUC of 0.825. The Scatter plots 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).

region but rather the accumulation of independent effects from residing sites. In other words, regions with the most, but unique site-level effects are prioritized. By doing so, our approach is advantageous in two ways: First, it 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 neighboring sites whilst avoiding the redundancy of information from site-level analysis.

Four Bonferroni significant DMR were identified in islands, which were located near the following genes: LTB, a cytokine shown to stimulate osteoclast activity; SKIV2L, with an RNA helicase activity,
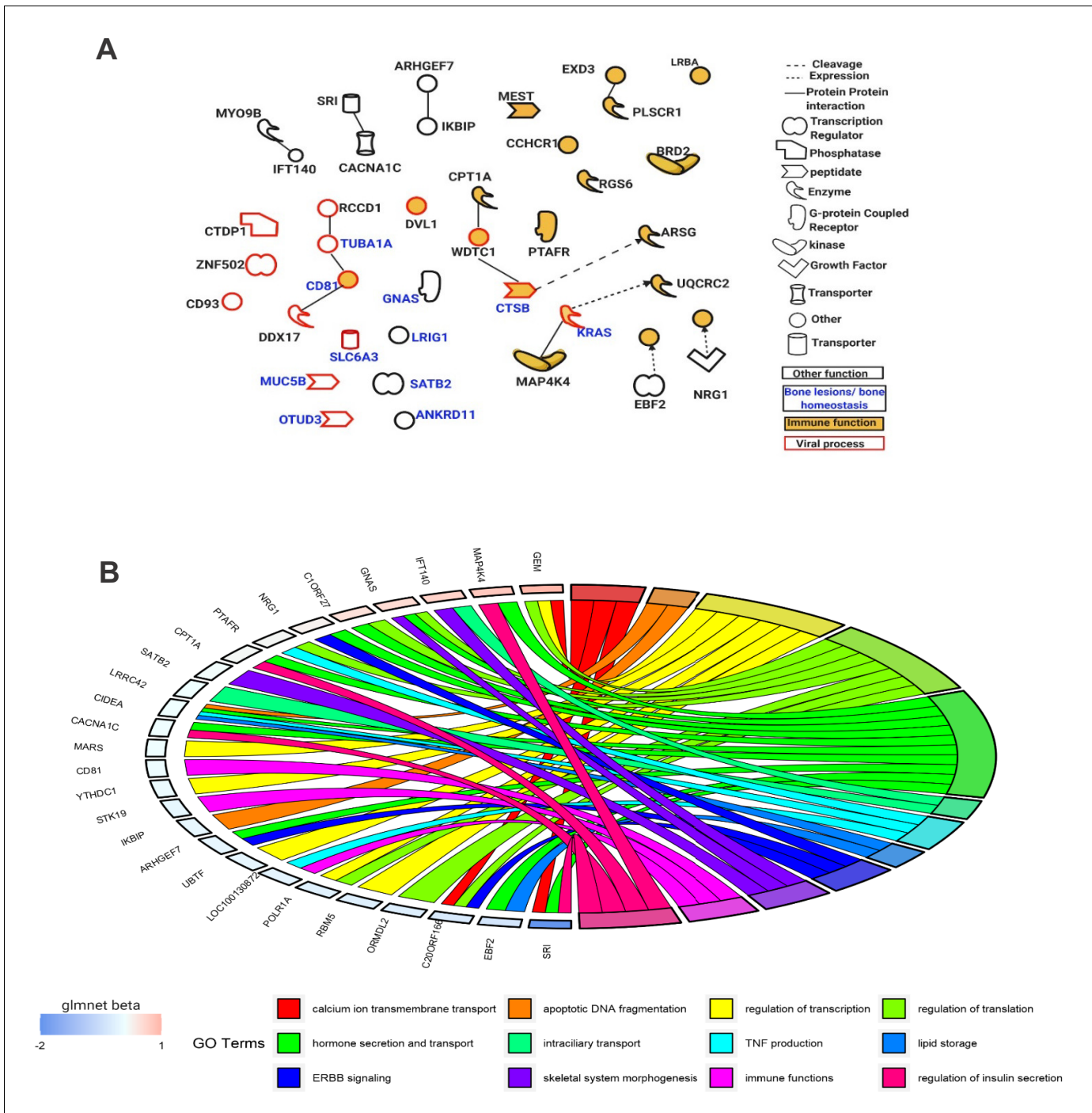
Figure 5. Functions of genes mapped near the *Best subset* of differentially methylated sites identified through the elastic-net regularization extension of the generalized linear model. (A) An IPA-based network showing a subset of these genes with functional interactions (edges) or mapping to one of 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.

thought to be involved in blocking translation of viral mRNA and has been implicated in regulating host responses to viral infections (*Eckard et al., 2014*); EBF3, which is involved in bone development and B cell differentiation (*Seike et al., 2018*); and CCND1, a Wnt target that was reported to be upregulated in response to mechanical loading of bone (*Holguin et al., 2016*).

Additionally, six Bonferroni significant DMR in gene bodies were identified. These were located within genes with functions related to mitosis and ciliogenesis (SDCCAG8) (*Insolera et al., 2014*): TGFB1-mediated signaling (RBPMS) (*Shanmugaapriya et al., 2016*); calcium signaling (CACNA1B) (*Blair et al., 2007*); protein ubiquitination (HSPA13) (*Kaye et al., 2000*); cytoskeletal organization (PARD3B) (*Kohjima et al., 2002*); and histone acetylation (BRD1) (*Mishima et al., 2014*).

The *Pooled* sites identified from the discovery set were able to discriminate cases and controls with a considerable accuracy when tested on the cross-validation 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*, *Supplementary file 5*). 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 (*Visconti et al., 2017*). 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 indicates possible differences between cases and controls in these genes, which are involved in mechanotransduction, a process that has been implicated in localization of bone lesions in PDB (*Gasper, 1979*). 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 (*Numan et al., 2015*) and should prompt further work to investigate host–environment interactions including studies of the microbiome in this complex but fascinating disease (*Ohlsson and Sjögren, 2018*).

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 (*Reid et al., 2011*; *Reid et al., 2005*; *Tan et al., 2017*). Unfortunately, PDB often remains clinically silent until it has reached an advanced stage by which point irreversible skeletal damage may already have occurred (*Gennari et al., 2019*). 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 PDB pathogenesis. This is primarily justified by the difficulty to collect bone tissue from PDB patients in a similarly sized cohort. Nevertheless, on comparison of our Bonferroni significant DMS and DMR (*Tables 2* and *3*), with a published list of highly concordant CpG sites between blood and femur bone tissue collected during hip replacement surgery (*Ebrahimi et al., 2021*), we noted considerable overlap. We found that CpGs annotated to 8 of the 14 genes from our DMS analysis were among the highly correlated sites between blood and bone (*Supplementary file 6*). For DMRs, of the 10 genes reported in our study (*Table 3*), 6 had at least one CpG with high correlation between blood and bone (*Supplementary file 6*). However, 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. Moreover, one needs to consider that blood also contains progenitors of bone cells and that white blood cells share a similar ancestry with osteoclasts.

Although the split-sample approach was meant to allow for validation of the results for increased statistical rigor, our cross-validation dataset is not totally independent from its discovery counterpart in that similar sources of noise and counfounding effects are present in both. However, the total cohort was obtained from a large number of centers across the UK representing most major cities,

which adds to the validity of our overall results. Another limiting aspect of our study was drawing functional relevance of our DMS and DMR by reference to tissue specific eQTMs from the BIOS QTL database, which were originally derived from blood. Therefore, the effects of the differential methylation from our candidates DMS and DMR on gene expression under PDB remain to be investigated. Finally, it is possible that the observed methylation changes reported in this study exist as a consequence of the disease; therefore, further prospective studies assessing their true potential as predictor biomarkers are warranted. Such studies could revolve around recruiting individuals with a genetic predisposition and/or family history of PDB for which the level of methylation of our 95 best subset sites can be routinely assessed. Such epigenetic measurements can then be linked to future disease onset if any, in the presence of appropriate controls.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Other	Infinium Human Methylation450 BeadChip	Illumina, USA		DNA Methylation array
Software, algorithm	RnBeads	R		Version 1.10.8
Software, algorithm	SIMCA	Umetrics, Sweden		Version 15
Software, algorithm	IPA	Qiagen, Germany		
Software, algorithm	GGM	R		Version 2.4
Software, algorithm	topGO	R		Version 2.4

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) (*Tan et al., 2017*). The PRISM 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 cross-validation 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 (*Tsai and Bell, 2015*). 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 cross-validation set.

DNA methylation profiling

Genomic DNA was extracted from peripheral blood using standard protocols. Bisulfite conversion was performed on 500 µg of DNA using Zymo EZ-96 DNA methylation Kit (RRID:SCR_008968, 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 (RRID:SCR_010958) was used for quality control (*Müller et al., 2019*). 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 (*Ambatipudi et al., 2016*). 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 (*Pidsley et al., 2013*) 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 $\log_2((\text{methylated signal} +1)/(\text{unmethylated signal} +1))$, 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 (*Houseman et al., 2012*), part of the *RnBeads* pipeline. The reference methylome was obtained from previously published methylation data measured from sorted blood cells comprising 47 samples (*Reinius et al., 2012*). 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) that 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* (*Müller et al., 2019*).

In all statistical models described below, the term *confounders* refers to the following covariates: age, sex, array, bisulfite conversion batch, array scan batch, blood cell composition from the *Houseman* method (*Houseman et al., 2012*), and the top 10 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 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 cross-validation set. Meta-analysis looking at the combined effect from both discovery and cross-validation was performed on the totality of probes using the R package *Metafor* (RRID:SCR_003450) (*Wolfgang, 2010*). The Bonferroni adjusted genome-wide significance threshold of $p=1.17\times10^{-7}$ (0.05/429,156) was used to identify Bonferroni significant DMS based on the meta-analysis p-values.

Differential methylation analysis of regions

DMR were analyzed using binomial regression, member of the family of the generalized linear models (equivalent to logistic regression), in two steps:

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

$$phenotype \sim confounders \tag{1}$$

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

$$phenotype \sim confounders + CpG\,site_1 + CpG\,site_2 + ... + CpG\,site_n \tag{2}$$

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 cross-validation 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. In other words, the Bonferroni adjustment of regions in the cross-validation was based on the subset of regions found Bonferroni significant in the discovery set. Visualization of the effect of individual sites from selected DMR was conducted using R package coMET (*Martin et al., 2015*).

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 p-values 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 DMRs 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* (RRID:SCR_015505) (*Friedman et al., 2010*), to identify the best subset of discriminatory sites (designated *Best subset*) of the list of *Pooled sites*. We trained an orthogonal projection to latent structure discriminant analysis (OPLS-DA) classifier (*Boccard and Rutledge, 2013*), implemented in the software SIMCA ver. 15 (RRID:SCR_014688, Umetrics, Sweden), on the discovery data from *Pooled* and *Best subset* sites separately. Each model was then tested on the cross-validation set, and its performance was further assessed based on the value from receiver operating characteristic 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 IPA (RRID:SCR_008653, Qiagen, Germany) as well as the GO R package *topGO* (RRID:SCR_014798) (*Alexa, 2020*) 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 (*Giovanni Maria, 2006*). The ggm partial correlations, based on the pooled sites, were used for drawing associations between GO biological process terms found enriched in the same set as follows: First, 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. More specifically, for any two GO terms, we considered the significantly differentially methylated sites from genes associated with either terms. We then tested for enrichments of pairs of sites with significant ggms out of all possible pairs of sites across the two terms. Likewise, Fisher's test p-values<0.05 after FDR multiple testing correction were used to create pairs of functionally related keywords. The software Cytoscape (RRID:SCR_003032) (*Shannon et al., 2003*) was used to visualize these associations.

eQTM analysis

To assess the effect of DNA methylation at CpG sites on the expression of nearby genes, we used data from the BIOS QTL browser (*Bonder et al., 2017*; *Bios QTL, 2021*).

Acknowledgements

Additional information

Competing interests

Stuart H Ralston: has received research funding from Amgen, Eli Lilly, Novartis, and Pfizer unrelated to the submitted work. The author has no other competing interests to declare. The other authors declare that no competing interests exist.

Funding

Author contributions

Ilhame Diboun, Formal analysis, Writing - original draft, Writing - review and editing; Sachin Wani, genotyping; Stuart H Ralston, Resources, Data curation, Writing - review and editing; Omar ME Albagha, Conceptualization, Formal analysis, Supervision, Investigation, Writing - review and editing

Author ORCIDs

Omar ME Albagha (iD) https://orcid.org/0000-0001-5916-5983

Ethics

Human subjects: 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.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.65715.sa1
Author response https://doi.org/10.7554/eLife.65715.sa2

Additional files

Supplementary files

- Supplementary file 1. List of replicated differentially methylated sites with FDR < 0.05.

- Supplementary file 2. List of replicated DMR at islands with FDR < 0.05.
- Supplementary file 3. List of replicated DMR at gene bodies with FDR < 0.05.
- Supplementary file 4. List of replicated DMR at promoters with FDR < 0.05.
- Supplementary file 5. List of Best subset sites from Glmnet analysis.
- Supplementary file 6. List of CpG, reported as correlated between bone and blood in Ebrahimi et al. (PMID: 32692944), mapping to the same genes as our significant DMS and DMR in Pagets disease.
- Supplementary file 7. List of expression quantitative trait-methylation (eQTM) sites from the Pooled sites. Highlighted in bold are the 8CpGs belonging to the best subset of sites (a subset of sites best explanatory of PDB).
- Transparent reporting form

Data availability

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

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Albagha OM, Diboun I, Ralston SH, Wani S	2020	Epigenetic analysis of Paget's disease of bone identifies differentially methylated loci that predict disease status	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163970	NCBI Gene Expression Omnibus, GSE163970

References

Albagha OM, Visconti MR, Alonso N, Langston AL, Cundy T, Dargie R, Dunlop MG, Fraser WD, Hooper MJ, Isaia G, Nicholson GC, del Pino Montes J, Gonzalez-Sarmiento R, di Stefano M, Tenesa A, Walsh JP, Ralston SH. 2010. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. *Nature Genetics* **42**:520–524. DOI: https://doi.org/10.1038/ng.562, PMID: 20436471

Albagha OM, Wani SE, Visconti MR, Alonso N, Goodman K, Brandi ML, Cundy T, Chung PY, Dargie R, Devogelaer JP, Falchetti A, Fraser WD, Gennari L, Gianfrancesco F, Hooper MJ, Van Hul W, Isaia G, Nicholson GC, Nuti R, Papapoulos S, et al. 2011. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. *Nature Genetics* **43**:685–689. DOI: https://doi.org/10.1038/ng.845, PMID: 21623375

Alexa ARJ. 2020. *topGO: Enrichment Analysis for Gene Ontology*. https://rdrr.io/bioc/topGO/

Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, Barrdahl M, Boeing H, Aleksandrova K, Trichopoulou A, Lagiou P, Naska A, Palli D, Krogh V, Polidoro S, Tumino R, Panico S, Bueno-de-Mesquita B, Peeters PH, Quirós JR, et al. 2016. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* **8**:599–618. DOI: https://doi.org/10.2217/epi-2016-0001, PMID: 26864933

Bios QTL. 2021. *Bios QTL*. https://molgenis26.target.rug.nl/downloads/biosqtlbrowser/

Blair HC, Schlesinger PH, Huang CL, Zaidi M. 2007. Calcium signalling and calcium transport in bone disease. *Sub-Cellular Biochemistry* **45**:539–562. DOI: https://doi.org/10.1007/978-1-4020-6191-2_21, PMID: 18193652

Boccard J, Rutledge DN. 2013. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock omics data fusion. *Analytica Chimica Acta* **769**:30–39. DOI: https://doi.org/10.1016/j.aca.2013.01.022, PMID: 23498118

Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, van Iterson M, van Dijk F, van Galen M, Bot J, Slieker RC, Jhamai PM, Verbiest M, Suchiman HE, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Arindrarto W, Kielbasa SM, et al. 2017. Disease variants alter transcription factor levels and methylation of their binding sites. *Nature Genetics* **49**:131–138. DOI: https://doi.org/10.1038/ng.3721, PMID: 27918535

Bornstein C, Brosh R, Molchadsky A, Madar S, Kogan-Sakin I, Goldstein I, Chakravarti D, Flores ER, Goldfinger N, Sarig R, Rotter V. 2011. SPATA18, a spermatogenesis-associated gene, is a novel transcriptional target of p53 and p63. *Molecular and Cellular Biology* **31**:1679–1689. DOI: https://doi.org/10.1128/MCB.01072-10, PMID: 21300779

Corral-Gudino L, Borao-Cengotita-Bengoa M, Del Pino-Montes J, Ralston S. 2013. Epidemiology of Paget's disease of bone: a systematic review and meta-analysis of secular changes. *Bone* **55**:347–352. DOI: https://doi.org/10.1016/j.bone.2013.04.024, PMID: 23643679

Courtial N, Smink JJ, Kuvardina ON, Leutz A, Göthert JR, Lausen J. 2012. Tal1 regulates osteoclast differentiation through suppression of the master regulator of cell fusion *DC-STAMP*. *The FASEB Journal* **26**: 523–532. DOI: https://doi.org/10.1096/fj.11-190850, PMID: 21990371

Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H, Baker CN, Bower L, Brown JM, Caddle LB, Chiani F, Clary D, Cleak J, Daly MJ, Denegre JM, Doe B, et al. 2016. High-throughput discovery of novel developmental phenotypes. *Nature* **537**:508–514. DOI: https://doi.org/10.1038/nature19356, PMID: 27626380

Ebrahimi P, Luthman H, McGuigan FE, Akesson KE. 2021. Epigenome-wide cross-tissue correlation of human bone and blood DNA methylation – can blood be used as a surrogate for bone? *Epigenetics* **16**:92–105. DOI: https://doi.org/10.1080/15592294.2020.1788325

Eckard SC, Rice GI, Fabre A, Badens C, Gray EE, Hartley JL, Crow YJ, Stetson DB. 2014. The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nature Immunology* **15**:839–845. DOI: https://doi.org/10.1038/ni.2948, PMID: 25064072

Friedman J, Hastie T, Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**:1–22. DOI: https://doi.org/10.18637/jss.v033.i01

Gasper TM. 1979. Paget's disease in a treadle machine operator. *BMJ* **1**:1217–1218. DOI: https://doi.org/10.1136/bmj.1.6172.1217-e, PMID: 445028

Gennari L, Rendina D, Falchetti A, Merlotti D. 2019. Paget's Disease of Bone. *Calcified Tissue International* **104**: 483–500. DOI: https://doi.org/10.1007/s00223-019-00522-3, PMID: 30671590

Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Philtjens S, Heeman B, Engelborghs S, Vandenbulcke M, De Baets G, Bäumer V, Cuijt I, Van den Broeck M, Peeters K, Mattheijssens M, Rousseau F, Vandenberghe R, De Jonghe P, Cras P, De Deyn PP, Martin JJ, et al. 2015. Loss of *TBK1* is a frequent cause of frontotemporal dementia in a belgian cohort. *Neurology* **85**:2116–2125. DOI: https://doi.org/10.1212/WNL.0000000000002220, PMID: 26581300

Giovanni Maria M. 2006. Independencies induced from a graphical markov model after marginalization and conditioning: the R package ggm. *Journal of Statistical Software* **15**:v015.i06. DOI: https://doi.org/10.18637/jss.v015.i06

Holguin N, Brodt MD, Silva MJ. 2016. Activation of Wnt Signaling by Mechanical Loading Is Impaired in the Bone of Old Mice. *Journal of Bone and Mineral Research* **31**:2215–2226. DOI: https://doi.org/10.1002/jbmr.2900

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**:86. DOI: https://doi.org/10.1186/1471-2105-13-86, PMID: 22568884

Insolera R, Shao W, Airik R, Hildebrandt F, Shi SH. 2014. SDCCAG8 regulates pericentriolar material recruitment and neuronal migration in the developing cortex. *Neuron* **83**:805–822. DOI: https://doi.org/10.1016/j.neuron.2014.06.029, PMID: 25088364

Kalogeropoulos M, Varanasi SS, Olstad OK, Sanderson P, Gautvik VT, Reppe S, Francis RM, Gautvik KM, Birch MA, Datta HK. 2010. Zic1 transcription factor in bone: neural developmental protein regulates mechanotransduction in osteocytes. *The FASEB Journal* **24**:2893–2903. DOI: https://doi.org/10.1096/fj.09-148908, PMID: 20354137

Kaye FJ, Modi S, Ivanovska I, Koonin EV, Thress K, Kubo A, Kornbluth S, Rose MD. 2000. A family of ubiquitin-like proteins binds the ATPase domain of Hsp70-like stch. *FEBS Letters* **467**:348–355. DOI: https://doi.org/10.1016/S0014-5793(00)01135-2, PMID: 10675567

Kohjima M, Noda Y, Takeya R, Saito N, Takeuchi K, Sumimoto H. 2002. PAR3beta, a novel homologue of the cell polarity protein PAR3, localizes to tight junctions. *Biochemical and Biophysical Research Communications* **299**: 641–646. DOI: https://doi.org/10.1016/S0006-291X(02)02698-0, PMID: 12459187

Kuo WJ, Digman MA, Lander AD. 2010. Heparan sulfate acts as a bone morphogenetic protein coreceptor by facilitating ligand-induced receptor hetero-oligomerization. *Molecular Biology of the Cell* **21**:4028–4041. DOI: https://doi.org/10.1091/mbc.e10-04-0348, PMID: 20861306

Kurihara N, Zhou H, Reddy SV, Garcia Palacios V, Subler MA, Dempster DW, Windle JJ, Roodman GD. 2006. Expression of measles virus nucleocapsid protein in osteoclasts induces Paget's disease-like bone lesions in mice. *Journal of Bone and Mineral Research* **21**:446–455. DOI: https://doi.org/10.1359/JBMR.051108, PMID: 16491293

Kurihara N, Hiruma Y, Yamana K, Michou L, Rousseau C, Morissette J, Galson DL, Teramachi J, Zhou H, Dempster DW, Windle JJ, Brown JP, Roodman GD. 2011. Contributions of the measles virus nucleocapsid gene and the SQSTM1/p62(P392L) mutation to Paget's disease. *Cell Metabolism* **13**:23–34. DOI: https://doi.org/10.1016/j.cmet.2010.12.002, PMID: 21195346

Langston AL, Campbell MK, Fraser WD, MacLennan GS, Selby PL, Ralston SH, PRISM Trial Group. 2010. Randomized trial of intensive bisphosphonate treatment versus symptomatic management in Paget's disease of bone. *Journal of Bone and Mineral Research* **25**:20–31. DOI: https://doi.org/10.1359/jbmr.090709, PMID: 19580457

Ly LL, Yoshida H, Yamaguchi M. 2013. Nuclear transcription factor Y and its roles in cellular processes related to human disease. *American Journal of Cancer Research* **3**:339–346. PMID: 23977444

Martin TC, Yet I, Tsai PC, Bell JT. 2015. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics* **16**:131. DOI: https://doi.org/10.1186/s12859-015-0568-2, PMID: 25928765

Mishima Y, Wang C, Miyagi S, Saraya A, Hosokawa H, Mochizuki-Kashio M, Nakajima-Takagi Y, Koide S, Negishi M, Sashida G, Naito T, Ishikura T, Onodera A, Nakayama T, Tenen DG, Yamaguchi N, Koseki H, Taniuchi I,

Iwama A. 2014. Histone acetylation mediated by Brd1 is crucial for Cd8 gene activation during early thymocyte development. *Nature Communications* **5**:5872. DOI: https://doi.org/10.1038/ncomms6872, PMID: 25519988

Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, Bock C. 2019. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biology* **20**:55. DOI: https://doi.org/10.1186/s13059-019-1664-9, PMID: 30871603

Nishikawa K, Nakashima T, Takeda S, Isogai M, Hamada M, Kimura A, Kodama T, Yamaguchi A, Owen MJ, Takahashi S, Takayanagi H. 2010. Maf promotes osteoblast differentiation in mice by mediating the age-related switch in mesenchymal cell differentiation. *Journal of Clinical Investigation* **120**:3455–3465. DOI: https://doi.org/10.1172/JCI42528

Numan MS, Amiable N, Brown JP, Michou L. 2015. Paget's disease of bone: an osteoimmunological disorder? *Drug Design, Development and Therapy* **9**:4695–4707. DOI: https://doi.org/10.2147/DDDT.S88845, PMID: 26316708

Obaid R, Wani SE, Azfer A, Hurd T, Jones R, Cohen P, Ralston SH, Albagha OME. 2015. Optineurin negatively regulates osteoclast differentiation by modulating NF-κB and interferon signaling: implications for Paget's Disease. *Cell Reports* **13**:1096–1102. DOI: https://doi.org/10.1016/j.celrep.2015.09.071, PMID: 26527009

Ohlsson C, Sjögren K. 2018. Osteomicrobiology: a new Cross-Disciplinary research field. *Calcified Tissue International* **102**:426–432. DOI: https://doi.org/10.1007/s00223-017-0336-6, PMID: 29079994

Ooi EL, Chan ST, Cho NE, Wilkins C, Woodward J, Li M, Kikkawa U, Tellinghuisen T, Gale M, Saito T. 2014. Novel antiviral host factor, TNK1, regulates IFN signaling through serine phosphorylation of STAT1. *PNAS* **111**: 1909–1914. DOI: https://doi.org/10.1073/pnas.1314268111, PMID: 24449862

Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**:293. DOI: https://doi.org/10.1186/1471-2164-14-293, PMID: 23631413

Ralston SH, Corral-Gudino L, Cooper C, Francis RM, Fraser WD, Gennari L, Guañabens N, Javaid MK, Layfield R, O'Neill TW, Russell RGG, Stone MD, Simpson K, Wilkinson D, Wills R, Zillikens MC, Tuck SP. 2019. Diagnosis and management of Paget's Disease of Bone in Adults: A Clinical Guideline. *Journal of Bone and Mineral Research* **34**:579–604. DOI: https://doi.org/10.1002/jbmr.3657, PMID: 30803025

Ralston SH, Albagha OM. 2014. Genetics of Paget's disease of bone. *Current Osteoporosis Reports* **12**:263–271. DOI: https://doi.org/10.1007/s11914-014-0219-y, PMID: 24988994

Reid IR, Miller P, Lyles K, Fraser W, Brown JP, Saidi Y, Mesenbrink P, Su G, Pak J, Zelenakas K, Luchi M, Richardson P, Hosking D. 2005. Comparison of a single infusion of zoledronic acid with risedronate for Paget's Disease. *New England Journal of Medicine* **353**:898–908. DOI: https://doi.org/10.1056/NEJMoa044241

Reid IR, Lyles K, Su G, Brown JP, Walsh JP, del Pino-Montes J, Miller PD, Fraser WD, Cafoncelli S, Bucci-Rechtweg C, Hosking DJ. 2011. A single infusion of zoledronic acid produces sustained remissions in paget disease: data to 6.5 years. *Journal of Bone and Mineral Research* **26**:2261–2270. DOI: https://doi.org/10.1002/jbmr.438, PMID: 21638319

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, Söderhäll C, Scheynius A, Kere J. 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLOS ONE* **7**:e41361. DOI: https://doi.org/10.1371/journal.pone.0041361, PMID: 22848472

Seike M, Omatsu Y, Watanabe H, Kondoh G, Nagasawa T. 2018. Stem cell niche-specific Ebf3 maintains the bone marrow cavity. *Genes & Development* **32**:359–372. DOI: https://doi.org/10.1101/gad.311068.117, PMID: 29563184

Seitz S, Barvencik F, Gebauer M, Albers J, Schulze J, Streichert T, Amling M, Schinke T. 2010. Preproenkephalin (Penk) is expressed in differentiated osteoblasts, and its deletion in hyp mice partially rescues their bone mineralization defect. *Calcified Tissue International* **86**:282–293. DOI: https://doi.org/10.1007/s00223-010-9344-5, PMID: 20204609

Shanmugaapriya S, van Caam A, de Kroon L, Vitters EL, Walgreen B, van Beuningen H, Davidson EB, van der Kraan PM. 2016. Expression of TGF-β signaling regulator RBPMS (RNA-Binding protein with multiple splicing) Is regulated by IL-1β and TGF-β superfamily members, and decreased in aged and osteoarthritic cartilage. *Cartilage* **7**:333–345. DOI: https://doi.org/10.1177/1947603515623991, PMID: 27688842

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504. DOI: https://doi.org/10.1101/gr.1239303, PMID: 14597658

Shibata S, Teshima Y, Niimi K, Inagaki S. 2019. Involvement of ARHGEF10, GEF for RhoA, in Rab6/Rab8-mediating membrane traffic. *Small GTPases* **10**:169–177. DOI: https://doi.org/10.1080/21541248.2017.1302550, PMID: 28448737

Tan A, Goodman K, Walker A, Hudson J, MacLennan GS, Selby PL, Fraser WD, Ralston SH, PRISM-EZ Trial Group. 2017. Long-Term randomized trial of intensive versus symptomatic management in Paget's Disease of Bone: The PRISM-EZ Study. *Journal of Bone and Mineral Research* **32**:1165–1173. DOI: https://doi.org/10.1002/jbmr.3066, PMID: 28176386

Tan A, Ralston SH. 2014. Paget's disease of bone. *QJM* **107**:865–869. DOI: https://doi.org/10.1093/qjmed/hcu075, PMID: 24757095

Tsai PC, Bell JT. 2015. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *International Journal of Epidemiology* **44**:1429–1441. DOI: https://doi.org/10.1093/ije/dyv041, PMID: 25972603

Vallet M, Soares DC, Wani S, Sophocleous A, Warner J, Salter DM, Ralston SH, Albagha OM. 2015. 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. *Human Molecular Genetics* **24**:3286–3295. DOI: https://doi.org/10.1093/hmg/ddv068, PMID: 25701875

Varanasi SS, Olstad OK, Swan DC, Sanderson P, Gautvik VT, Reppe S, Francis RM, Gautvik KM, Datta HK. 2010. Skeletal site-related variation in human trabecular bone transcriptome and signaling. *PLOS ONE* **5**:e10692. DOI: https://doi.org/10.1371/journal.pone.0010692, PMID: 20502692

Visconti MR, Usategui-Martín R, Ralston SH. 2017. Antibody response to paramyxoviruses in Paget's Disease of Bone. *Calcified Tissue International* **101**:141–147. DOI: https://doi.org/10.1007/s00223-017-0265-4, PMID: 28361207

Watts GD, Wymer J, Kovach MJ, Mehta SG, Mumm S, Darvish D, Pestronk A, Whyte MP, Kimonis VE. 2004. Inclusion body myopathy associated with paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature Genetics* **36**:377–381. DOI: https://doi.org/10.1038/ng1332, PMID: 15034582

Wolfgang V. 2010. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**: 48. DOI: https://doi.org/10.18637/jss.v036.i03

Wong SW, Huang BW, Hu X, Ho Kim E, Kolb JP, Padilla RJ, Xue P, Wang L, Oguin TH, Miguez PA, Tseng HC, Ko CC, Martinez J. 2020. Global deletion of optineurin results in altered type I IFN signaling and abnormal bone remodeling in a model of paget's disease. *Cell Death & Differentiation* **27**:71–84. DOI: https://doi.org/10.1038/s41418-019-0341-6, PMID: 31076632

Zhang Y, Lu J, Liu X. 2018. MARCH2 is upregulated in HIV-1 infection and inhibits HIV-1 production through envelope protein translocation or degradation. *Virology* **518**:293–300. DOI: https://doi.org/10.1016/j.virol.2018.02.003, PMID: 29573664