nature genetics

Frequent mutation of the major cartilage collagen gene *COL2A1* in chondrosarcoma

Patrick S Tarpey[1,8], Sam Behjati[1,2,8], Susanna L Cooke[1], Peter Van Loo[1,3], David C Wedge[1], Nischalan Pillay[4,5], John Marshall[1], Sarah O'Meara[1], Helen Davies[1], Serena Nik-Zainal[1], David Beare[1], Adam Butler[1], John Gamble[1], Claire Hardy[1], Jonathon Hinton[1], Ming Ming Jia[1], Alagu Jayakumar[1], David Jones[1], Calli Latimer[1], Mark Maddison[1], Sancha Martin[1], Stuart McLaren[1], Andrew Menzies[1], Laura Mudie[1], Keiran Raine[1], Jon W Teague[1], Jose M C Tubio[1], Dina Halai[4], Roberto Tirabosco[4], Fernanda Amary[4], Peter J Campbell[1,6,7], Michael R Stratton[1], Adrienne M Flanagan[4,5] & P Andrew Futreal[1]

Chondrosarcoma is a heterogeneous collection of malignant bone tumors and is the second most common primary malignancy of bone after osteosarcoma. Recent work has identified frequent, recurrent mutations in *IDH1* or *IDH2* in nearly half of central chondrosarcomas. However, there has been little systematic genomic analysis of this tumor type, and, thus, the contribution of other genes is unclear. Here we report comprehensive genomic analyses of 49 individuals with chondrosarcoma (cases). We identified hypermutability of the major cartilage collagen gene *COL2A1*, with insertions, deletions and rearrangements identified in 37% of cases. The patterns of mutation were consistent with selection for variants likely to impair normal collagen biosynthesis. In addition, we identified mutations in *IDH1* or *IDH2* (59%), *TP53* (20%), the RB1 pathway (33%) and Hedgehog signaling (18%).

We subjected 49 cases of untreated chondrosarcoma, including 30 central, 4 peripheral and 14 dedifferentiated tumors and 1 tumor arising in an individual with synovial chondromatosis, along with matching normal tissues to whole-exome sequencing as previously described[1]. Dedifferentiated chondrosarcoma is characterized by biphasic histology represented by conventional chondrosarcoma abutting a high-grade non-cartilaginous component. For dedifferentiated chondrosarcoma, we obtained sequence from the high-grade non-cartilaginous component of the tumor in each case. Case distribution by grade is shown in **Supplementary Table 1**. Exomic sequence coverage at a minimum depth of 30× was 65–70%, and data were processed as previously described[1]. We confirmed sequence variants as somatic using custom targeted capture and sequencing combined with manual data inspection. Candidate genes were assessed in an extension series of 26 cases (**Supplementary Table 1**). Allele-specific copy number was assessed with SNP6.0 arrays and direct analyses of exome reads as previously described[1,2].

In total, we identified 1,428 somatic mutations, with somatic mutation burden ranging from 1 to 115 per case (**Supplementary Table 2**). Mutations comprised 944 missense, 61 nonsense, 37 essential splice-site and 301 synonymous changes, 80 indels and 5 substitutions in microRNAs. Somatic mutation burden showed significant association with higher tumor grade. High-grade tumors (grades 2 and 3 and dedifferentiated) had on average more than double the number of somatic mutations per sample as grade 1 chondrosarcoma (Wilcoxon rank-sum test $P = 0.00017$).

The most notable finding of this study was the identification of mutations in *COL2A1*, which encodes the α chain of type II collagen fibers, the major collagen constituent of articular cartilage. We initially identified somatic mutations in eight cases from the exome screen that were all small indels. This striking pattern of mutations prompted a more thorough investigation of *COL2A1* to look for rearrangements in addition to point mutations. We tiled the entire gene footprint with baits for custom capture and subjected all exome cases as well as the 26 follow-up tumors to capture and sequencing. In total, 44% (33/75) of cases had at least 1 mutation that was predicted to affect coding sequence (**Fig. 1**, **Supplementary Fig. 1** and **Supplementary Table 3**). The mutations consisted of splice-site changes ($n = 2$), indels ($n = 23$), missense changes ($n = 2$) and large-scale rearrangements ($n = 36$). We identified no synonymous mutations (**Supplementary Table 3**). The majority of rearrangements had both breakpoints located within the footprint of *COL2A1* (**Supplementary Fig. 2**), and none was predicted to result in an in-frame fusion event. Whole-genome shotgun sequencing at low depth in six of the chondrosarcoma cases showed that breakpoints were clustered in *COL2A1* and that chromosome 12 was not highly rearranged elsewhere (**Supplementary Fig. 2**). In 13 cases, we identified a single *COL2A1* mutation; however,

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [2]Department of Paediatrics, University of Cambridge, Cambridge, UK. [3]Human Genome Laboratory, Department of Human Genetics, VIB and KU Leuven, Leuven, Belgium. [4]Histopathology, Royal National Orthopaedic Hospital National Health Service (NHS) Trust, Stanmore, UK. [5]University College London (UCL) Cancer Institute, London, UK. [6]Department of Haematology, Addenbrooke's Hospital, Cambridge, UK. [7]Department of Haematology, University of Cambridge, Cambridge, UK. [8]These authors contributed equally to this work. Correspondence should be addressed to P.A.F. (afutreal@mdanderson.org).

Received 29 December 2012; accepted 16 May 2013; published online 16 June 2013; doi:10.1038/ng.2668

© 2013 Nature America, Inc. All rights reserved.

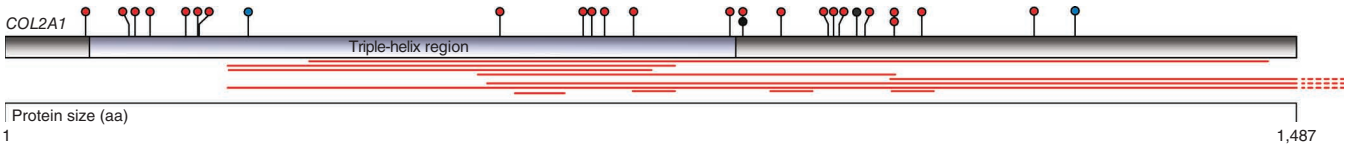

Figure 1 Type and location of *COL2A1* mutations in chondrosarcoma. Somatic mutations identified in the primary and extension investigations are indicated as circles; different colors indicate truncating (red), essential splice-site (black) and missense (blue) mutations. Large deletions are depicted by red bars beneath the schematic.

20 tumors had more than 1 mutation in this gene. A one-sided Fisher's exact test indicated that, of all cases, high-grade tumors (grades 2 and 3 and dedifferentiated) were significantly more likely to contain a *COL2A1* mutation than low-grade tumors (grade 1) ($P = 0.029$). This association between *COL2A1* mutation and tumor grade, however, was not significant when only central chondrosarcoma cases were considered ($P = 0.093$).

Although finding 20 cases with more than 1 *COL2A1* mutation is formally compatible with a loss-of-function mechanism, the presence of multiple cases with a single heterozygous mutation suggests that functional consequences at the protein level may be more complex. We therefore performed immunohistochemistry for COL2A1 protein on 43 available tumors (22 mutation positive and 21 wild type; **Supplementary Table 4**). We observed abnormal staining (absent or focal compared to the diffuse staining seen in normal cartilage; **Supplementary Fig. 3**) in 31 of 43 tumors, with absence of staining in 8 tumors and focal staining in 23. Of the eight cases with no staining, six encoded mutant COL2A1, with only one having two mutations identified in the gene. Of the 23 cases showing focal staining, 12 encoded mutant COL2A1 (2 cases with more than 1 mutation), and 11 had no detectable mutation. It is possible that *COL2A1* mutations in a proportion of the seemingly wild-type cases have not been detected, in particular, given the complex nature of the rearrangements that predominate, or have otherwise aberrant type II collagen deposition for reasons that remain obscure. These data argue that complete loss of protein is not the predominant outcome of *COL2A1* mutation.

We also sequenced the entire *COL2A1* gene in 56 osteosarcomas, 24 chordomas, 10 other cartilaginous tumors and 73 meningiomas to compare patterns of mutation. The results showed an accumulation of mutations across the gene footprint of *COL2A1* that was restricted to chondrosarcomas (**Supplementary Table 5**). This pattern suggested that *COL2A1* is hypermutable in chondrosarcoma. Given that *COL2A1* is transcribed at a substantial rate in chondrocytes, the data may support the notion of a transcription-associated mutation or recombination (TAM or TAR) mechanism[3]. Elevated *COL2A1* mutation rate does not preclude mutations from having biological consequences. Analyses of the patterns of mutation provided evidence of selection. No silent coding mutations were found. There was a preponderance of frameshifting indels compared to substitution mutations in exons (**Supplementary Table 6a**). In *COL2A1* introns, these mutation classes were in equal proportion. Considering only indels, there were very few mutations in coding sequence that were multiples of three (in frame) compared to the noncoding portions of the gene (**Supplementary Table 6b**). These data suggest positive selection for mutations predicted to substantially disrupt *COL2A1* coding potential. To further explore the specificity of *COL2A1* mutation for chondrosarcoma, mutations across all collagen genes were assessed in multiple bone and cartilage tumors, including chondroblastomas ($n = 6$), chondromyxoid fibromas ($n = 2$), chordomas ($n = 24$) and osteosarcomas ($n = 56$). The results clearly indicate an enrichment for nonsynonymous mutations that was restricted to *COL2A1* in chondrosarcomas (**Supplementary Fig. 4**).

Mature collagen fibrils are formed via the assembly of pro-collagen α chains into triple-helix formations that are stabilized through post-translational modifications. Constitutive mutations of *COL2A1* lead to a variety of skeletal and ocular disorders that range from lethal perinatal dwarfism, due to *de novo* missense mutations in achondrogenesis type II (MIM 200610), to Stickler syndrome type I (MIM 108300), caused primarily by familial truncating mutations. The variety of *COL2A1* somatic mutations reported here in chondrosarcoma would likely lead to disruption of the collagen maturation process through the production of aberrant pro-collagen α chains, which might have dominant-negative effects, given the heterozygous nature of many of the mutations. In a mouse model carrying a dominant-negative *Col2a1* mutation identified in the human disease spondyloepiphyseal dysplasia congenita (MIM 183900), depletion of collagen fibrils was found to significantly impair chondrocyte differentiation[4]. However, there is no reported increase in predisposition to chondrosarcoma in these human developmental disorders nor in their mouse models, similar to the lack of cancer predisposition found for the germline alleles of fibroblast growth factor receptor genes that give rise to other skeletal disorders. Of note, targeted *Pten* deficiency in mouse chondrocytes leads to dyschondroplasia resembling human enchondroma[5]. A dedifferentiated chondrosarcoma (PD6363a) found to have homozygous deletion of *PTEN* did not have *COL2A1*, *IDH1* or *IDH2* mutation. Lastly, of various extracellular matrix components, type 2 collagen was found to restore cartilaginous features of otherwise dedifferentiated primary chondrocytes in monolayer culture[6]. Thus, there is a rational basis to hypothesize that *COL2A1* mutations in chondrosarcoma might not be merely passenger events but might bring about fundamental perturbation of matrix deposition and signaling that may contribute to oncogenesis through the abrogation of normal differentiation programs. These data indicate that the development of *in vivo* approaches for mechanistic investigation of *COL2A1* in chondrosarcoma will be important.

As we previously reported, *IDH1* and *IDH2* mutations are prevalent in central chondrosarcoma[7] and absent from peripheral chondrosarcoma. We detected an *IDH1* or *IDH2* mutation in 29 of 49 tumors (59%) (**Fig. 2**). A Fisher's exact test of difference in the proportion of tumors of different grades bearing an *IDH1* or *IDH2* mutation did not give a significant result ($P = 0.22$). Given the active development of inhibitors for mutant IDH proteins[8], chondrosarcoma would be a compelling opportunity, as there are currently no effective treatments other than radical surgical resection. In addition, evaluation of 2-hydroxyglutarate, the oncometabolite produced by mutant IDH proteins, as a biomarker[9] in chondrosarcoma management is warranted.

We also identified frequent involvement of the RB1 pathway in chondrosarcoma (**Figs. 2** and **3a**). Copy number analyses identified 13 tumors with homozygous deletions of *CDKN2A*, consistent with previous reports[10] (**Supplementary Fig. 5** and **Supplementary Table 7**). We also identified a *CDKN2A* missense mutation encoding p.Asp108Gly, with loss of the wild-type allele, as well as two further truncating mutations in the extension tumor series. We found focal

© 2013 Nature America, Inc. All rights reserved.

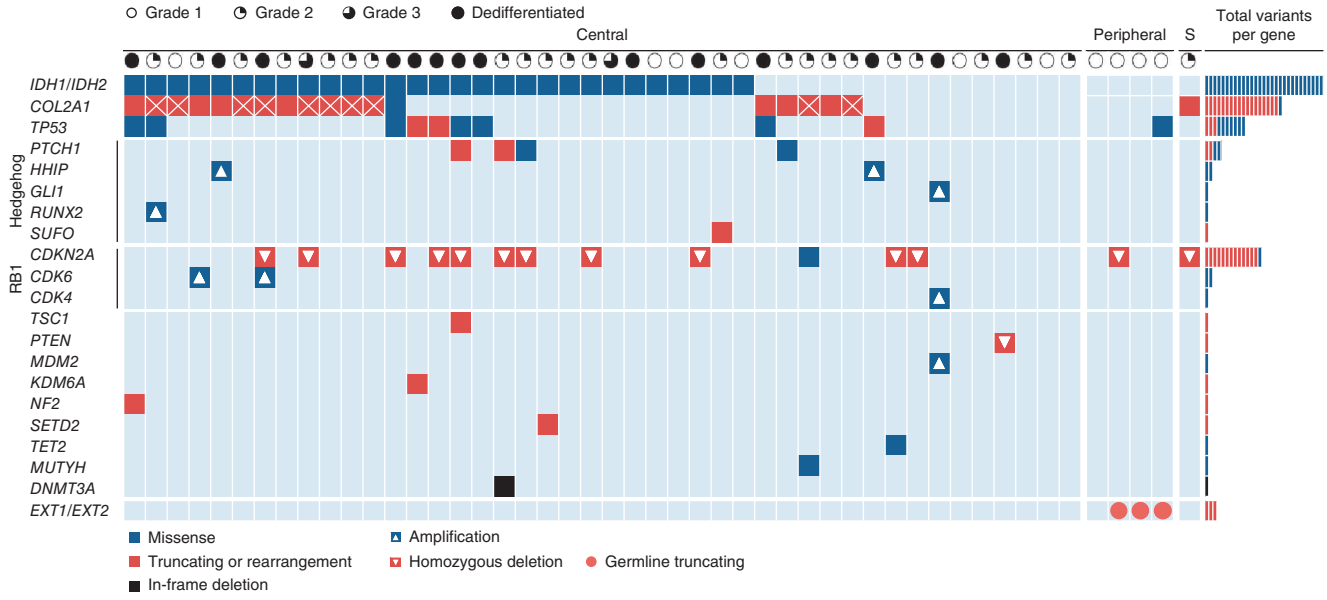

Figure 2 Known and likely driver variants identified in 49 primary tumors. Mutated genes are shown in rows, and the 49 tumors are presented in columns. Clinical grade for each tumor is indicated above the column. Tumors are classified as central or peripheral, and one tumor was a malignant chondrosarcoma that arose in a case with synovial chondromatosis (S). Statistical analysis to identify non-random accumulation of nonsynonymous substitutions confirmed *IDH1* as being significantly mutated (Online Methods). Types of variants are depicted with shaded squares. Boxes with crosses indicate samples with more than one *COL2A1* mutation.

amplifications of *CDK4* (one case) and *CDK6* (two cases), with one of the cases with *CDK6* amplification also having a *CDKN2A* mutation. *CDK4* was amplified together with *MDM2*, as has previously been reported[11]. *CDK6*, to the best of our knowledge, has not previously been found to be amplified in chondrosarcoma. Altogether, 33% (16/49) of exome-series cases had mutations affecting RB regulatory constituents. *TP53* mutation have a key role in chondrosarcoma, with 20% (10/49) of cases having coding mutations. Mutations of other known cancer-related genes included ones in *SETD2*, *KDM6A*, *NF2*, *TET2*, *DNMT3A* and *TSC1*. Germline *EXT1* or *EXT2* mutation was identified in three grade 1 peripheral chondrosarcoma. PD7299a, the case with the largest substitution burden, had a somatic *MUTYH* missense mutation previously reported in the context of colorectal cancer susceptibility in familial adenomatous polyposis 2 (MIM 608456)[12–15].

We found evidence for the involvement of the Indian Hedgehog (IHH) signaling pathway in chondrosarcoma (**Figs. 2** and **3b**). IHH signaling is crucial for normal chondrocyte differentiation. Mutations resulting in constitutive Hedgehog signaling are causal in benign cartilage tumors[16–18]. We identified four mutations of *PTCH1* in the exome screen (two missense and two truncating) and a *PTCH1* missense mutation in the extension tumor series (**Fig. 2** and **Supplementary Table 2**). A single inactivating *SUFU* mutation and a single *GLI1* amplification were also identified in separate cases. Further, we identified focal amplifications of *RUNX2* and *HHIP*, which are both closely linked to IHH signaling[16,17]. Taken together, there is mutational evidence for IHH pathway activation in 18% (9/49) of exome-series cases. Hedgehog pathway inhibition via the small molecule vismodegib targeting SMO has recently been shown to have significant activity in advanced basal cell carcinoma, including patients with basal-cell nevus syndrome due to Hedgehog pathway mutation[18]. Assessment of Hedgehog pathway inhibitors in the subset of patients with chondrosarcoma in whom mutations in this pathway have been identified may be an important line of therapeutic investigation.

Combined with our previous efforts identifying *IDH1* and *IDH2* mutations in central chondrosarcoma, the comprehensive analysis presented here begins to paint a more complete picture of chondrosarcoma genomics for the most common and most lethal subtypes. Aberrant RB and IHH pathways and fundamental perturbation of matrix biology via *COL2A1* mutation, as discovered here, have been shown to likely contribute to chondrosarcoma. In addition to exploiting *IDH1* and *IDH2* mutations in therapeutic and biomarker contexts, delineating the contributions of RB and IHH pathway mutations presents opportunities for focused therapeutic exploration. Further, if the production of mature collagen fibrils is indeed compromised by mutant *COL2A1*, then there may be an opportunity to explore therapeutic strategies exploiting cellular and endoplasmic reticulum stress responses to improperly folded proteins. This study demonstrates

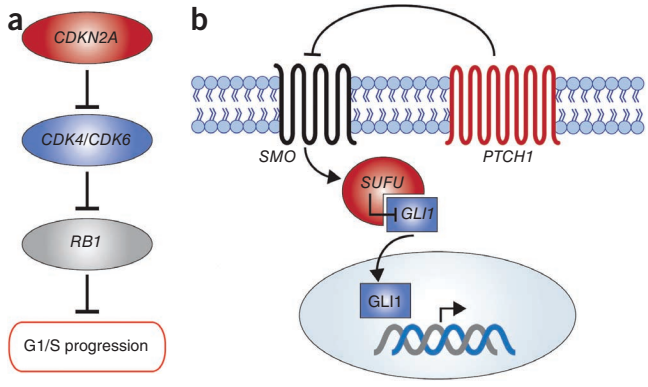

Figure 3 Genes mutated in chondrosarcoma from the RB1 and Hedgehog pathways. (**a**) RB1 pathway, indicating genes with mutations in the current tumor series. (**b**) Hedgehog pathway, indicating known cancer-related genes mutated in this tumor series. Genes in blue are activated by mutations, and genes in red are inactivated.

© 2013 Nature America, Inc. All rights reserved.

the need to comprehensively characterize all cancer types and subtypes to generate maximum opportunity for comparative oncogenomics and, crucially, expedited development of therapeutic and diagnostic strategies.

URLs. European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega/; Sanger Institute SNP array data, http://www.sanger.ac.uk/cgi-bin/genetics/CGP/cghviewer/CghHome.cgi.

METHODS

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

Accession codes. Sequencing data and SNP6 array data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI), under accessions EGAD00001000358 and EGAD00010000432, respectively.

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

ACKNOWLEDGMENTS

We are grateful to the patients for participating in the research and to the clinicians and support staff of the London Sarcoma Service involved in their care. This work was supported by funding from the Wellcome Trust (grant 077012/Z/05/Z) and the Skeletal Cancer Action Trust (SCAT), UK. Material was obtained from the Royal National Orthopaedic Hospital Musculoskeletal Research Program and Biobank. Support was provided to A.M.F. by the National Institute for Health Research, the University College London Hospital Biomedical Research Centre and the UCL Experimental Cancer Centre. P.J.C. is personally funded through a Wellcome Trust Senior Clinical Research Fellowship (grant WT088340MA). P.V.L. is a postdoctoral researcher of the Research Foundation–Flanders (FWO). S.B. is funded through the Wellcome Trust PhD Programme for Clinicians.

AUTHOR CONTRIBUTIONS

P.S.T. and S.B. analyzed the sequence data. S.L.C. and J.M.C.T. performed rearrangement analysis. P.V.L. analyzed the SNP6 array data. D.C.W. performed statistical analyses. S. McLaren, D.H. and S.O. coordinated sample processing and technical investigations. S. Martin coordinated sample acquisition and processing. C.H., C.L., L.M. and M.M. performed technical investigations. A.B., J.G., J.H., M.M.J., A.J., D.J., A.M., J.M., K.R., J.W.T., H.D., S.N.-Z. and D.B. performed informatic investigations. F.A., R.T., N.P. and A.M.F. provided samples and clinical data and performed immunohistochemistry. P.J.C., M.R.S., A.M.F. and P.A.F. directed the research and contributed to the manuscript. P.A.F. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Stephens, P.J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
2. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
3. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.* **13**, 204–214 (2012).
4. Barbieri, O. *et al.* Depletion of cartilage collagen fibrils in mice carrying a dominant negative *Col2a1* transgene affects chondrocyte differentiation. *Am. J. Physiol. Cell Physiol.* **285**, C1504–C1512 (2003).
5. Yang, G. *et al.* PTEN deficiency causes dyschondroplasia in mice by enhanced hypoxia-inducible factor 1α signaling and endoplasmic reticulum stress. *Development* **135**, 3587–3597 (2008).
6. Chiu, L.H. *et al.* Differential effect of ECM molecules on re-expression of cartilaginous markers in near quiescent human chondrocytes. *J. Cell Physiol.* **226**, 1981–1988 (2011).
7. Amary, M.F. *et al. IDH1* and *IDH2* mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J. Pathol.* **224**, 334–343 (2011).
8. Popovici-Muller, J. *et al.* Discovery of the first potent inhibitors of mutant IDH1 that lower tumor 2-HG *in vivo*. *ACS Med. Chem. Lett.* **3**, 850–855 (2012).
9. Fathi, A.T. *et al.* Prospective serial evaluation of 2-hydroxyglutarate, during treatment of newly diagnosed acute myeloid leukemia, to assess disease activity and therapeutic response. *Blood* **120**, 4649–4652 (2012).
10. Hallor, K.H. *et al.* Genomic profiling of chondrosarcoma: chromosomal patterns in central and peripheral tumors. *Clin. Cancer Res.* **15**, 2685–2694 (2009).
11. Schrage, Y.M. *et al.* Central chondrosarcoma progression is associated with pRb pathway alterations: CDK4 down-regulation and p16 overexpression inhibit cell growth *in vitro*. *J. Cell Mol. Med.* **13**, 2843–2852 (2009).
12. Aceto, G. *et al.* Mutations of *APC* and *MYH* in unrelated Italian patients with adenomatous polyposis coli. *Hum. Mutat.* **26**, 394 (2005).
13. Ahn, J. *et al.* Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (*EXT1*). *Nat. Genet.* **11**, 137–143 (1995).
14. Hopyan, S. *et al.* A mutant PTH/PTHrP type I receptor in enchondromatosis. *Nat. Genet.* **30**, 306–310 (2002).
15. Stickens, D. *et al.* The *EXT2* multiple exostoses gene defines a family of putative tumour suppressor genes. *Nat. Genet.* **14**, 25–32 (1996).
16. Bruce, S.J. *et al.* Inactivation of *Patched1* in the mouse limb has novel inhibitory effects on the chondrogenic program. *J. Biol. Chem.* **285**, 27967–27981 (2010).
17. Wuelling, M. & Vortkamp, A. Transcriptional networks controlling chondrocyte proliferation and differentiation during endochondral ossification. *Pediatr. Nephrol.* **25**, 625–631 (2010).
18. Tang, J.Y. *et al.* Inhibiting the hedgehog pathway in patients with the basal-cell nevus syndrome. *N. Engl. J. Med.* **366**, 2180–2188 (2012).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

Subject samples. Informed consent was obtained from all subjects, and ethical approval was obtained from the Cambridgeshire 2 Research Ethics Service (reference 09/H0308/165). Collection and use of subject samples were approved by the appropriate institutional review board (IRB) of each institution. In addition, this study and use of its collective materials had specific IRB approval.

Exome enrichment and sequencing. Genomic libraries were prepared using the Illumina Paired-End Sample Prep kit following the manufacturer's instructions. Enrichment was performed as described previously[19] using the Agilent SureSelect Human All Exon 50 Mb kit following the manufacturer's recommended protocol, with the exception that enrichment was not performed before PCR amplification. Each exome was sequenced using the 75-bp paired-end protocol on an Illumina HiSeq DNA Analyzer to produce approximately 10 Gb of sequence per exome. Sequencing reads were aligned to the human genome (NCBI Build 37) using the Burrows-Wheeler Aligner (BWA) algorithm with default settings[20]. Reads that were unmapped, PCR-derived duplicates and reads mapping outside the targeted region of the genome were excluded from the analysis. Average exome sequence coverage at 30× or higher was 70%.

Enrichment and sequencing of target regions. A custom-made bait set (Agilent) was designed to capture and enrich genomic regions of putative somatic variants and selected genes for follow-up, including *CDKN1A*, *CDKN2A*, *COL11A1*, *COL11A2*, *COL2A1*, *COL9A1*, *COL9A2*, *COL9A3*, *DNMT3A*, *EXT1*, *EXT2*, *IDH1*, *IDH2*, *NF2*, *PTCH1*, *RB1*, *SETD2*, *SUFU*, *TET2* and *TP53*. The bait set was applied to the 49 primary tumor samples, to a chondrosarcoma follow-up series of 26 cases and to non-chondrosarcoma tumors. Enrichment and sequencing were performed as for the exomes.

Rearrangement screen. Six tumors (PD7054a, PD7061a, PD9388a, PD9390a, PD9391a and PD9394a) underwent low-sequence-depth whole-genome shotgun sequencing to identify rearrangements, as previously described[21], with the exception that a read length of 50 bp was used instead of 37 bp.

Variant detection. The CaVEMan (cancer variants through expectation maximization) algorithm was used to call single-nucleotide substitutions[1]. This algorithm uses a naive Bayesian classifier to estimate the posterior probability of each possible genotype (wild type compared to germline or somatic mutation) at each base. We applied several post-processing filters to the set of initial CaVEMan mutation calls to remove variants reported in poor quality sequence and to improve the specificity of the output.

To call insertions and deletions, we used split-read mapping implemented as a modification of the Pindel algorithm[22]. Pindel searches for reads with one end anchored on the genome and the other end mapped with high confidence in two (split) portions, spanning a putative indel. Post-processing filters were applied to the output to improve specificity. To call rearrangements, we applied the Brass (breakpoint via assembly) algorithm, which identifies rearrangements by grouping discordant read pairs that represent the same breakpoint event[23]. Mutations were annotated to Ensembl version 58.

Variant validation. All putative somatic variants identified in the primary screen of 49 tumors were subjected to targeted capture and were sequenced. Variants for which additional reads could not be generated were validated by manual inspection of sequencing data from the initial exome screen. *COL2A1* rearrangements were validated, algorithmically or manually, by defining the exact location of the breakpoint at nucleotide-level resolution and by extracting split reads across the breakpoint. Additionally, 14 arbitrarily chosen breakpoints (39%) were validated by PCR (primer sequences available upon request). Rearrangements other than those affecting *COL2A1* were not further validated.

Detection of copy number variation. SNP array hybridization on the SNP6.0 platform was carried out according to Affymetrix protocols and as described at the Sanger Institute website (see URLs) for 36 of 49 tumors from the exome screen. For the remaining 13 cases, copy number data were derived directly from exome sequencing data.

Copy number analysis was performed using ASCAT (version 2.2)[2], taking into account non-neoplastic cell infiltration and tumor aneuploidy, and resulted in integral allele-specific copy number profiles for the tumor cells[5]. Homozygous deletions were called if there were zero copies in the tumor cells. Amplifications were defined as previously described[6]. Measures of non-neoplastic cell infiltration and tumor aneuploidy were derived by ASCAT[5].

Immunohistochemistry. Immunohistochemistry was performed using standard techniques, including the chromogen diaminobenzidine, while employing the Bond-Max Autostainer (Novocastra), as previously described[24]. A monoclonal mouse primary antibody to human COL2A1 was used (Santa Cruz Biotechnology, sc-59958; 1:50 dilution). Semiquantitative assessment of staining was performed independently by two histopathologists.

Statistical analyses. The significance of the association between tumor grade and overall somatic mutation burden was tested using the Wilcoxon rank-sum test. The significance of the association between *COL2A1* mutation status and tumor grade was tested using one-sided Fisher's exact tests. The significance of the difference in the proportion of tumors of different grades bearing an *IDH1* or *IDH2* mutation was determined by Fisher's exact test and confirmed by χ^2 test for trend. The statistical significance of recurrence in mutated genes was assessed using a permutation method. First, the regions that were pulled down by baits were identified as those with coverage in at least three of four non-cancerous samples, subject to the same pulldown procedure used in this study. Random mutations were generated *in silico* in these regions for each sample, with the number of simulated mutations equal to three times the number of substitutions observed in that sample. Simulated mutations were annotated using the Variant Effect Predictor[25]. Nonsynonymous mutations (those with an annotation of 'NON_SYNONYMOUS_CODING', 'ESSENTIAL_SPLICE_SITE', 'STOP_GAINED' or 'STOP_LOST') were randomly selected in the same numbers as observed in our samples. This procedure was repeated 100,000 times, and, for each mutated gene, the number of runs that contained at least as many mutated samples as were observed in our data set, n, was noted. The probability that each gene would be mutated in the observed number of samples by chance was then determined by $P = n/100{,}000$. P values were adjusted using the Bonferroni correction, after which the only recurrently mutated gene with significantly more mutated samples than expected by chance was *IDH1*.

19. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
20. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
21. Stephens, P.J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
22. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
23. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
24. Shalaby, A. *et al.* The role of epidermal growth factor receptor in chordoma pathogenesis: a potential therapeutic target. *J. Pathol.* **223**, 336–346 (2011).
25. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).

© 2013 Nature America, Inc. All rights reserved.