nature genetics

Genome-wide association study identifies two susceptibility loci for osteosarcoma

Sharon A Savage[1], Lisa Mirabello[1], Zhaoming Wang[2], Julie M Gastier-Foster[3,4], Richard Gorlick[5], Chand Khanna[6], Adrienne M Flanagan[7,8], Roberto Tirabosco[8], Irene L Andrulis[9], Jay S Wunder[9], Nalan Gokgoz[9], Ana Patiño-Garcia[10], Luis Sierrasesúmaga[10], Fernando Lecanda[10], Nilgün Kurucu[11], Inci Ergurhan Ilhan[11], Neriman Sari[11], Massimo Serra[12], Claudia Hattinger[12], Piero Picci[12], Logan G Spector[13], Donald A Barkauskas[14], Neyssa Marina[15,16], Silvia Regina Caminada de Toledo[17], Antonio S Petrilli[17], Maria Fernanda Amary[8], Dina Halai[8], David M Thomas[18], Chester Douglass[19], Paul S Meltzer[6], Kevin Jacobs[2], Charles C Chung[2], Sonja I Berndt[1], Mark P Purdue[1], Neil E Caporaso[1], Margaret Tucker[1], Nathaniel Rothman[1], Maria Teresa Landi[1], Debra T Silverman[1], Peter Kraft[19], David J Hunter[19], Nuria Malats[20], Manolis Kogevinas[21–24], Sholom Wacholder[1], Rebecca Troisi[1], Lee Helman[6], Joseph F Fraumeni Jr[1], Meredith Yeager[2], Robert N Hoover[1] & Stephen J Chanock[1]

Osteosarcoma is the most common primary bone malignancy of adolescents and young adults. To better understand the genetic etiology of osteosarcoma, we performed a multistage genome-wide association study consisting of 941 individuals with osteosarcoma (cases) and 3,291 cancer-free adult controls of European ancestry. Two loci achieved genome-wide significance: a locus in the *GRM4* gene at 6p21.3 (encoding glutamate receptor metabotropic 4; rs1906953; $P = 8.1 \times 10^{-9}$) and a locus in the gene desert at 2p25.2 (rs7591996 and rs10208273; $P = 1.0 \times 10^{-8}$ and 2.9×10^{-7}, respectively). These two loci warrant further exploration to uncover the biological mechanisms underlying susceptibility to osteosarcoma.

Osteosarcoma is the most common primary malignant bone tumor in children and young adults, affecting approximately four persons per million each year in the United States[1]. Peak incidence correlates with the pubertal growth spurt, occurring earlier in females than in males. It is more common at sites of rapid bone growth. Tall stature and high birth weight are proven risk factors[2], and osteosarcoma is a syndrome-associated malignancy in the Li-Fraumeni, Rothmund-Thomson and hereditary retinoblastoma cancer susceptibility syndromes. Several small case-control studies have reported preliminary associations of common genetic variants with osteosarcoma risk in biologically plausible pathways[3–11] (for example, growth and DNA repair), but statistical power has been limited by small sample sizes[12].

We developed an international, multi-institutional collaborative effort to conduct a genome-wide association study (GWAS) of osteosarcoma (**Supplementary Table 1**). We extracted germline genomic DNA from either blood or buccal cells obtained from osteosarcoma case series using standard methods. Participating subjects provided informed consent under the auspices of local institutional review boards (IRBs). We selected control subjects from previously scanned cancer-free adults over the age of 55 years drawn from large case-control and cohort studies, using an approach that was successfully applied for Ewing sarcoma[13] and pediatric acute lymphoblastic leukemia (ALL)[14].

We conducted genotyping of all cases using the Illumina OmniExpress SNP microarray, with samples over time being divided into two stages, which we call stage 1a and 1b, owing to the staggered

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [2]Cancer Genomics Research Laboratory, National Cancer Institute, Division of Cancer Epidemiology and Genetics, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland, USA. [3]Nationwide Children's Hospital, Columbus, Ohio, USA. [4]Department of Pathology and Pediatrics, The Ohio State University, Columbus, Ohio, USA. [5]Albert Einstein College of Medicine, The Children's Hospital at Montefiore, Bronx, New York, USA. [6]Center for Cancer Research, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [7]University College London (UCL) Cancer Institute, London, UK. [8]Royal National Orthopaedic Hospital National Health Service (NHS) Trust, Stanmore, UK. [9]Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada. [10]Department of Pediatrics, University Clinic of Navarra, Universidad de Navarra, Pamplona, Spain. [11]Department of Pediatric Oncology, A.Y. Ankara Oncology Training and Research Hospital, Ankara, Turkey. [12]Laboratory of Experimental Oncology, Orthopaedic Rizzoli Institute, Bologna, Italy. [13]Department of Pediatrics, Division of Epidemiology and Clinical Research, University of Minnesota, Minneapolis, Minnesota, USA. [14]Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [15]Stanford University, Palo Alto, California, USA. [16]Department of Pediatric Hematology-Oncology, Lucile Packard Children's Hospital, Palo Alto, California, USA. [17]Pediatric Oncology Institute, Grupo de Apoio ao Adolescente e à Criança com Câncer (GRAACC), Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil. [18]Sir Peter MacCallum Department of Oncology, University of Melbourne, East Melbourne, Victoria, Australia. [19]Harvard School of Public Health, Boston, Massachusetts, USA. [20]Centro Nacional de Investigaciones Oncológicas, Madrid, Spain. [21]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [22]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [23]Network Biomedical Research Centre in Epidemiology and Public Health (CIBERESP), Barcelona, Spain. [24]National School of Public Health, Athens, Greece. Correspondence should be addressed to S.A.S. (savagesh@mail.nih.gov).

Received 14 January; accepted 26 April; published online 2 June 2013; doi:10.1038/ng.2645

Table 1 Summary of combined association results for 941 osteosarcoma cases and 3,291 controls

dbSNP	Locus	Position[a]	Allele[b]	Stage	EAF		*P*	OR (95% CI)
					Cases	Controls		
rs1906953	6p21.31	34144424	C/T	Discovery	0.19	0.15	4.0×10^{-5}	1.45 (1.21–1.73)
				Replication	0.21	0.12	6.5×10^{-6}	2.00 (1.47–2.72)
				Combined			8.0×10^{-9}	1.57 (1.35–1.83)
rs7591996	2p25.2	6378872	A/C	Discovery	0.50	0.45	4.8×10^{-5}	1.30 (1.15–1.47)
				Replication	0.55	0.43	5.7×10^{-6}	1.69 (1.35–2.13)
				Combined			1.0×10^{-8}	1.39 (1.23–1.54)
rs10208273	2p25.2	6441445	G/A	Discovery	0.36	0.30	3.5×10^{-6}	1.37 (1.20–1.56)
				Replication	0.34	0.29	2.7×10^{-2}	1.31 (1.03–1.67)
				Combined			2.9×10^{-7}	1.35 (1.21–1.52)
rs17206779	5q12.3	64483533	T/C	Discovery	0.53	0.47	2.2×10^{-5}	1.32 (1.16–1.49)
				Replication	0.54	0.46	6.7×10^{-3}	1.37 (1.09–1.72)
				Combined			5.1×10^{-7}	1.33 (1.19–1.47)

The discovery stage consisted of the combined scans of stage 1a (596 cases) and stage 1b (98 cases) for a total of 694 scanned osteosarcoma cases and 2,703 cancer-free adult controls, all of European ancestry. The independent replication set consisted of 247 osteosarcoma cases genotyped by TaqMan and 588 controls previously scanned on the Illumina Omni 2.5M or OmniExpress SNP microarray. Combined results report the data from the meta-analysis of stage 1 and the TaqMan replication (941 cases and 3,291 controls). EAF, effect (coded) allele frequency.
[a]Build 36 position. [b]Reference/effect alleles.

receipt of the samples in this multi-institutional study design (Online Methods and **Supplementary Fig. 1**). In stage 1a, we genotyped 910 available osteosarcoma cases from 5 studies. On the basis of quality control filtering (including for locus and sample missing rates, sample heterozygosity, departure from Hardy-Weinberg equilibrium and sex discrepancies) and assessment of underlying population substructure (STRUCTURE and principal-components analyses (PCA)), we included 596 cases of European ancestry in the primary GWAS analysis (**Supplementary Figs. 2** and **3**); 97 African-American and 99 Hispanic cases were identified and excluded from the present analysis. In stage 1b, we genotyped a later collection of samples consisting of 218 osteosarcoma cases using the OmniExpress SNP microarray (**Supplementary Table 1**). We applied the same quality control filtering and population substructure analyses as for stage 1a, yielding an additional 98 cases of European ancestry for the final analysis. The choice of SNPs for follow-up was based on association results for the first set (stage 1a), but, for the final analysis, we present the results of the full set of 694 cases (namely, combined stages 1a and 1b).

Overall, 698,968 SNPs passed quality control metrics for the full set of 694 cases for final analysis. The numbers of SNPs overlapping those genotyped in pooled controls were 510,856, 510,856, 310,384 and 304,092 for the US, Spanish, Italian and UK components, respectively, and we used these SNPs in the association analyses (Online Methods).

We combined data from already scanned controls drawn from two large cohort studies (the Prostate, Lung, Colon and Ovarian Cancer Prevention Trial (PLCO)[15] and the American Cancer Society Cancer Prevention Study II (CPSII)[16] scanned on the Illumina Omni 2.5M SNP microarray) and three European studies (the Spanish Bladder Cancer Study[17] scanned on the Illumina HumanHap 1M SNP microarray, the Environment and Genetics Lung Cancer Etiology Study (EAGLE)[18] scanned on the Illumina HumanHap550 SNP microarray and the Wellcome Trust Case Control Consortium (WTCCC)[19] scanned on the Illumina HumanHap550 SNP microarray). We selected 2,703 controls on the basis of comparable quality control metrics for comparison with the scanned cases. We conducted a systematic assessment of the underlying population substructure to remove highly admixed individuals (with non-European admixture of >20%) from a model to investigate cases of European ancestry (**Supplementary Figs. 2** and **3**).

We performed association analyses using a 1-degree-of-freedom trend test adjusted by sex and eigenvectors. We designed optimized TaqMan assays based on SNP associations observed in the first set of cases in the scan (stage 1a; **Supplementary Table 2**). We genotyped this set of 30 SNPs in a follow-up set of 247 osteosarcoma cases (**Supplementary Table 1**) and 588 controls drawn from PLCO and the Nurse Health Study (NHS). We applied a fixed-effect meta-analysis to the combined results of the two scanned sets comprising stage 1 (596 cases in stage 1a and 98 cases in stage 1b, yielding a total of 694 unique cases) and the replication stage of 30 SNPs with association $P < 1.0 \times 10^{-4}$. In a fixed-effect meta-analysis of all cases and controls of European ancestry (941 cases and 3,291 cancer-free adult controls), we observed two regions at 6p21.3 and 2p25.2 with associations that achieved genome-wide significance (**Table 1** and **Supplementary Table 2**).

The locus at 6p21.3, marked by rs1906953, was associated with susceptibility to osteosarcoma ($P = 8.0 \times 10^{-9}$; odds ratio (OR) = 1.57, 95% confidence interval (CI) = 1.35–1.83) (**Table 1**). The rs1906953 SNP is located in intron 7 (g.34036446C >T) of the *GRM4* gene encoding glutamate receptor metabotropic 4 (**Fig. 1a**). The most significantly associated SNP in this region resides within *GRM4* in a distinct haplotype block from the human leukocyte antigen (HLA) class II region, which is more than 1 Mb telomeric at 6p21.3. It is also notable that between the HLA class II region and the region associated with osteosarcoma risk is a distinct region containing *BAK1*, which in published GWAS analyses has been conclusively associated with platelet counts[20–22], chronic lymphocytic leukemia[23] and testicular germ cell tumors[24].

The C reference allele of rs1906953 in *GRM4* was highly conserved, and its frequency was higher in African populations; the global minor allele frequency (MAF) was 0.29 in 1000 Genomes Project data (Phase 1 genotype data from 1,094 individuals)[25], with greater frequency of the T allele in individuals of Asian or African ancestry (MAF = 0.46 and 0.56, respectively). Although there were four surrogates for this SNP in individuals of European ancestry ($r^2 > 0.6$ within ± 500 kb) in 1000 Genomes Project data, the index SNP rs1906953 mapped to a DNase I hypersensitivity region in the Encyclopedia of DNA Elements (ENCODE)[26] data set, raising the possibility that the variant contributing to osteosarcoma risk is within a region of open chromatin (**Supplementary Table 3**) and could contain active regulatory elements. One intronic surrogate SNP (rs73410895) was predicted to alter known regulatory motifs, including HMG-IY, Pou3f2 and Pou1f1 (**Supplementary Table 3**).

GRM4 is a plausible candidate gene in osteosarcoma that has been implicated in intracellular signaling and inhibition of the cyclic AMP (cAMP) signaling cascade. In mice, a cAMP-dependent protein kinase (*Prkar1a*) is an osteosarcoma tumor suppressor gene[27,28], suggesting that the cAMP pathway is important in osteosarcoma. Although glutamate signaling is best characterized in the central nervous system (CNS), where it is known to be involved in the excitability of gonadotropin-releasing hormone neurons, it also occurs in bone[29]. The GRM4 receptor is expressed in bone osteoblast (bone-building) and

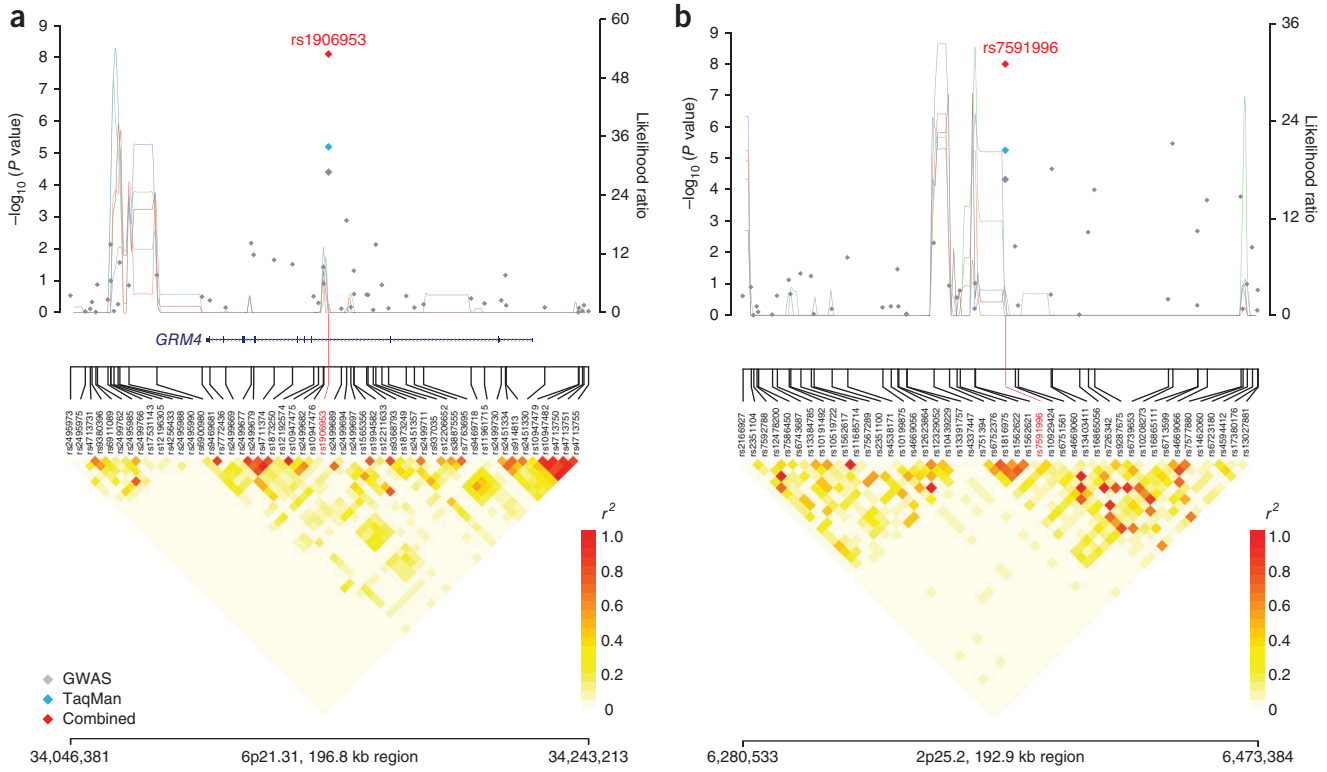

Figure 1 Regional plots of loci associated with osteosarcoma. (**a**,**b**) Regional plots of association results, recombination hotspots and LD for the 6p21.31 (**a**) and 2p25.2 (**b**) osteosarcoma susceptibility loci. Association results from a trend test are represented as the $-\log_{10} P$ values of the SNPs (left *y* axis) plotted according to their chromosomal positions (*x* axis). LD structure based on data from controls ($n = 2,703$) was visualized using snp.plotter software. The line graph shows likelihood ratio statistics (right *y* axis) for recombination hotspots calculated with SequenceLDhot software; different colors represent the results from 5 tests of 100 controls without resampling. Physical locations are based on NCBI Build 36 of the human genome. Gene annotation is based on NCBI RefSeq genes from the UCSC Genome Browser.

osteoclast (bone-resorbing) cells, suggesting that glutamate signaling is involved in cell differentiation and regulation during bone formation and resorption[30]. *GRM4* is expressed in human osteosarcoma cells[31] and is associated with poor prognosis in colorectal cancer[32], pediatric CNS tumors[33], rhabdomyosarcoma and multiple myeloma[34], as well as with cancer cell proliferation *in vitro*[35].

We observed a second association signal in an intergenic region at 2p25.2 (**Fig. 1b**), with the rs7591996 SNP achieving genome-wide significance (OR = 1.39, 95% CI = 1.23–1.54; $P = 1.0 \times 10^{-8}$). A second SNP, rs10208273, was moderately correlated with the first signal ($r^2 = 0.32$ in HapMap 3 release 2) and approached genome-wide significance (OR = 1.35, 95% CI = 1.21–1.52; $P = 2.9 \times 10^{-7}$) (**Fig. 1b** and **Table 1**). The risk allele frequency for rs7591996 was 0.50 in osteosarcoma cases of European ancestry and 0.45 in controls. rs7591996 had a global MAF of 0.38 in 1000 Genomes Project data; the risk allele was the minor allele in CEU individuals (Utah residents of Northern and Western European ancestry) and the major allele in individuals of Asian or African ancestry (MAF = 0.35 and 0.14, respectively). Both rs7591996 and rs10208273 occur in a region of relatively low linkage disequilibrium (LD) (**Fig. 1b**). Data from the 1000 Genomes Project for European populations documented at least 27 surrogate SNPs ($r^2 > 0.6$ within ± 500 kb). According to ENCODE[26] data, the two index SNPs did not localize to a region of active regulatory elements or transcription factor binding sites. However, several of the surrogate SNPs altered known regulatory motifs and transcription factor binding sites (**Supplementary Table 3**) and may affect gene expression. These findings suggest that further sequencing and fine mapping will be required to determine which variants will be optimal for the functional studies needed to explain disease association.

Finally, we report a third locus that was promising but did not yet achieve genome-wide significance, rs17206779 (OR = 0.75, 95% CI = 0.68–0.84; $P = 5.1 \times 10^{-7}$), located in the *ADAMTS6* gene encoding ADAM metallopeptidase with thrombospondin type 1 motif 6 (g.64447777C > T) at 5q12.3. It is notable that variations in genes encoding members of the ADAMTS protein family have been associated with height[36], a known risk factor for osteosarcoma[2]. Further data are required to confirm the association of this locus and to then fine map the region before conducting functional studies.

To further explore the two significantly associated regions reported in this study, we imputed SNPs on the basis of 1000 Genomes Project data (March 2012 release) together with the DCEG Imputation Reference Set version 1 (ref. 37) using the IMPUTE2 program[38] across a 1-Mb region centered on each index SNP (Online Methods). A subsequent association analysis did not identify new signals that were substantially stronger than those of the genotyped SNPs for either the 2p25.2 or 5q12.3 region (**Supplementary Fig. 4**). Although there seemed to be stronger signals in the imputed data for the HLA class II region (**Supplementary Fig. 4a**), there is a strong recombination hotspot separating it from the region marked by the rs1906953 index SNP. It will be critical to pursue more detailed mapping and genotyping of the HLA class II region to determine whether there is an independent signal in addition to the one that resides in *GRM4*.

An assessment of the loci previously reported in candidate gene association studies[3–12] did not show overlap with any of the reported

loci approaching genome-wide significance (**Supplementary Table 4**). It is notable that the SNPs in the 8q24.3 region and the multicancer-associated 8q24.21 region seemed to be promising candidates[10]; the strongest signal among a set of highly correlated SNPs at 8q24.21, rs11777807, had association $P = 4.6 \times 10^{-5}$, and rs369051 at 8q24.3 had association $P = 1.16 \times 10^{-4}$. Further studies are needed to determine whether these are stable associations or were perhaps due to chance.

Through a multistage GWAS of osteosarcoma, we have identified two susceptibility regions at 2p25.2 and 6p21.3, with the latter harboring a plausible candidate gene, *GRM4*. It is noteworthy that the loci we have identified demonstrate estimated ORs greater than 1.5, which are higher than those observed for most variants associated with cancer risk in adults. Our findings are consistent with the high risk estimates reported for Ewing sarcoma[13] and ALL[14], other rare pediatric tumors. Further investigation of these associated loci is warranted to uncover the biological mechanisms underlying susceptibility to osteosarcoma.

URLs. Cancer Genetic Markers of Susceptibility (CGEMS) portal, http://cgems.cancer.gov/; CGF, http://cgf.nci.nih.gov/; GLU, http://code.google.com/p/glu-genetics/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; STRUCTURE, http://pritch.bsd.uchicago.edu/structure.html; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; SequenceLDhot, http://www.maths.lancs.ac.uk/~fearnhea/Hotspot/; snp.plotter, http://cbdb.nimh.nih.gov/~kristin/snp.plotter.html; PHASE v2.1, http://www.stat.washington.edu/stephens/phase/download.html; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank G. Maganoli for tissue banking, M. Fanelli for DNA isolation and C. Ferrari for updating clinicopathological data at the Orthopaedic Rizzoli Institute. We thank A. Griffin and D. Marsilio for data collection and T. Selander and the Biospecimen Repository staff at Mount Sinai Hospital. We acknowledge the advice of F. Real at the Spanish National Cancer Research Centre (CNIO). We thank F. Tesser Gamba at the Pediatric Oncology Institute at GRAACC-UNIFESP, and we also thank the International Sarcoma Kindred Study.

This study was funded by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health and the Bone Cancer Research Trust UK. Research is supported by the Chair's Grant U10 CA98543 and Human Specimen Banking Grant U24 CA114766 to the Children's Oncology Group from the National Cancer Institute, US National Institutes of Health. Additional support for research is provided by a grant from the WWWW (QuadW) Foundation to the Children's Oncology Group. This work was supported by grants to I.L.A. and J.S.W. from the Ontario Research Fund and Canadian Foundation for Innovation. This study was also supported by biobank grants from the Regione Emilia-Romagna and by the infrastructure and personnel of the Royal National Orthopaedic Hospital Musculoskeletal Research Programme and Biobank. Support was also provided to A.M.F. by the National Institute for Health Research UCL Hospitals (UCLH) Biomedical Research Centre and the UCL Experimental Cancer Centre. Funding was also provided by PI10/01580, the Fondo de Investigación Sanitaria (FIS), the Instituto de Salud Carlos III (ISCIII) and the Caja de Ahorros de Navarra (CAN) Programme 'Tú eliges, tú decides' to A.P.-G. and L.S. and by an Asociación Española Contra el Cáncer (AECC) project to F.L.

AUTHOR CONTRIBUTIONS

S.A.S. and S.J.C. designed the project. J.M.G.-F., R.G., C.K., A.M.F., R. Tirabosco, I.L.A., J.S.W., N.G., L.G.S., D.A.B., N. Marina, A.P.-G., L.S., F.L., M.S., C.H., P.P., N.K., I.E.I., N.S., S.R.C.d.T., A.S.P., M.F.A., D.H., D.M.T., C.D., P.S.M., S.I.B., M.P.P., N.E.C., M.T., N.R., M.T.L., D.T.S., P.K., D.J.H., N. Malats, M.K., S.W., R. Troisi, L.H., J.F.F. and R.N.H. performed sample collection and clinical characterization. K.J., C.C.C., M.Y. and Z.W. performed genotyping. Z.W. and L.M. performed statistical analysis. The manuscript was written by S.A.S., L.M., Z.W. and S.J.C. and reviewed by all coauthors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Mirabello, L., Troisi, R. & Savage, S. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. *Cancer* **115**, 1531–1543 (2009).
2. Mirabello, L. *et al.* Height at diagnosis and birth-weight as risk factors for osteosarcoma. *Cancer Causes Control* **22**, 899–908 (2011).
3. Patio-Garcia, A., Sotillo-Pieiro, E., Modesto, C. & Sierrases-Maga, L. Analysis of the human tumour necrosis factor-α (TNFα) gene promoter polymorphisms in children with bone cancer. *J. Med. Genet.* **37**, 789–792 (2000).
4. Ruza, E., Sotillo, E., Sierrasesúmaga, L., Azcona, C. & Patiño-Garcia, A. Analysis of polymorphisms of the vitamin D receptor, estrogen receptor, and collagen Iα1 genes and their relationship with height in children with bone cancer. *J. Pediatr. Hematol. Oncol.* **25**, 780–786 (2003).
5. Savage, S.A. *et al.* Germ-line genetic variation of *TP53* in osteosarcoma. *Pediatr. Blood Cancer* **49**, 28–33 (2007).
6. Savage, S.A. *et al.* Analysis of genes critical for growth regulation identifies Insulin-like Growth Factor 2 Receptor variations with possible functional significance as risk factors for osteosarcoma. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1667–1674 (2007).
7. Koshkina, N.V. *et al.* Exploratory analysis of *Fas* gene polymorphisms in pediatric osteosarcoma patients. *J. Pediatr. Hematol. Oncol.* **29**, 815–821 (2007).
8. Toffoli, G. *et al.* Effect of *TP53* Arg72Pro and *MDM2* SNP309 polymorphisms on the risk of high-grade osteosarcoma development and survival. *Clin. Cancer Res.* **15**, 3550–3556 (2009).
9. Hu, Y.S. *et al.* Association between TGFBR1*6A and osteosarcoma: a Chinese case-control study. *BMC Cancer* **10**, 169 (2010).
10. Mirabello, L. *et al.* Genetic variation at chromosome 8q24 in osteosarcoma cases and controls. *Carcinogenesis* **31**, 1400–1404 (2010).
11. Mirabello, L. *et al.* A comprehensive candidate gene approach identifies genetic variation associated with osteosarcoma. *BMC Cancer* **11**, 209 (2011).
12. Savage, S.A. & Mirabello, L. Using epidemiology and genomics to understand osteosarcoma etiology. *Sarcoma* **2011**, 548151 (2011).
13. Postel-Vinay, S. *et al.* Common variants near *TARDBP* and *EGR2* are associated with susceptibility to Ewing sarcoma. *Nat. Genet.* **44**, 323–327 (2012).
14. Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1006–1010 (2009).
15. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
16. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the *ABO* locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41**, 986–990 (2009).
17. Rothman, N. *et al.* A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.* **42**, 978–984 (2010).
18. Landi, M.T. *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* **85**, 679–691 (2009).
19. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
20. Li, J. *et al.* GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum. Mol. Genet.* **22**, 1457–1464 (2013).
21. Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201–208 (2011).
22. Qayyum, R. *et al.* A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. *PLoS Genet.* **8**, e1002491 (2012).
23. Slager, S.L. *et al.* Common variation at 6p21.31 (*BAK1*) influences the risk of chronic lymphocytic leukemia. *Blood* **120**, 843–846 (2012).
24. Rapley, E.A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* **41**, 807–810 (2009).
25. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
27. Molyneux, S.D. *et al. Prkar1a* is an osteosarcoma tumor suppressor that defines a molecular subclass in mice. *J. Clin. Invest.* **120**, 3310–3325 (2010).
28. Griffin, K.J. *et al.* A transgenic mouse bearing an antisense construct of regulatory subunit type 1A of protein kinase A develops endocrine and other tumours: comparison with Carney complex and other PRKAR1A induced lesions. *J. Med. Genet.* **41**, 923–931 (2004).

29. Cowan, R.W., Seidlitz, E.P. & Singh, G. Glutamate signaling in healthy and diseased bone. *Front. Endocrinol.* (*Lausanne*) **3**, 89 (2012).
30. Skerry, T.M. The role of glutamate in the regulation of bone mass and architecture. *J. Musculoskelet. Neuronal Interact.* **8**, 166–173 (2008).
31. Kalariti, N., Lembessis, P. & Koutsilieris, M. Characterization of the glutametergic system in MG-63 osteoblast-like osteosarcoma cells. *Anticancer Res.* **24**, 3923–3929 (2004).
32. Chang, H.J. *et al.* Metabotropic glutamate receptor 4 expression in colorectal carcinoma and its prognostic significance. *Clin. Cancer Res.* **11**, 3288–3295 (2005).
33. Brocke, K.S. *et al.* Glutamate receptors in pediatric tumors of the central nervous system. *Cancer Biol. Ther.* **9**, 455–468 (2010).
34. Stepulak, A. *et al.* Expression of glutamate receptor subunits in human cancers. *Histochem. Cell Biol.* **132**, 435–445 (2009).
35. Luksch, H. *et al.* Silencing of selected glutamate receptor subunits modulates cancer growth. *Anticancer Res.* **31**, 3181–3192 (2011).
36. Hirschhorn, J.N. & Lettre, G. Progress in genome-wide association studies of human height. *Horm. Res.* **71** (suppl. 2), 5–13 (2009).
37. Wang, Z. *et al.* Improved imputation of common and uncommon SNPs with a new reference set. *Nat. Genet.* **44**, 6–7 (2012).
38. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).

ONLINE METHODS

Genome-wide SNP genotyping. Genome-wide SNP genotyping of osteosarcoma cases was conducted using the Illumina OmniExpress BeadChip at the NCI Cancer Genomics Research Laboratory (CGR) in the Division of Cancer Epidemiology and Genetics (DCEG) at the National Cancer Institute. Genotype analysis occurred in two stages because of the sequential receipt of samples. Stage 1a consisted of 910 unique cases from the Children's Oncology Group (COG), Toronto Study, PamplonaCUN, Rizzoli (Italy), Ankara (Turkey) and UCL. Adult, cancer-free controls were drawn from previously scanned studies in the United States, namely, PLCO and CPS-II of the American Cancer Society, both scanned on the Illumina Omni 2.5M SNP microarray. European controls were drawn from EAGLE (Italy) scanned on the HumanHap550 array, WTCCC (UK) scanned on the HumanHap550 array and the NCI Spanish Bladder Cancer Study scanned on the HumanHap 1M SNP array. Stage 1b consisted of 218 unique cases scanned on the OmniExpress BeadChip drawn from the UCL, Rizzoli and PamplonaCUN studies plus 2 additional studies from the Sir Peter MacCallum Department of Oncology (Australia) and Brazil. Cancer-free adult controls were drawn from PLCO and NHS, scanned on the Illumina Omni 2.5M and OmniExpress arrays, respectively.

Each participating study obtained informed consent from study participants and approval of the study from its IRB; studies also obtained IRB certification permitting data sharing in accordance with the US National Institutes of Health (NIH) Policy for Sharing of Data Obtained in NIH-Supported or -Conducted Genome-Wide Association Studies (GWAS). The CGEMS data portal provides access to individual-level data from the National Cancer Institute scan only to investigators from certified scientific institutions after approval of their submitted Data Access Request.

Quality control assessment. We conducted systematic quality control that included steps specific for the scanning of different arrays at distinct times. For SNP assays, exclusions included SNPs with less than a 90% completion rate and those with extreme deviation from fitness for Hardy-Weinberg proportion ($P < 1 \times 10^{-7}$). There were 29 duplicated cases in stages 1a and 1b; concordance rates were 99.96%.

Analysis of stage 1a. In the quality control analysis of stage 1a, samples were excluded on the basis of (i) completion rates lower than 94% ($n = 28$ samples); (ii) abnormal heterozygosity values of less than 20% or greater than 31% ($n = 8$); (iii) expected duplicates ($n = 23$ pairs); (iv) abnormal X-chromosome heterozygosity ($n = 1$); and (v) phenotype exclusion (due to ineligibility or incomplete information) ($n = 57$). Genotypes for all subject pairs were also examined for close relationships (presence of first- and second-degree relatives) using the GLU (Genotyping Library and Utilities version 1.0) qc.ibds module with an IBD0 threshold of 0.70; no first-degree relatives were identified in the cases scanned.

Using a set of 12,000 unlinked SNPs (pairwise $r^2 < 0.004$) common to the GWAS chips used herein[33], we excluded 264 subjects with less than 80% European ancestry, as determined using STRUCTURE analysis[34] and PCA[35]; a majority of these represent 97 cases of African-American ancestry and 99 cases of Hispanic ancestry. The final association analysis for stage 1a included 596 cases and 2,703 controls of European ancestry. After quality control filtering, data from 698,968 SNPs were available. The numbers of SNPs overlapping those of pooled controls were 510,856, 510,856, 310,384 and 304,092 for the US, Spanish, Italian and UK components, respectively, and these SNPs were used in the downstream association analyses.

Analysis of stage 1b. Similar quality control metrics were applied to the second set of additional scanned cases, stage 1b. For the current analysis, we included 98 additional cases in the association analysis of individuals of European ancestry.

Quality control metrics for previously scanned cancer-free adult controls. Pooled PLCO controls were scanned on the Illumina Omni 2.5M array. We excluded (i) samples with missing rates of >6%; (ii) SNPs with missing rates of >10%; (iii) samples with mean heterozygosity of >21% or <16%; (iv) one sample from each cryptic related pair; and (v) non-CEU admixed individuals, identified by STRUCTURE analysis. Pooled SPBC controls were scanned on the Illumina 1M SNP array; details of genotype quality control were previously described[17]. Pooled EAGLE controls were scanned on the Illumina 550K SNP array; details of genotype quality control were previously published[18]. Pooled WTCCC controls were previously described[39].

Replication and TaqMan genotyping. We conducted a follow-up analysis of the 30 SNPs selected for replication (based on stage 1a data) using TaqMan genotyping assays (ABI) that were optimized in the CGR pipeline. We analyzed a total of 247 cases and 550 controls; the studies included 99 histologically confirmed osteosarcoma cases and 65 hospital-based cancer-free pediatric controls from the National Osteosarcoma Etiology Study Group plus 148 cases drawn from the studies used in the scan in which there was insufficient DNA for array analysis. Cancer-free controls were also selected from the PLCO and NHS studies (previously analyzed on the Illumina Omni 2.5M and OmniExpress arrays, respectively).

Statistical analysis. Associations between SNPs and risk of osteosarcoma were estimated using unconditional logistic regression by OR and 95% CI, using multivariate unconditional logistic regression assuming a trend genetic model (with the effect of the variant included in a log-additive model with 1 degree of freedom). When included in the null model, PCA identified four significant ($P < 0.05$) eigenvectors. The main effect model was adjusted by sex, country and four eigenvectors, identified on the basis of significance ($P < 0.05$) observed in the null model. For the replication study, the analysis was adjusted only for sex.

For stage 1 (1a and 1b), in addition to joint analysis, a fixed-effect meta-analysis was performed to combine results for cases from each country (component) to facilitate the filtering out of artifacts by checking for heterogeneity across countries. Similar meta-analysis was also performed to combine scan (discovery) association results with TaqMan (replication) association results for the 30 SNPs that were analyzed in both stages.

The estimated inflation factor λ and the adjusted $\lambda_{1,000}$ inflation factor for the test statistic from combined stage 1a and 1b data were 1.036 and 1.033, respectively. Recombination hotspots were identified in the vicinity of osteosarcoma-associated loci at 6p21.31 and 2p25.2 using SequenceLDhot[40], a program that uses an approximate marginal likelihood method[41] and calculates likelihood ratio statistics at a set of possible hotspots. We tested 5 unique sets of 100 control samples. We used the PHASE v2.1 program to calculate background recombination rates[42,43]. The LD heatmap was visualized in r^2 using the snp.plotter R program[44].

Imputation. To further interrogate the loci associated with osteosarcoma, we imputed additional SNPs within 1 Mb of either side of each implicated SNP using IMPUTE2 software and reference data from both the 1000 Genomes Project[13] and the DCEG Imputation Reference Set version 1 (ref. 37). We used SNPTEST to analyze the posterior SNP dosages from IMPUTE2, adjusting for sex, country and four eigenvectors, as described[45].

Data analysis. Data analysis and management were performed with GLU, PLINK, EIGENSTRAT, IMPUTE2 and SNPTEST.

39. Purdue, M.P. *et al*. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat. Genet.* **43**, 60–65 (2011).
40. Fearnhead, P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics* **22**, 3061–3066 (2006).
41. Fearnhead, P. & Donnelly, P. Approximate likelihood methods for estimating local recombination rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64**, 657–680 (2002).
42. Crawford, D.C. *et al*. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36**, 700–706 (2004).
43. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
44. Luna, A. & Nicodemus, K.K. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics* **23**, 774–776 (2007).
45. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

doi:10.1038/ng.2645