Genome-wide association identifies three new susceptibility loci for Paget's disease of bone

Please cite this article: Albagha, O. M. E., Wani, S. E., Visconti, M. R., Alonso, N., Goodman, K., Brandi, M. L., Cundy, T., Chung, P. Y. J., Dargie, R., Devogelaer, J-P., Falchetti, A., Fraser, W. D., Gennari, L., Gianfrancesco, F., Hooper, M. J., Van Hul, W., Isaia, G., Nicholson, G. C., Nuti, R., Papapoulos, S., del Pino Montes, J., Ratajczak, T., Rea, S. L., Rendina, D., Gonzalez-Sarmiento, R., Di Stefano, M., Ward, L. C., Walsh, J. P., Ralston, S. H., & Genetic Determinants Pagets Dis (2011). Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. *Nature Genetics*, DOI: 10.1038/ng.845

LETTER

Genome-wide association identifies three new susceptibility loci for Paget's disease of bone

Omar ME Albagha[1], Sachin Wani[1], Micaela R Visconti[1], Nerea Alonso[1], Kirsteen Goodman[2], Maria Luisa Brandi[3], Tim Cundy[4], Pui Yan Jenny Chung[5], Rosemary Dargie[6], Jean-Pierre Devogelaer[7], Alberto Falchetti[3], William D Fraser[8], Luigi Gennari[9], Fernando Gianfrancesco[10], Michael J Hooper[11], Wim Van Hul[5], Gianluca Isaia[12], Geoff C Nicholson[13], Ranuccio Nuti[9], Socrates Papapoulos[14], Javier del Pino Montes[15], Thomas Ratajczak[16], Sarah L Rea[16], Domenico Rendina[17], Rogelio Gonzalez-Sarmiento[15], Marco di Stefano[12], Lynley C Ward[16], John P Walsh[16], & Stuart H Ralston[1,2] for the Genetic Determinants of Paget's Disease (GDPD) Consortium

[1]Rheumatic Diseases Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK; [2]Edinburgh Clinical Trials Unit, University of Edinburgh, Western General Hospital, Edinburgh UK; [3]Department of Internal Medicine, University of Florence, Florence, Italy; [4]Department of Medicine, University of Auckland, Auckland, New Zealand; [5]Department of Medical Genetics, University of Antwerp, Antwerp, Belgium, [6]University Department of Medicine, Glasgow Royal Infirmary, Glasgow, UK; [7]Department of Rheumatology, Saint-Luc University Hospital, Université Catholique de Louvain, Brussels, Belgium, [8]Department of Clinical Chemistry, Royal Liverpool University Hospital, Liverpool, UK; [9]Department of Internal Medicine, Endocrine Metabolic Sciences and Biochemistry, University of Siena, Siena, Italy; [10]Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", Italian National Research Council, Naples, Italy; [11]Department of Medicine, University of Sydney, Sydney, Australia; [12]Medical and Surgical Department, Geriatric Section, University of Torino, Torino, Italy; [13]Department of Clinical and Biomedical Sciences, Barwon Health, Geelong Hospital, University of Melbourne, Melbourne, Australia; [14]Department of Endocrinology & Metabolic Diseases, Leiden University Medical Centre, Leiden, The Netherlands; [15]Unidad de Medicina Molecular, Departamento de Medicina, Universidad de Salamanca, Servicio de Reumatología, Hospital Universitario de Salamanca, RETICEF, Salamanca, Spain; [16]Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, and School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia; [17]Department of Clinical and Experimental Medicine Federico II University of Naples, Naples, Italy.

Correspondence to: Dr Omar M E Albagha, PhD (omar.albagha@ed.ac.uk) or Professor Stuart H Ralston MD, (stuart.ralston@ed.ac.uk) Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK Phone: +44-131-651-1022; Fax: +44+44-131-651-1085.

Paget's disease of bone (PDB) is a common disorder characterised by focal abnormalities of bone remodelling. We previously identified variants at the *CSF1*, *OPTN* and *TNFRSF11A* loci as risk factors for PDB by genome wide association study[1]. Here, we extended this study, identified three new loci and confirmed their association with PDB in 2,215 cases and 4,370 controls from seven independent populations. The new associations were with rs5742915 within *PML* on 15q24 (OR=1.34; $P = 1.6 \times 10^{-14}$); rs10498635 within *RIN3* on 14q32 (OR= 1.44; $P = 2.55 \times 10^{-11}$) and rs4294134 within *NUP205* on 7q33 (OR=1.45; $P = 8.45 \times 10^{-10}$). Our data also confirmed the association of *TM7SF4* (rs2458413; OR=1.40; $P = 7.38 \times 10^{-17}$) with PDB. The seven loci explained a substantial proportion of familial risk of PDB (~13%). These studies provide new insights into the genetic architecture and pathophysiology of PDB.

Paget's disease of bone (PDB) is a common skeletal disorder with a strong genetic component that affects up to 2% of individuals of European ancestry aged 55 years and above[2,3]. Mutations of the *SQSTM1* gene are known to cause a high penetrance form of PDB which is clinically severe[4] and occurs in about 40% of patients with a family history of the disorder[5,6]. We recently identified additional susceptibility alleles for PDB at the *CSF1, OPTN,* and *TNFRSF11A* loci by a genome wide association study (GWAS) involving 692 PDB cases and 1,001 controls with replication cohort of 481 cases and 520 controls[1]. In order to identify additional susceptibility loci for the disease, we performed an extended GWAS involving a total of 749 PDB cases of British descent in whom *SQSTM1* mutations had been excluded and 2,930 British controls derived from the 1958 Birth Cohort[7] with replication in a further 1,474 cases and 1,671 controls from six independent populations.

After applying quality control measures and excluding samples of non-European ancestry, the extended cohort (henceforth referred to as the GWAS stage) comprised 741 cases and 2,699 controls with genotype information for 290,115 SNPs, providing a 4-fold increase in power to detect loci of moderate effect size (odds ratio ≥ 1.4) compared with our previous study[1]. To increase SNP coverage, we performed genome wide SNP imputation for the GWAS stage samples using phased haplotype data from the HapMap project as a reference. The results of association testing of genotyped and imputed SNPs (total 2,487,078 SNPs) from the GWAS stage are shown in **Fig. 1**. A locus on chromosomes 8q22.3 showed genome-wide evidence of association with PDB ($P < 5.0 \times 10^{-8}$) in addition to the previously identified genome wide significant loci on 1p13.3, 10p13 and 18q21.33[1].

In the second stage of this study we analysed the highest ranking SNPs observed in the GWAS stage (*P* values of 5×10^{-5} or less) for replication after excluding those in linkage disequilibrium (LD; $r^2 > 0.8$ or $D' > 0.95$) with the highest ranking SNP from each region. A

total of 27 SNPs were genotyped in the replication cohorts which consisted of 1,474 *SQSTM1*-negative PDB cases from six different geographic regions and 1,671 unaffected controls from the same regions that were matched with the cases by gender as described in the online methods section and **Supplementary Table 1**. A meta-analysis of data from the GWAS stage and individual replication cohorts was performed and the results are summarised in **Supplementary Table 2**. This strengthened the association with PDB for the *CSF1, OPTN,* and *TNFRSF11A* loci which were identified in our previous study[1] and confirmed the association with 8q22.3 locus which was suggestively associated with PDB in our previous GWAS[1] and was confirmed to be associated with PDB in a small study of Belgian and Dutch subjects[8]. Furthermore, three additional genome wide significant loci on 7q33, 14q32.12, and 15q24.1 were identified in the combined data set ($P < 5 \times 10^{-8}$; **Table 1 and Fig. 2**).

The strongest signal on 8q22.3 was with rs2458413 (combined *P*-value = 7.38×10^{-17}; OR = 1.4). There was no significant heterogeneity between the study groups (**Table 1, Fig. 3 and Supplementary Table 3**) and the direction of association was similar in all cohorts. The associated region spans ~220kb but SNPs with the highest association signal appear to cluster within an 18-kb LD block spanning the entire Transmembrane 7 superfamily member 4 gene (*TM7SF4;* **Fig. 2 and Supplementary Fig. 1**). This gene encodes dendritic cell-specific transmembrane protein (DC-STAMP)[9] which is a strong functional candidate gene for PDB since it is required for the fusion of osteoclast precursors to form mature osteoclasts[10]. Previous studies have shown that RANKL induced DC-STAMP expression is essential for osteoclast formation[11] and a recent study showed that the connective tissue growth factor CCN2 stimulates osteoclast fusion through interaction with DC-STAMP[12]. Since osteoclasts from patients with PDB are larger in size and contain more nuclei than normal osteoclasts, it seems likely that the genetic variants that predispose to PDB do so by enhancing *TM7SF4* expression or by causing gain-of-function at the protein level but further studies will be required to investigate these possibilities.

The first new locus for PDB susceptibility was on 7q33 tagged by rs4294134 (combined *P*-value = 8.45×10^{-10}; OR = 1.45). The direction of association was similar in all study cohorts and analysis of the combined data set showed no evidence for heterogeneity between study groups (**Table 1, Fig. 3 and Supplementary Table 3**). The associated region spans ~350kb (**Fig. 2**) but the strongest signal was with rs4294134, located within the 22nd intron of *NUP205*. This gene encodes nucleoporin 205kDa which is one of the main components of the nuclear pore complex involved in the regulation of transport between the cytoplasm and nucleus[13]. All SNPs with $P < 1 \times 10^{-5}$ in the 350kb associated region were in moderate to strong LD with rs4294134 ($r^2 \geq 0.5$; $D' \geq 0.95$) with the exception of two SNPs (rs3110788

and rs3110794) which were poorly correlated with rs4294134 ($r^2 \leq 0.21$; $D' \geq 0.95$; **Fig. 2**). Conditional analysis in the GWAS stage indicated that the association signal appeared to be driven by rs4294134 ($P = 8.8 \times 10^{-3}$) after adjusting for rs3110788 ($P = 0.31$) and rs3110794 ($P = 0.10$). None of the genes located in this region are known to affect bone metabolism and further studies will be required to identify the functional variant(s) responsible for association with PDB.

The second new susceptibility locus was located on 14q32.12 and was tagged by rs10498635. This SNP showed borderline evidence of association with PDB in our previous study ($P = 9.69 \times 10^{-8}$)[1] but reached genome-wide significance in the present study (combined P-value = 2.55×10^{-11}; OR = 1.44). Association testing showed no evidence for heterogeneity between the study groups (**Table 1, Fig. 3 and Supplementary Table 3).** The 62kb-associated region is bounded by two recombination hotspots and contains the gene *RIN3* (**Fig. 2**) that encodes the Ras and Rab interactor 3, a protein that plays a role in vesicular trafficking through interaction with small GTPases such as Ras and Rab[14,15]. The function of *RIN3* in bone metabolism is currently unknown, but it could play a role in bone resorption in view of the importance that small GTPases play in vesicular trafficking and in osteoclast function[16,17]. It is of interest to note that mutations affecting the VCP, a protein also involved in vesicular trafficking, cause the syndrome of inclusion body myopathy with early-onset Paget's disease and frontotemporal dementia (IBMPFD)[18].

The third new susceptibility locus was located on 15q24.1 and the strongest association was with rs5742915 (combined P-value = 1.60×10^{-14}; OR = 1.34; **Table 1, Fig. 3 and Supplementary Table 3**). The associated region is bounded by two recombination hot spots and spans ~200kb but a gap spanning ~40kb was observed in this region with no SNP coverage in the illumina arrays or the HapMap CEU population. The associated SNPs were clustered within the promyelocytic leukaemia gene (*PML*; **Fig. 2**) and the strongest signal was observed for rs5742915, which results in a phenylalanine to leucine amino acid change at codon 645 (F645L) of the PML protein. The function of *PML* in bone metabolism is unclear but it is known to be involved in TGF-β signalling[19]. Accordingly Lin *et al* showed that cells from *pml* knock out mice were resistant to TGF-β-dependent growth arrest and apoptosis and had impaired induction of TGF-β target genes[19]. Since TGF-β is known to play a role in the regulation of bone remodelling, it is possible that the association between PDB and *PML* could be mediated by an effect on TGF-β signalling, but further research will be required to investigate this possibility. The *GOLGA6A* gene is also located in the associated region and encodes a protein that belongs to golgin, a family of coiled-coil proteins associated with the Golgi apparatus and play a role in membrane fusion and as structural supports for the Golgi

cisternae. This gene is located in the 40kb gap region that contains a large low-copy repeat sequence. Although the *GOLGA6A* has no known role in bone metabolism, mutations in other members of the golgin family have been shown to cause a lethal skeletal dysplasia[20], and a severe form of osteoporosis[21].

We were also able to replicate our previously reported association between variants at the *CSF1, OPTN*, and *TNFRSF11A* loci and PDB in the present study[1]. The results of meta-analysis of the combined data set for these loci are shown in **Table 1** and **Supplementary Fig. 2** which provide conclusive evidence for association of variants at *CSF1* ($P = 7.06 \times 10^{-35}$), *OPTN* ($P = 4.37 \times 10^{-38}$), and *TNFRSF11A* ($P = 7.98 \times 10^{-21}$) with PDB. Evidence of heterogeneity between study groups was observed for rs1561570 ($I^2 = 65.7\%$; $P_{het} = 0.01$) at *OPTN* but this was due to differences in effect size rather than the direction of effect and the association remained genome wide significant after accounting for heterogeneity ($P = 4.34 \times 10^{-12}$; OR = 1.68). The heterogeneity was caused by larger effect size observed in the Dutch cohort (**Supplementary Fig. 2**) possibly due to the small sample size of this cohort. These observations provide highly robust evidence for association between these loci and PDB and extend those recently reported[8] in the Dutch and Belgian populations which were also included in the present study.

We next wanted to determine if the identified loci on 15q24.1, 7q33 and 14q32.12 interacted with each other or with the previously identified loci on 1p13.3, 8q22.3, 10p13 and 18q21.33 to affect the risk of PDB. Pair-wise interaction analysis showed weak evidence for interaction between 7q33 (rs4294134) with 8q22.3 (rs2458413; $P = 0.03$) and 10p13 (rs1561570; $P = 0.02$). However, these interactions were not significant after adjusting for multiple testing and none of the other loci showed evidence for interaction ($P > 0.05$) suggesting a multiplicative model of association with PDB risk. In order to estimate the effect size of the identified loci on the development of PDB, we calculated the proportion of familial risk explained by the genome wide significant loci in the replication sample assuming a sibling relative risk for PDB of 7.0[22]. This showed that the proportion of familial risk explained was ~13% which is much greater than observed for other bone diseases like osteoporosis[23]. We also estimated the cumulative population attributable risk of these loci in the replication cohort and found it to be 86% and we found that the risk of PDB increased with increasing number of risk allele scores defined by the seven loci ($OR_{per\text{-}riskallele} = 1.44$, 95% CI = 1.38 – 1.51, $P = 5.4 \times 10^{-57}$). When allele scores were weighted according to their estimated effect size we found that subjects in the top 10% of the allele score distribution (D10; n = 315) had 10.1 fold (95%CI; 7.0 – 14.6; $P = 2.4 \times 10^{-39}$) increase in risk of developing PDB compared to those in the bottom 10% of the distribution (D1; n = 315) from

the replication dataset (**Fig.4**). Although these data suggest that a large part of the genetic risk of PDB in patients without *SQSTM1* mutations is accounted for by these loci, we acknowledge that the functional variants need to be identified before we can precisely estimate the contribution that these loci make to the risk of developing PDB. To assess the functional effect of the identified SNPs on gene expression, we tested the association between top PDB-associated SNPs (or those in LD; D'≥0.8) from each of the seven loci and cis-allelic expression of genes located in the associated regions using publicly available expression quantitative trait loci (eQTL) data. This showed highly significant associations for transcripts of *TM7SF4* (rs2458415; expression *P*-value = 1.22×10^{-18}) and *OPTN* (rs1561570; expression *P*-value = 6.61×10^{-62}) in peripheral blood monocytes[24] suggesting that the association with PDB risk for these loci could be mediated by influencing gene expression levels.

In addition to the loci mentioned above, additional variants were identified that showed suggestive evidence for association with PDB. For example a locus on chromosome Xq24 showed borderline evidence for association with PDB (rs5910578 within *SLC25A43* gene; combined $P = 1.26 \times 10^{-7}$; OR = 1.34) as did another locus on chromosome 6p22.3 (rs1341239 near *PRL* gene; combined $P = 3.83 \times 10^{-6}$; OR= 1.20; **Supplementary Table 2**). Given that we observed 6 genotyped variants with $P < 1 \times 10^{-5}$ in the GWAS stage after removal of confirmed SNPs and associated variants when we only expect 3 by chance (**Supplementary Fig. 3**), it is likely that some of the associations observed are true but our study was not sufficiently powered to detect them at a genome wide significance level ($P < 5 \times 10^{-8}$).

This study has been successful in identifying seven loci that contribute substantially to the risk of developing PDB. The identified loci have relatively large effect sizes compared with other common diseases such as osteoporosis and rheumatoid arthritis. This indicates that susceptibility to PDB is most probably mediated by inheritance of a relatively small number of genes with large effect sizes as opposed to a large number of genes with small effect sizes as seen in other complex diseases. Many of the susceptibility variants lie within or close to genes that are known to play important roles in regulating osteoclast differentiation and function whereas other variants lie within genes not previously implicated in the regulation of bone metabolism. Whilst further work will be required to identify the functional variants, the present study has provided new insights into the genetic architecture of PDB and has identified several genes that previously were not suspected to play a role in bone metabolism. Finally, the large effect size of the variants identified means that it may be possible in the future to identify people at risk of developing PDB by genetic profiling.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of the many participants who provided samples for the analysis. Dr Lee Murphy and Ms Angie Fawkes of the Wellcome Trust Clinical Research Facility for technical support with the Illumina genotyping. Mrs Aman Khatib for her assistance in data management. The study was supported in part by grants to SHR from the Arthritis Research Campaign (grants 13724, 17646 & 15389) and a grant to OMEA and SHR from the Paget's Association. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk and funding for the project was provided by the Wellcome Trust under award 076113 and 085475. We thank other members of the GDPD consortium for sample collection and clinical phenotyping of the **Italian cohorts**: Silvano Adami and Maurizio Rossini, University of Verona, Vellegio Hospital, Verona, Italy; Maurizio Benucci, Rheumatology Unit, Nuovo Ospedale, Florence, Italy; Simona Bergui and Giancarlo Isaia, Department of Internal Medicine, University of Torino, Torino, Italy; Daniela Merlotti, Department of Internal Medicine, University of Siena, Siena, Italy; Ombretta Di Munno, Rheumatology Unit, University of Pisa, Pisa, Italy; Sergio Ortolani, Centre for Metabolic Bone Disease, Instituto Auxologico Italiano, Milan, Italy; Massimo Fabio Uliviera, Ospedale Maggiore, Milan, Italy; Francesca Torricelli, Unit of Genetic Diagnosis, Carregi Hospital, Florence, Italy; Giuseppe Mossetti and Antonio Del Puento, Federico II, University of Naples, Naples, Italy. For the **Belgian and Dutch cohorts**: Steven Boonen, Department of Experimental Medicine, Leuven University, Leuven, Belgium; Stefan Goemaere & Hans-Georg Zmierczak, Unit for Osteoporosis and Metabolic Bone Diseases, Ghent University Hospital, Ghent, Belgium; Piet Geusens, Biomedical Research Unit, University of Hasselt, Diepenbeek, Belgium, & Maastricht University, The Netherlands; Marcel Karperien, Department of Tissue Regeneration, University of Twente, Enschede, The Netherlands; Jan Van Offel and Filip Vanhoenacker, University Hospital of Antwerp, Edegem, Belgium; Leon Verbruggen, Department of Rheumatology, UZ Brussels, Brussels, Belgium; René Westhovens, Department of Rheumatology, University Hospital Gasthuisberg KU Leuven, Leuven, Belgium; Elisabeth Marelise Eekhoff, Department of Endocrinology, VU University Medical Centre, Amsterdam, The Netherlands; this work was supported by grants from the 'Fonds voor Wetenschappelijk onderzoek' (FWO, G.0065.10N) and a network of excellence grant (EuroBoNeT) from the European Union (FP6) to W.V.H. For the **Australian cohort:** Bronwyn GA Stuckey, Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands WA and School of Medicine and Pharmacology, University of Western

Australia, Crawley WA, Australia; Bryan K Ward, Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands WA, Australia.

AUTHOR CONTRIBUTIONS

O.M.E.A. contributed to study design and funding, oversaw the genotyping, performed data management, quality control, statistical and bioinformatics analyses, and wrote the first draft of the manuscript. S.H.R. designed the study, obtained funding, coordinated the sample collection and phenotyping, and revised the manuscript. K.G., M.L.B., T.C., P.Y.J.C, R.D., J.-P. D., A.F., W.D.F., L.G, F.G., M.J.H., W.V.H, G.I., G.C.N., R.N., S.P., J.d.P.M., T.R., S.L.R, D.R., R.G.-S., M.d.S., L.C.W., and J.P.W. contributed toward clinical sample collection and phenotyping. M.R.V., N.A., S.W., R.G.-S., P.Y.J.C., and F.G. contributed to sample preparation and carried out DNA sequencing to identify samples with *SQSTM1* mutations. All authors critically reviewed the article for important intellectual content and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

O.M.E.A. and S.H.R. have filed patent applications on the use of genetic profiling to identify patients at risk of Paget's disease. The other authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1. Loci for susceptibility to PDB detected by genome wide association study. Manhattan plot of association test results of GWAS stage data showing chromosomal position of 2,487,078 genotyped or imputed SNPs plotted against genomic-control adjusted $-\log_{10} P$. The red horizontal line represents the threshold for genome wide significance ($P < 5 \times 10^{-8}$).

Figure 2. Regional association plots of loci showing genome wide significant association with PDB. Details of loci on chromosome (**a**) 7q33, (**b**) 15q24.1, (**c**) 8q22.3 and (**d**) 14q32.12 showing the chromosomal position (based on NCBI human genome build 36) of SNPs in each region plotted against $-\log_{10} P$ values. Genotyped (squares) and imputed (circles) SNPs are colour-coded according to the extent of linkage disequilibrium with the SNP showing the highest association signal (represented as purple diamonds) from each region in the combined analysis The estimated recombination rates (cM/Mb) from HapMap CEU release 22 are shown as light blue lines and blue arrows represent known genes in each region. The associated regions were defined based on LD with the highest association signal ($r^2 > 0.2$) within a window of 500kb.

Figure 3. Forest plots of overall effect size for SNPs associated with PDB risk from the identified loci on (**a**) 7q33 (rs4294134), (**b**) 8q22.3 (rs2458413), (**c**) 14q32.12 (rs10498635), and (**d**) 15q24.1 (rs5742915). The overall effect size was estimated using meta-analysis of the GWAS sample and the six replication samples. The black squares represent the effect estimates for the individual cohorts and the horizontal lines represent the 95% confidence interval of the estimates. The sizes of the squares are proportionate to the weight of the estimate. The diamonds and triangles represent the overall estimate under fixed effect and random effects model, respectively. The dotted vertical lines represent the overall fixed effect estimates.

Figure 4. Cumulative contribution of genome-wide significant loci to the risk of PDB. Risk allele scores defined by the seven loci associated with PDB risk is plotted against the odds ratio (OR) for PDB. Risk alleles were weighted according to their estimated effect size and weighted risk allele scores were divided into ten equal parts (deciles) using data from the replication cohorts. The OR for PDB risk was calculated for each decile in reference to the fifth decile (D5). Vertical bars represent 95% confidence intervals.

Table 1. Summary of the seven loci showing genome-wide significant association with Paget's disease of bone.

			GWAS Stage		Replication		Combined Overall Effect				
					Fixed effect		Fixed effect		Heterogeneity		
Chr	SNP	RA	*P*	OR (95%CI)	*P*	OR (95%CI)	*P*	OR (95%CI)	P_{het}	I^2	*Closest Gene*
1	rs10494112	G	5.83×10^{-17}	1.75 (1.54 – 1.99)	4.93×10^{-19}	1.69 (1.50 – 1.89)	7.06×10^{-35}	1.72 (1.57 – 1.87)	0.97	00.0	*CSF1, EPS8LS*
7	**rs4294134**	**G**	**1.20×10^{-05}**	**1.50 (1.25 – 1.79)**	**2.29×10^{-05}**	**1.42 (1.20 – 1.66)**	**8.45×10^{-10}**	**1.45 (1.29 – 1.63)**	**0.83**	**00.0**	***NUP205***
8	rs2458413	A	7.85×10^{-11}	1.51 (1.34 – 1.71)	1.09×10^{-07}	1.32 (1.19 – 1.46)	7.38×10^{-17}	1.40 (1.29 – 1.51)	0.10	44.3	*TM7SF4*
10	rs1561570[a]	T	9.56×10^{-18}	1.71 (1.51 – 1.93)	2.09×10^{-21}	1.64 (1.48 – 1.81)	4.37×10^{-38}	1.67 (1.54 – 1.80)	0.01	65.7	*OPTN*
14	**rs10498635**	**C**	**1.51×10^{-05}**	**1.45 (1.23 – 1.71)**	**5.64×10^{-07}**	**1.42 (1.29 – 1.63)**	**2.55×10^{-11}**	**1.44 (1.29 – 1.60)**	**0.62**	**00.0**	***RIN3***
15	**rs5742915**	**C**	**1.40×10^{-07}**	**1.38 (1.22 – 1.54)**	**3.99×10^{-08}**	**1.32 (1.20 – 1.46)**	**1.60×10^{-14}**	**1.34 (1.25 – 1.45)**	**0.56**	**00.0**	***PML***
18	rs3018362	A	1.87×10^{-11}	1.50 (1.34 – 1.69)	1.27×10^{-10}	1.40 (1.26 – 1.55)	7.98×10^{-21}	1.45 (1.34 – 1.56)	0.46	00.0	*TNFRSF11A*

RA, risk allele; OR, odds ratio for the risk allele; CI, confidence interval; I^2, heterogeneity statistics; P_{het}, *P*-value for heterogeneity. Newly identified loci are shown in bold letters. [a]rs1561570 showed significant heterogeneity but random-effect results were genome-wide significant ($P = 4.34 \times 10^{-12}$; OR = 1.68).

ONLINE METHODS

GWAS stage study subjects. This study describes an extension to our previously reported GWAS of PDB in which we used genotype data from 692 PDB cases from our previously described study[1], and extended the case group by genotyping an additional 57 PDB cases. The additional cases were selected from recently recruited subjects in the PRISM study[25]; a randomised trial of two different treatment strategies for PDB patients from the UK. We also increased the size of the control group by using genotype data from 2,930 subjects from the British 1958 Birth Cohort genotyped by the Wellcome Trust Case-Control Consortium[7]. This control group represents a better match to our PDB cases than the previous controls which were recruited from Scotland[1] since, like the PRISM participants, they were recruited from all over the UK. The extended samples size used in this study provided 90% power to detect disease associated allele with MAF = 0.2 and genotype relative risk of 1.4 assuming a multiplicative model and a disease with population prevalence of 2%. This represents a substantial increase in power compared to our previous study[1] where we had 20% power to detect alleles with genotyped relative risk of 1.4.

GWAS stage genotyping and quality control. Genotyping and quality control for the 692 PDB cases were performed using Illumina HumanHap300-Duo arrays as described previously[1]. The additional 57 PDB cases were genotyped using Illumina Human660W Quad version 1 arrays and quality control measures were applied as previously described[1]. Briefly; SNPs with call rate < 95% were excluded and samples with call rate < 90% (n=1); excess heterozygosity (n=1); and non-European ancestry (n= 6; **Supplementary Fig. 4**) were removed before analysis. The genotyping of the British 1958 Birth Cohort was previously performed by the Wellcome Trust Case-Control Consortium using the Illumina Human 1.2M Duo custom array (www.wtccc.org.uk.)[7]. For the control group, SNPs with call rate <95% were excluded and we removed 231 samples because they failed at least one of the following quality control criteria: low call rate, non-European ancestry, gender mismatch, or cryptic relatedness. Population ancestry was determined using multidimensional scaling analysis of identity-by-state (IBS) distances matrix as previously described[1]. After quality control, we analysed 741 PDB cases and 2,699 controls with genotype data for 290,115 SNPs which were common to the three different genotyping arrays. To ensure consistent genotyping between different platforms, a subset of samples were genotyped using at least two different platforms and cross-platform genotype concordance rate was > 99.7% (**Supplementary Table 4**). Additionally, the genotype cluster plots for all SNPs showing association with PDB at $P < 1.0 \times 10^{-4}$ were visually inspected in cases and controls and only high quality genotype data were

included in the analysis. Furthermore, genotype call rate for the top associated SNPs was consistent between cases and controls (**Supplementary Table 5**).

Replication samples. The replication study groups were derived from clinic-based PDB patients and gender-matched controls selected from the same region. Patients with *SQSTM1* mutations were excluded and all study participants provided informed consent. The first replication cohort comprised 175 PDB patients from the UK; 8 PDB cases from Sydney Australia and 215 PDB cases from Western Australia. These patients were of British descent and were matched with 485 unaffected British controls. The second replication cohort (Italian replication cohort 1) comprised 354 PDB cases and 390 unaffected controls enrolled from various referral centres in Italy who took part in the GenPage project[26]. The third replication cohort (Italian replication cohort 2) comprised 205 Italian PDB cases and 238 unaffected controls enrolled from referral centres in Northern, Central and Southern Italy as previously described[27]. The fourth replication cohort comprised 246 sporadic PDB patients recruited from various referrals centres in Belgium and these were matched with 263 controls with no clinical evidence of PDB as previously described[8]. The fifth replication cohort comprised 85 PDB patients and 93 controls recruited from various centres in the Netherlands as described[8,28]. The sixth replication comprised 186 sporadic PDB cases recruited from the Salamanca region in the Castilla-Leon region of Spain and 202 unaffected controls from the same region.

Replication sample genotyping and quality control. Genotyping of replication samples was performed by Sequenom (Hamburg, Germany) using the MassARRAY iPLEX platform. To minimize genotyping bias due to variations between runs; DNA from cases and controls from the six different replication cohorts were distributed into 384-well plates so that each plate had the same number of cases and controls. We included 4000 known genotypes as a quality control measure and the concordance rate between the genotype calls was > 99.8%. We removed 64 samples due to low call rate (< 90%) and the call rate for all genotyped SNPs was >95%.

Imputation. Genome-wide genotype imputation for autosomal SNPs was performed using MACH[29] and the HapMap European (CEU) phased haplotype data from release 22 were used as a reference. We excluded SNPs with poor imputation quality based on the estimated correlation between imputed and true genotypes ($r^2 < 0.3$). Additionally, a subset (2%) of known genotypes were masked during imputation and then imputed genotypes were

compared with true genotypes and the average per allele imputation error rate was 2.9%. Imputed SNPs were tested for association using PorbABEL software[30] implementing a logistic regression model in which the allelic dosage of imputed SNP was used to adjust for uncertainty in imputed genotypes.

Statistical analysis. Statistical analyses were performed using PLINK (Version 1.07) [31] and R (v2.11.1). In GWAS stage, genotyped SNPs were tested for association with PDB using standard allelic (1.d.f) χ^2 statistic. We also performed association testing using regression models in which we adjusted for gender, population clusters (as determined by multidimensional scaling analysis) but results were essentially identical to those obtained from the standard allelic test reported here (data not shown). The genomic inflation factor λ_{GC} was calculated based on the 90% least significant SNPs as described previously[32]. The observed test statistic values were corrected using the genomic control method ($\lambda_{GC} = 1.05$; **Supplementary Fig. 3**). Logistic regression was used to test for independent effects of SNPs where the allelic dosage of the conditioning SNP was entered as a covariate in the regression model. To assess if the reported associations were confounded by age, age of onset or recruitment centre, we performed a regression analysis using case-only data from the GWAS stage to test if any of these factors were associated with the top hits using linear regression models. The results of this analysis showed no evidence to suggest that the reported association is confounded by age, age of onset, or recruitment centre ($P > 0.10$). The cut off point for genome wide significance was set as $P < 5 \times 10^{-8}$ as recently proposed[33]. Association testing of replication data was performed in each replication cohort using standard (1.d.f) χ^2 statistic. To assess combined genetic effects, we performed meta-analysis of all studies using the inverse-variance method assuming fixed-effect model. We also tested random-effects model using DerSimonian-Laird method[34] and between-study heterogeneity was assessed using the Cochran's Q and I^2 metrics. Heterogeneity was considered significant if $P_{het} < 0.05$. The population attributable risk (PAR) for markers showing association with PDB was calculated according to the following formula:
PAR = p(OR-1)/[p(OR-1)+1]; where p is the frequency of the risk allele in controls and OR is the risk allele odds ratio. The cumulative PAR was calculated as follows: Cumulative PAR = 1- ($\Pi_{1\rightarrow n}$ (1-PAR$_i$)); where n is the number of variants and PAR$_i$ is the individual PAR for the *ith* SNP. The proportion of familial risk attributable to the identified loci was calculated as previously described[35] assuming a multiplicative model of association and a sibling relative risk $\lambda_s = 7.0$ as estimated from previous epidemiological studies[22]. Regional association plots were generated using the locuszoom tool[36].

eQTL analysis. SNPs showing genome wide significant association with PDB (or those in strong LD; D'≥0.8) were tested for association with cis-allelic expression of gene transcripts located in the associated regions using publicly available eQTL data[24,37-40]. Only cis-acting allelic associations located within 250 kb of either 5' or 3' end of the associated gene with expression *P*-value < 1 x 10^{-5} were considered. To avoid false detection, we excluded expression data if the gene probe contained a polymorphic SNP or was located in a highly repetitive sequence.

References

1. Albagha, O. M. *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.* **42**, 520-524 (2010).

2. Cooper, C. *et al.* The epidemiology of Paget's disease in Britain: is the prevalence decreasing? *J.Bone Miner.Res.* **14**, 192-197 (1999).

3. Morales-Piga, A. A., Rey-Rey, J. S., Corres-Gonzalez, J., Garcia-Sagredo, J. M., & Lopez-Abente, G. Frequency and characteristics of familial aggregation of Paget's disease of bone. *J.Bone Miner.Res.* **10**, 663-670 (1995).

4. Visconti, M. R. *et al.* Mutations of SQSTM1 are associated with severity and clinical outcome in Paget disease of bone. *J.Bone Miner.Res.* **25**, 2368-2373 (2010).

5. Laurin, N., Brown, J. P., Morissette, J., & Raymond, V. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. *Am.J.Hum.Genet.* **70**, 1582-1588 (2002).

6. Hocking, L. J. *et al.* Domain-specific mutations in sequestosome 1 (SQSTM1) cause familial and sporadic Paget's disease. *Hum.Mol.Genet.* **11**, 2735-2739 (2002).

7. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).

8. Chung, P. Y. *et al.* The majority of the genetic risk for Paget's disease of bone is explained by genetic variants close to the CSF1, OPTN, TM7SF4, and TNFRSF11A genes. *Hum.Genet.* (2010).

9. Hartgers, F. C. *et al.* DC-STAMP, a novel multimembrane-spanning molecule preferentially expressed by dendritic cells. *Eur.J.Immunol.* **30**, 3585-3590 (2000).

10. Yagi, M. *et al.* DC-STAMP is essential for cell-cell fusion in osteoclasts and foreign body giant cells. *J.Exp.Med.* **202**, 345-351 (2005).

11. Kukita, T. *et al.* RANKL-induced DC-STAMP is essential for osteoclastogenesis. *J.Exp.Med.* **200**, 941-946 (2004).

12. Nishida, T., Emura, K., Kubota, S., Lyons, K. M., & Takigawa, M. CCN family 2/connective tissue growth factor (CCN2/CTGF) promotes osteoclastogenesis via induction of and interaction with dendritic cell-specific transmembrane protein (DC-STAMP). *J.Bone Miner.Res.* (2010).

13. Grandi, P. *et al.* Nup93, a vertebrate homologue of yeast Nic96p, forms a complex with a novel 205-kDa protein and is required for correct nuclear pore assembly. *Mol.Biol.Cell* **8**, 2017-2038 (1997).

14. Saito, K. *et al.* A novel binding protein composed of homophilic tetramer exhibits unique properties for the small GTPase Rab5. *J.Biol.Chem.* **277**, 3412-3418 (2002).

15. Kajiho, H. *et al.* RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J.Cell Sci.* **116**, 4159-4168 (2003).

16. Coxon, F. P. & Rogers, M. J. The role of prenylated small GTP-binding proteins in the regulation of osteoclast function. *Calcif.Tissue Int.* **72**, 80-84 (2003).

17. Van Wesenbeeck, L. *et al.* Involvement of PLEKHM1 in osteoclastic vesicular transport and osteopetrosis in incisors absent rats and humans. *J.Clin.Invest* **117**, 919-930 (2007).

18. Watts, G. D. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat.Genet.* **36**, 377-381 (2004).

19. Lin, H. K., Bergmann, S., & Pandolfi, P. P. Cytoplasmic PML function in TGF-beta signalling. *Nature* **431**, 205-211 (2004).

20. Smits, P. *et al.* Lethal skeletal dysplasia in mice and humans lacking the golgin GMAP-210. *N.Engl.J.Med.* **362**, 206-216 (2010).

21. Hennies, H. C. *et al.* Gerodermia osteodysplastica is caused by mutations in SCYL1BP1, a Rab-6 interacting golgin. *Nat.Genet.* **40**, 1410-1412 (2008).

22. Siris, E. S., Ottman, R., Flaster, E., & Kelsey, J. L. Familial aggregation of Paget's disease of bone. *J.Bone Miner.Res.* **6**, 495-500 (1991).

23. Rivadeneira, F. *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat.Genet.* **41**, 1199-1206 (2009).

24. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One.* **5**, e10693 (2010).

25. Langston, A. L. *et al.* Randomized trial of intensive bisphosphonate treatment versus symptomatic management in Paget's disease of bone. *J.Bone Miner.Res.* **25**, 20-31 (2010).

26. Falchetti, A. *et al.* Genetic epidemiology of Paget's disease of bone in italy: sequestosome1/p62 gene mutational test and haplotype analysis at 5q35 in a large representative series of sporadic and familial Italian cases of Paget's disease of bone. *Calcif.Tissue Int.* **84**, 20-37 (2009).

27. Gennari, L. *et al.* SQSTM1 gene analysis and gene-environment interaction in Paget's disease of bone. *J.Bone Miner.Res.* **25**, 1375-1384 (2010).

28. Eekhoff, M. E. *et al.* Paget's disease of bone in The Netherlands: a population-based radiological and biochemical survey-the Rotterdam Study. *J.Bone Miner.Res.* **19**, 566-570 (2004).

29. Li, Y. & Abecasis, G. R. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am.J.Hum.Genet.* **S79**, 2290 (2006).

30. Aulchenko, Y. S., Struchalin, M. V., & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC.Bioinformatics* **11**, 134 (2010).

31. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am.J.Hum.Genet.* **81**, 559-575 (2007).

32. Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat.Genet.* **37**, 1243-1246 (2005).

33. Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet.Epidemiol.* **32**, 381-385 (2008).

34. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin.Trials* **7**, 177-188 (1986).

35. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat.Genet.* **42**, 579-589 (2010).

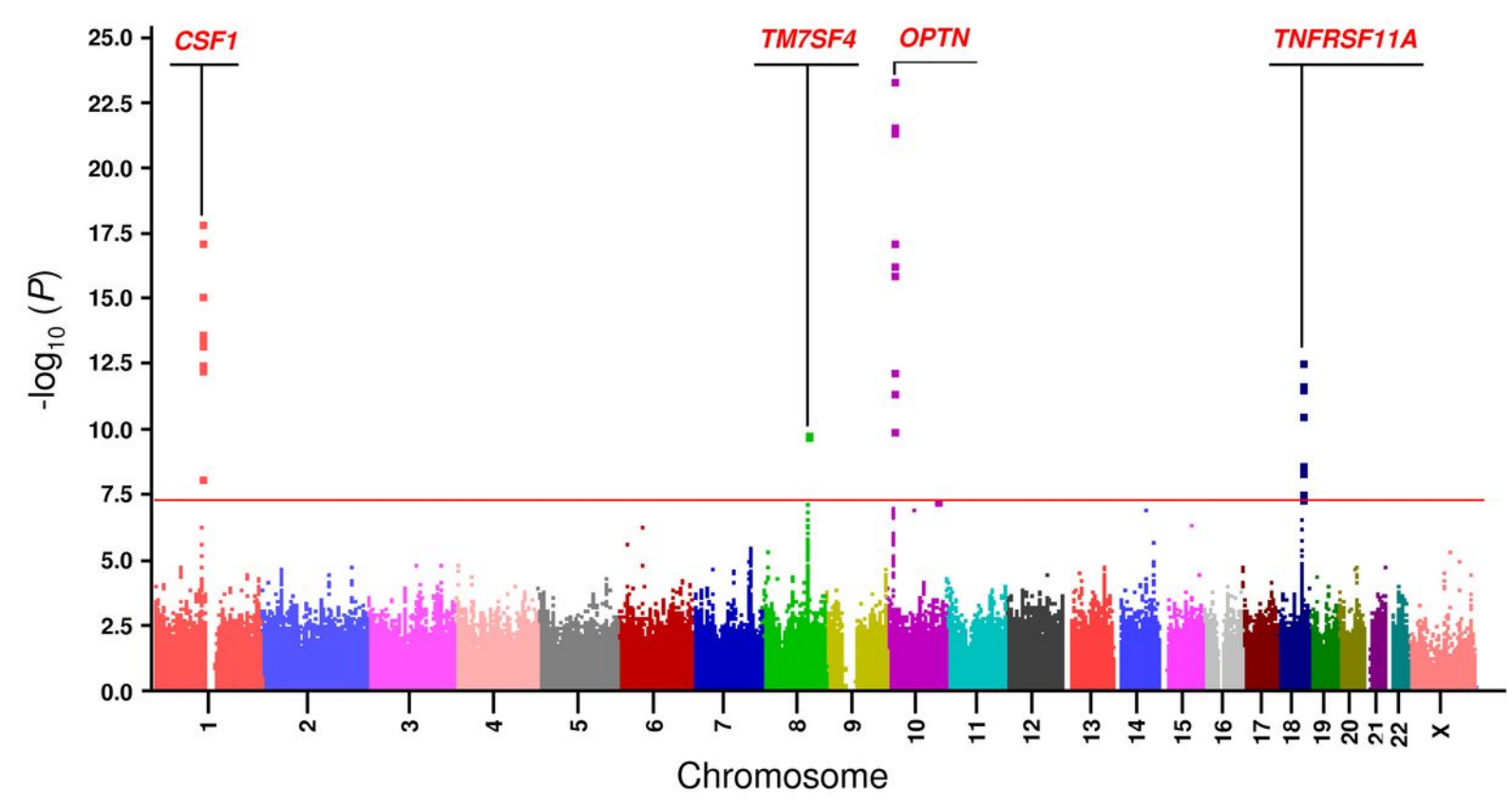
36. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).

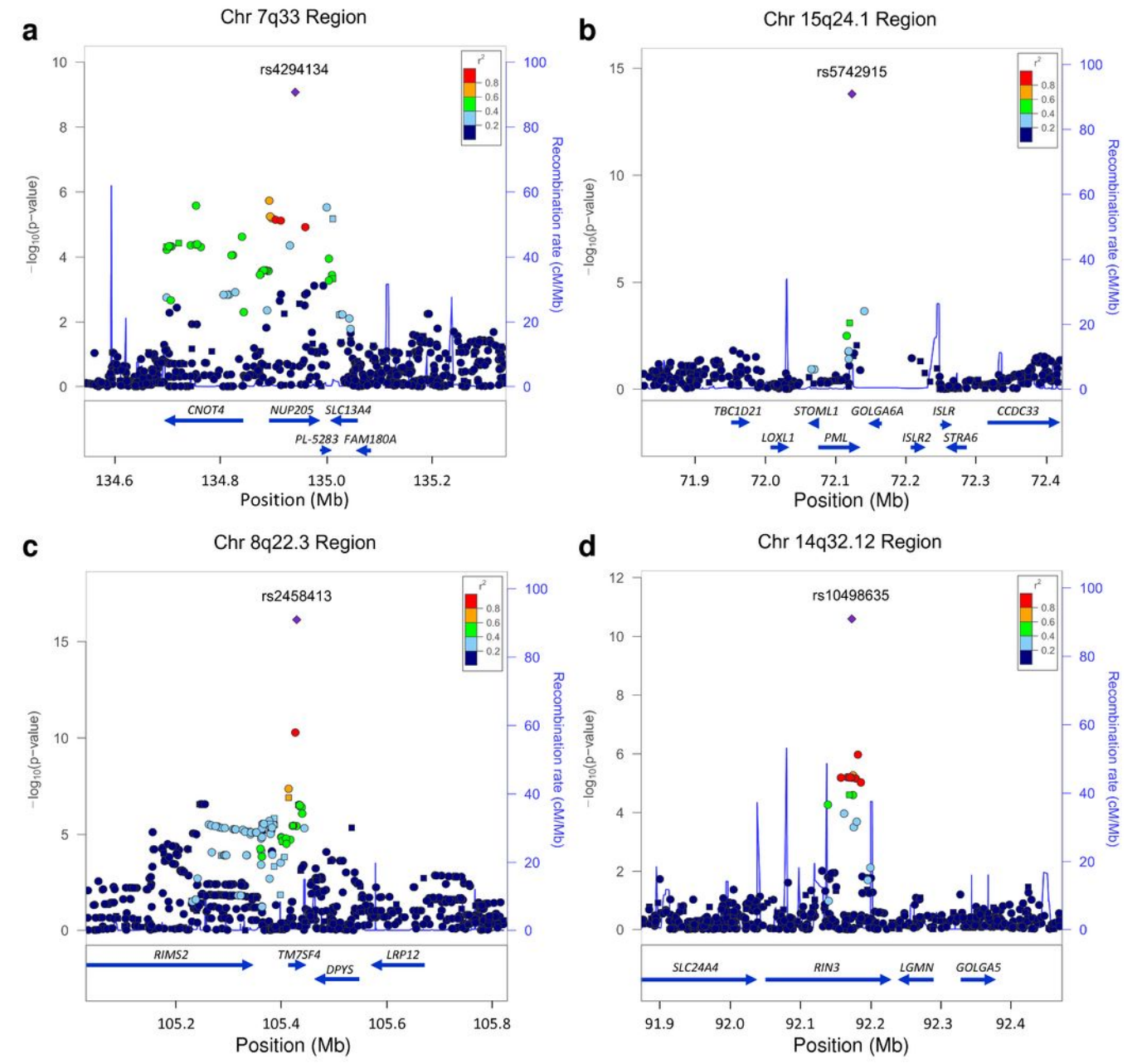
37. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).

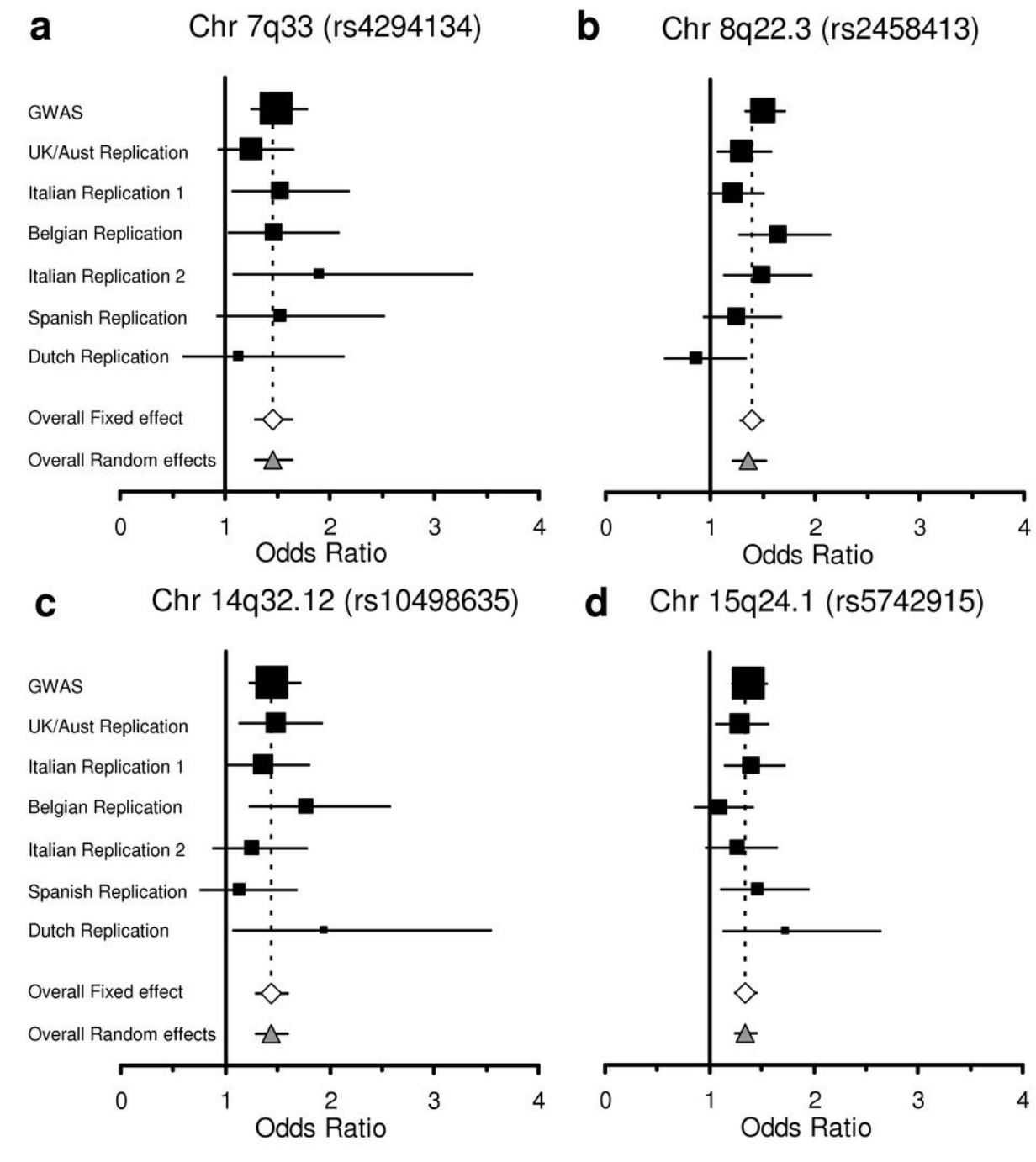
38. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat.Genet.* **39**, 1217-1224 (2007).

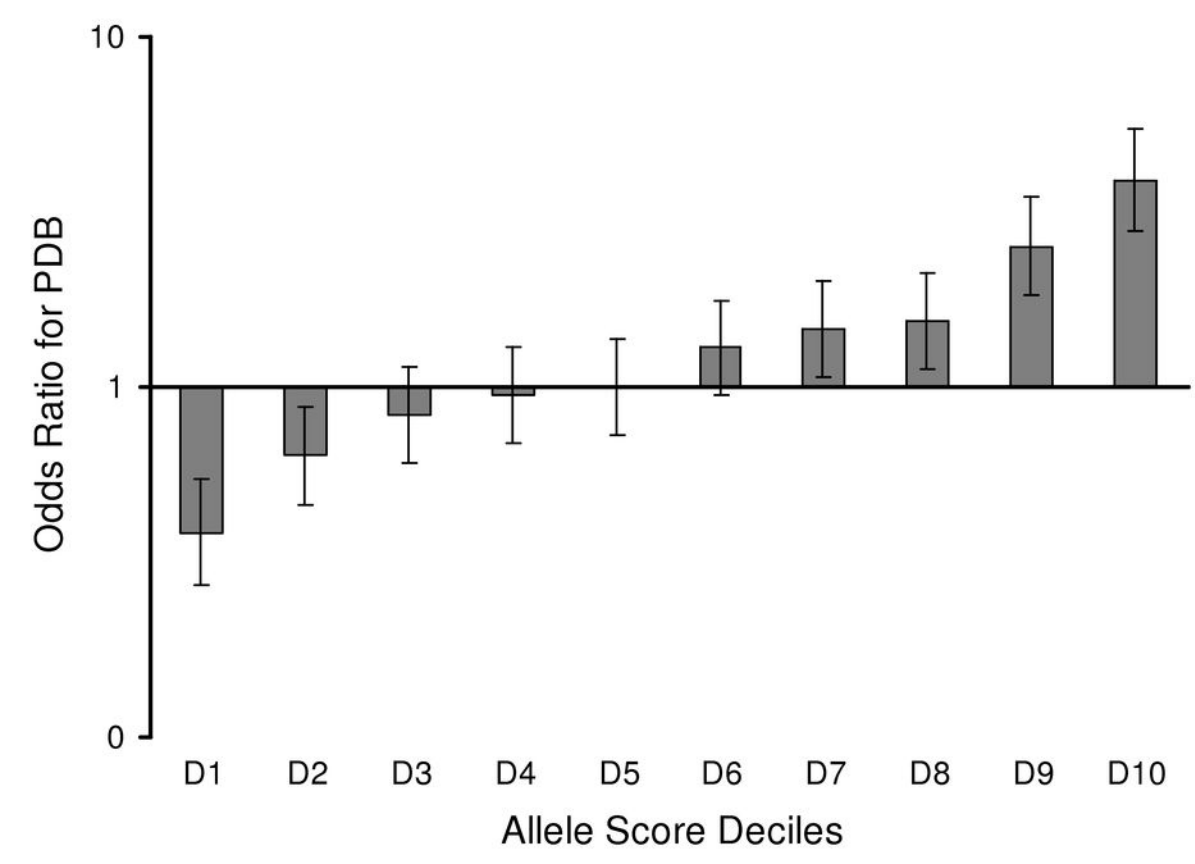
39. Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).

40. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777 (2010).


CSF1
TM7SF4
OPTN
TNFRSF11A
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-log10 (P)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome


a
Chr 7q33 Region
rs4294134
−log10(p-value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
CNOT4
NUP205
SLC13A4
PL-5283
FAM180A
134.6
134.8
135.0
135.2
Position (Mb)
b
Chr 15q24.1 Region
rs5742915
TBC1D21
STOML1
GOLGA6A
ISLR
CCDC33
LOXL1
PML
ISLR2
STRA6
71.9
72.0
72.1
72.2
72.3
72.4
Position (Mb)
c
Chr 8q22.3 Region
rs2458413
RIMS2
TM7SF4
LRP12
DPYS
105.2
105.4
105.6
105.8
Position (Mb)
d
Chr 14q32.12 Region
rs10498635
SLC24A4
RIN3
LGMN
GOLGA5
91.9
92.0
92.1
92.2
92.3
92.4
Position (Mb)


a
Chr 7q33 (rs4294134)
GWAS
UK/Aust Replication
Italian Replication 1
Belgian Replication
Italian Replication 2
Spanish Replication
Dutch Replication
Overall Fixed effect
Overall Random effects
0
1
2
3
4
Odds Ratio
b
Chr 8q22.3 (rs2458413)
0
1
2
3
4
Odds Ratio
c
Chr 14q32.12 (rs10498635)
GWAS
UK/Aust Replication
Italian Replication 1
Belgian Replication
Italian Replication 2
Spanish Replication
Dutch Replication
Overall Fixed effect
Overall Random effects
0
1
2
3
4
Odds Ratio
d
Chr 15q24.1 (rs5742915)
0
1
2
3
4
Odds Ratio


10
1
0
Odds Ratio for PDB
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
Allele Score Deciles