Title: Genomic and transcriptomic characterisation of undifferentiated pleomorphic sarcoma of bone

Naser M Ali[1], Stefania Niada[2*], Anna T Brini[2*], Mark R Morris[3], Sathishkumar Kurusamy[3], Abdullah Alholle[1], David Huen[3], Cristina R Antonescu[4], Franck Tirode[$], Vaiyapuri Sumathi[5#], Farida Latif[1#]

1. Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, U.K.

2. IRCCS Galeazzi, Department of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, Italy.

3. Research Institute in Healthcare Science, Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton, UK.

4. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY.

5. Department of Musculoskeletal Pathology, The Royal Orthopaedic Hospital. Robert Aitken Institute of Clinical Research, University of Birmingham, Edgbaston, Birmingham. U.K.

$ Univ Lyon, Université Claude Bernard Lyon 1, CNRS 5286, INSERM U1052, Cancer Research Center of Lyon, Department of Translational Research and Innovation, Centre Léon Bérard, Lyon, France

        * Co-2nd author, # Co-last author

Corresponding authors:

Vaiyapuri Sumathi, M.D., F.R.C.Path., Department of Musculoskeletal Pathology, The Royal Orthopaedic Hospital, Robert Aitken Institute of Clinical Research, University of Birmingham, Birmingham, B15 2TT, UK. Email: Vaiyapuri.sumathi@nhs.net

Farida Latif, PhD, Emeritus Professor, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK. Email: f.latif@bham.ac.uk

Running title: Genetic alterations landscape in UPSb tumours

Word count:

Abstract

Undifferentiated pleomorphic sarcoma of bone (UPSb), is a rare primary bone sarcoma that lacks a specific line of differentiation. There is very little information about the genetic alterations leading to tumourigenesis or malignant transformation. Distinguishing between UPSb and other malignant bone sarcomas, including dedifferentiated chondrosarcoma and osteosarcoma, can be challenging due to overlapping features. To explore the genomic and transcriptomic landscape of UPSb tumours, whole-exome sequencing (Giacomini, #274) and RNA Sequencing (RNA-Seq) were performed on UPSb tumours. All tumours lacked hotspot mutations in *IDH1/2* 132 or 172 codons, thereby excluding the diagnosis of dedifferentiated chondrosarcoma. Recurrent somatic mutations in *TP53* were identified in 4/14 samples (29%). Moreover, recurrent mutations in histone chromatin remodelling genes, including *H3F3A, ATRX* and *DOT1L,* were identified in 5/14 samples (36%), highlighting the potential role of deregulated chromatin remodelling pathways in UPSb tumourigenesis. The majority of recurrent mutations in chromatin remodelling genes identified here are reported in COSMIC, including the *H3F3A* G35 and K36 hotspot residues. Copy number alteration analysis identified gains and losses in genes that have been previously altered in UPSb or UPS of soft tissue. Eight somatic gene fusions were identified by RNA-Seq, two of which, *CLTC-VMP1* and *FARP1-STK24*, were reported previously in multiple cancers. Genomic characterization of these fusions revealed potential tumour suppressing or oncogenic roles. Hierarchical clustering analysis, using

RNA-Seq data, distinctly clustered UPSb tumours from osteosarcoma and other sarcomas, thus molecularly distinguishing UPSb from other sarcomas. RNA-Seq expression profiling analysis and quantitative RT-PCR showed an elevated expression in *FGF23* which can be a potential molecular biomarker in UPSb. To our knowledge, this study represents the first comprehensive WES and RNA-Seq analysis of UPSb tumours revealing novel protein-coding recurrent gene mutations, gene fusions and identifying a potential UPSb molecular biomarker, thereby broadening the understanding of the pathogenic mechanisms and highlighting the possibility of developing novel targeted therapeutics.

Keywords: whole exome sequencing, RNA sequencing, undifferentiated pleomorphic sarcoma of the bone, chromatin remodelling genes, gene fusions, sarcomas, FGF23, CNV, gene expression and hierarchical clustering analyses.

Introduction

Undifferentiated high-grade pleomorphic sarcoma of bone (UPSb) is a rare aggressive bone sarcoma that lacks a specific line or pattern of differentiation [1]. These tumours represent <2% of all primary malignant bone neoplasms and rarely occur in young adults [1,2]. Tumours commonly arise in long bones of lower extremities, particularly the femur followed by tibia and humerus, with a metastatic rate of at least 50%, especially to lungs [2,3]. The morphological appearance of the tumours is heterogeneous, consisting of atypical spindle and pleomorphic cells that lack matrix production. As UPSb is a diagnosis exclusion, thorough and extensive sampling of the tumour to rule out osteosarcoma or de-differentiated chondrosarcoma is mandatory [4]. This might pose diagnostic difficulties, particularly in a limited biopsy sample. The recommended treatment generally involves neoadjuvant therapy followed by wide surgical excision [5]. The chemosensitivity and survival rate of UPSb are similar to osteosarcoma but distinction from dedifferentiated

chondrosarcoma, which has a dismal prognosis, is important [5]. To date, there are no molecular studies or stringent diagnostic criteria to distinguish between these bone sarcomas. The genetics of UPSb is poorly understood. Previous studies have reported low frequency of *TP53* mutations, *MDM2* amplification [6], and various genomic gains and losses, including *CDKN2A, RB1* and *TP53* [2]. Nevertheless, no extensive high-throughput studies have been conducted to achieve a comprehensive understanding of the aetiology of these tumours. To gain further comprehensive insights into the molecular landscape and pathogenic mechanisms of UPSb, we performed integrative analysis using whole exome sequencing (WES) and RNA sequencing (RNA-Seq).

Materials and methods

Tumour samples

A retrospective search of the pathology database at the Royal Orthopaedic Hospital for resected samples of UPSb was carried out. In total, fourteen cases with the diagnosis of UPSb were identified (additional details in materials and methods supplementary material).

Whole exome sequencing and copy number alteration analysis

DNA from fresh frozen and FFPE (10 x 10µm sections) tissues was extracted and purified using the DNA Isolation (Roche Diagnostic Ltd, UK) and Arcturus PicoPure DNA Extraction kits (ThermoFisher Scientific, UK), respectively, following the manufacturer's instructions. A total of 1-3 µg of DNA from 12 tumours and 9 corresponding normals were sent to Oxford Gene Technology (OGT, Oxford) for WES. Exons were captured using the Agilent SureSelect Human All Exon V5 kit (Agilent, Santa Clara, USA), following manufacturer's protocol, and were massively sequenced (100-bp paired-end) using the Illumina HiSeq 2000 platform (Illumina, San Diego, CA). Two tumours and one corresponding normal were exome-sequenced at SMCL Next Generation Sequencing

(NGS) Hub (Cambridge, UK). Exons were captured by Illumina Nextera Rapid Capture Exome kit and sequenced using Illumina HiSeq 4000, generating 2 X 150 bp reads. Additional information is in materials and methods supplementary material.

CNV analysis (from WES data) was performed on ten normal-paired UPSb tumours, using CNVkit, applying the tool's default settings https://cnvkit.readthedocs.io/en/stable/).

RNA Sequencing experiment

Total RNA from eight fresh frozen tissues was extracted using either standard TRIzol-chloroform method or Qiagen RNeasy Mini kit (Qiagen, Manchester, UK) following the manufacturer's instructions. The quality and concentration of RNA was assessed using Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). A total of 50-100ng of DNase-treated RNA from each tumour was massively sequenced at the Genomics Birmingham facility (Institute of Cancer & Genomic Sciences, University of Birmingham). Additional information is in materials and methods supplementary material.

RNA-Seq differential gene expression profiling and hierarchical clustering analyses

The 149 RNAseq samples from the CINSARC dataset [1] were retrieved from the Sequence Read Archive Bioproject PRJNA282597 (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP057793) and 16 osteosarcoma samples were randomly chosen from the SRA bioproject PRJNA345424 (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP090849) [2]. Gene expression values were extracted by Kallisto v0.42.5 [3] with GRCh38 release 79 genome annotation and transformed into log2(tpm+2) prior to sample aggregation and normalization using the quantile method of the limma R package, within the R version 3.1.2 [4]. Clustering was computed using the Cluster v2.0.3 package. Pairwise comparisons of expression profiles

were performed using F-Test variance comparison prior to Student's t-Test statistics with a Bonferroni correction. The pairwise comparisons of UPSb versus all other tumours were performed iteratively with 20 random sampling of 10 samples from the larger group. Gene ontology analyses were performed with DAVID v6.7 [5].

Results

Exome sequencing identifies recurrent mutations in *TP53* and histone remodelling genes

To investigate the mutational landscape of UPSb, WES was performed on 14 tumours and ten matched-normal tissue samples, aiming to achieve a coverage depth of at least 50X in each sample. Using stringent criteria to call high-confidence somatic mutations (see materials and methods), 794 single nucleotide variants (SNVs), representing the majority of mutations (85%), and 138 indels (insertions and deletions) were detected, corresponding to a median of 2.2 mutations (range from 0 to 10) per coding megabase (Supplementary material, Table S1 & Figure S1). This overall somatic mutation burden is comparable with a low mutation burden average of 1.06 per megabase identified in 206 soft tissue sarcomas, including 44 UPS of soft tissue (UPSst) [12]. Somatic SNVs comprised of: 748 nonsynonymous substitutions, 33 nonsense and 13 splice sites mutations. A total of 123 SNVs and indels were validated by Sanger sequencing, achieving 96.6% validation rate. To identify potential cancer driver genes, we focused on somatic recurrent genes that are mutated in more than one tumour. A total of 31 recurrent genes harbouring heterozygous somatic mutations were identified in 2/14 (14%) tumours, except for *TP53* which was mutated in 4/14 (29%) samples (Figure 1A & Figure 2A). The majority of recurrent genes were confirmed by Sanger sequencing (examples in Supplementary material, Figure S2). All four *TP53* somatic missense substitutions

(R158H, V216M, Y236C C242G) are described 'somatic' in COSMIC database and predicted deleterious by two independent *in-silico* tools (Supplementary material, Table S2). V216M missense substitution has been reported previously in a UPSb tumour with a progressive disease behaviour [6]. In addition to mutations in *TP53* gene, which participates in the integrity of histone remodelling complexes [13], recurrent somatic mutations in histone remodelling genes including *H3F3A, ATRX* and *DOT1L* were detected in 5/14 (36%) samples, suggesting a potential role of defective chromatin remodelling genes in UPSb tumourigenesis (Figure 2 B, C & D, Supplementary material, Table S2). Somatic G34V missense substitution and V35_K36insL in-frame insertion in the previously known *H3F3A* hotspot residues were identified in two UPSb tumours. Mutations in *ATRX* and *DOT1L* were missense substitutions. One *ATRX* mutation, E351V, is reported somatic in COSMIC database. We identified two mutations in *DOT1L*, G307V and Q595H occurring in Histone-lysine N-methyltransferase catalytic DOT1 domain and STAT1 binding motifs, respectively. To further investigate the cancer-related genes, all recurrent candidate genes were assessed against COSMIC Cancer Gene Census (CCGC) and InTOgen cancer driver genes databases. Seven genes (*TP53*, *ATRX*, *H3F3A*, *ZFHX3*, *CSMD3*, *PRPRT, TRIO*) were classified as cancer drivers (Figure 1A). Eight recurrent genes (*TP53*, *ATRX*, *DOT1L*, *GCGR*, *COL4A2*, *KCNQ3*, *PKLR*, *SLC12A1*) were identified as potential druggable genes by the drug–gene interaction database (DGIdb) (Figure 1A).

Somatic copy number alteration analysis using WES data

Somatic copy number alteration analysis (CNV) was performed on ten normal-tumour UPSb samples. Supplementary material, Figure S3 shows the overall somatic CNV heatmap of UPSb tumours. In comparison with a study Niini et al. (Niini, 2011 #265) that assessed the CNV profile in UPSb, we comparingly identified CNV alterations in the following genes: *ING1*, *CGK4*, *MDM2*, *MYC*, *PDGFRA*, *KIT*, *KDR*, *PDGFA*, *PDGFB*, *VEGFA* (Supplementary material, Table S3). We also found somatic CNV losses in *RB1* in

5/10 (50%) cases, gains and losses in *VGLL3* and *CDKN2A*, losses in *YAP1* as well as loss and gain in *CCNE1* (Supplementary material, Table S3), genes that been previously implicated in UPSst (Cancer Genome Atlas Research Network. Electronic address, 2017 #271).

RNA-Seq identifies two previously reported and six novel gene fusions

RNA-Seq was conducted on eight tumours, achieving a mean of 59,722,616 reads with 90.07% successful alignment rate to reference genome. Using two bioinformatic fusion callers and following a series of filtering steps (see materials and methods), eight gene fusion candidates were identified in four tumours, two of which were previously described in the literature (*CLTC-VMP1* and *FARP1-STK24*) [7,15] whereas the remaining fusions are novel (Figure 1B & Supplementary material, Table S4). In addition, the previously reported *CTSC-RAB38* read-through chimera in cancer and non-cancer tissues [16] was identified in 2/8 tumours (25%). Using RT-PCR, all eight fusions were detected in the tumour cDNA but not in corresponding normal tissue samples, suggesting that the fusions are highly tumour-specific. Focusing on gene fusions with potential druggability and/or involvement in tumourigenesis, genomic analysis breakpoints using LR-PCR were performed on five out of the eight fusions to determine the precise genomic breakpoints (Supplementary material, Table S4). Using LR-PCR, all five fusions were genomically characterised and validated in the tumour DNA but were not detected in corresponding normal tissues, confirming the somatic status of the fusions. The *APOL1-MYH9* and *PKNOX2-MMP20* occur as a result of paracentric chromosome inversion whereas *ASAP2-ADAM17* forms by an interstitial chromosomal deletion of ~97 kilobase. The two gene fusions previously reported in other cancer samples and cell lines are *CLTC-VMP1* and *FARP1-STK24* (Supplementary material, Figure S4 & S5). *CLTC-VMP1* has been

described previously in BT-549 and HCC1954 breast cancer cell lines [7], hypopharynx tumour [17] and large-cell lung carcinoma [18]. The *CLTC-VMP1* gene fusion results from joining the first 14 exons of *CLTC* to the last two exons of *VMP1* (Supplementary material, Figure S4 A & B). Genomic breakpoint analysis revealed a ~158 kilobase interstitial deletion within *CLTC* and *VMP1* genes, leading to a complete deletion of *PTRH2* gene residing within the fusion gene partners (Supplementary material, Figure S4 C & D). The *CLTC-VMP1* chimeric transcript is an out-of-frame fusion, for which the predicted translation product includes the first 764 amino acids of *CLTC* gene that extends to the beginning of exon 11 of *VMP1* (at codon 324) and reaching a premature stop codon after 75 amino acids (Figure 3A). The *FARP1-STK24* chimeric fusion is formed as a result of a ~219 kilobase interstitial deletion within gene partners, linking the first three and five exons of *FARP1* and *STK24,* respectively (Supplementary material, Figure S5 A, B, C & D). The chimeric transcript is in-frame, consisting of the first 88 and 211 amino acid residues of *FARP1* and *STK24*, respectively (Figure 3B). In *FARP1-STK24*, the majority of the STK24 protein kinase domain is retained yet missing the STK24 regulatory region (Figure 3B). This gene fusion has been previously described in an invasive breast cancer tumour [15]. To infer the potential biological relevance of the novel chimeric transcripts, gene fusion partner genes were checked against CCCG and InTOgen database, classifying *CLTC* and *MYH9* (in *APOL1-MYH9)* as 'cancer driver genes' (Figure 1B). DGIdb identified three drug-gene interactions for three fusions partner genes (underlined) in the following gene fusions: *FARP1-STK24*, *ASAP2-ADAM17*, *PKNOX2-MMP20* (Figure 1B).

Hierarchical clustering and expression profiling analyses using RNA-Seq data

Using RNA-Seq data, unsupervised hierarchical clustering analysis of UPSb tumours, other sarcomas and SRP090849 datasets distinctly clustered the UPSb tumours together in two groups, UPSb-G1 and UPSb-G2, thus molecularly distinguishing UPSb from osteosarcoma and other sarcomas (Figure 4A). Checking the clinicopathological

information of the two UPSb groups, UPSb-G2 (T1, T2, T5, T6) primarily has spindle cell morphology, in comparison with UPSb-G1 (T9, T10, T13, T14), which shows a mixture of spindle, pleomorphic and epithelioid morphology (an example of each group in Supplementary material, Figure S6). Supervised expression analysis of UPSb versus other sarcoma subtypes highlighted *FGF23,* fibroblast growth factor 23*,* as a specifically expressed gene in UPSb tumours (Figure 4B). David Gene Ontology analyses of specific UPSb genes identified a strong enrichment of immune response genes, suggesting their potential involvement in UPSb tumourigenesis (Figure 4C).

Confirmation of elevated FGF23 expression in UPSb using quantitative RT-PCR

To confirm the elevated expression of FGF23 by RNA-Seq, quantitative RT-PCR was carried out on four UPSb tumours and four normal tissue controls. An elevated expression of FGF23 was significantly observed in all tumours, comparing to low FGF23 expression levels in normal samples (*p-value*=0.0286) (Supplementary material, Figure S7).

<u>**Discussion**</u>

In this study, we examined the somatic genetic alterations present in UPSb tumours using WES and RNA-Seq technologies as well as comparing transcriptomic profiles of UPSb with other sarcomas. We performed WES on 14 tumours and focused on genes that are altered in more than one tumour, identifying a total of 31 recurrent genes. *TP53* was the most frequently mutated gene (29% of tumours). In 36% of the tumours, we identified mutations in chromatin remodelling genes (*H3F3A, ATRX, DOT1L*) which have not been previously described in the UPSb subtype. A significant co-occurrence of G34 *H3F3A* mutation with *ATRX/DAXX* and *TP53* mutations has been observed in nearly 100% of glioblastoma tumours [14]. Notably, no correlation of *H3F3A* mutations was observed in UPSb tumours harbouring *TP53*, *ATRX* or *DOT1L* mutations.

Heterozygous R132 and R172 hotspot point mutations in *IDH1* and *IDH2*, respectively, are commonly present in 61-87% of chondrosarcoma cases, including dedifferentiated chondrosarcoma [4,19]; whereas, these changes are absent in 222 osteosarcoma samples [20]. In an earlier study by Chen et al. [4], we investigated the *IDH1/2* mutation status in the 14 UPSb tumours used in this study and found no R132 and R172 mutations in any tumour, ruling out a diagnosis of dedifferentiated chondrosarcoma. Unlike the established association of *IDH1/2* in chondrosarcoma, the somatic genomic profile of osteosarcoma is complex and involves multiple genes, mainly *TP53* and *RB1* [21]. These genes are well-known cancer driver genes implicated in multiple cancers and therefore cannot be sensitively used to exclude the diagnosis of osteosarcoma.

The *TP53* gene is the most frequently mutated gene in various human cancers and 90% of *TP53* mutations are missense changes with potential gain-of-function characteristics [22,23]. Previously, *TP53* mutations were identified in 22% of UPSb tumours by conventional PCR and Sanger sequencing [6]. In this study, using massive-parallel sequencing for the first time on this tumour subtype, four *TP53* missense mutations (R158H, COSMIC ID: COSM1640853; V216M, COSM10667; Y236C, COSM10731; C242G, COSM3717645), all occurring in the DNA-binding domain of p53 protein (Figure 2A), were indentified. Mutations in the p53 DNA-binding domain can reduce the protein's binding specifically to DNA sequence motifs in p53-regulated genes. The R158H, V216M

and Y236C mutations are reported among the 50 most common somatic missense mutations in *TP53*, highlighting their potential pathological role in tumourigenesis [22]. We could not find significant differences in the total number of mutations in samples harbouring *TP53* mutations (n=4) comparing to the remaining samples (n=10). Correlations between *TP53* mutations and clinical implications are difficult to establish due to the clinical heterogeneity of the patients and small sample size; hence, additional investigations to elucidate their prognostic information are required.

Recurrent mutations in *H3F3A*, *ATRX* and *DOT1L* were detected in 5/14 tumours. Highly specific cancer-driving hotspot mutations in *H3F3A* (G34) and *H3F3B* (K36) were identified in 92% of giant cell tumour of the bone and 95% of chondroblastoma cases [24]. In this study, we identified recurrent somatic mutations (G34V, COSM502595 and V35_K36insL, COSM5574356) in *H3F3A* affecting the previously reported amino acid residues (Figure 2B). These hotspot sites are well-conserved amino acid residues of the amino-terminus tail that undergoes post-translational modifications [14]. Lysine 36 is a principal methylation site that typically promotes gene transcription when methylated or acetylated [14]. Histone 3.3 lysine to methionine substitution (K36M) reduces the methylation of lysine residue through inhibition of SET domain-containing enzymes [25]. This reduction of methylation at K36 was also observed in cell lines carrying the G34V substitution [25]. Since V35_K36insL is an-frame insertion and not a methionine substitution, further investigations will be necessary to elucidate any pathogenic mechanism of this in-frame insertion.

H3F3A/H3F3B driver mutations were described in giant cell tumour of the bone and chondroblastoma which are considered benign tumours or benign but locally aggressive tumours, respectively [24,26] as well as in malignant giant cell tumour of the bone [27,28]. A recent study by Amary et al. [27] identified *H3F3A* G34 substitutions in 13/385 (3.37%) of primary malignant bone tumours, classified as either osteosarcoma or malignant giant cell tumour of bone. In this study, we report recurrent *H3F3A* alterations in 2/14 (14.3%)

UPSb tumours, a higher percentage than in previously reported malignant tumours, with the caveat of a smaller cohort size. Hence, the possibility of a malignant phenotype or evolution should be considered in tumours harbouring *H3F3A* alterations.

ATRX is a member of the SWI/SNF2 (SWItch/Sucrose Non-Fermentable) ATP-dependent chromatin remodelling protein complex [29] which regulates the expression of thousands of genes through remodelling of chromatin structure [30,31]. *ATRX* has an established role in the carcinogenesis of multiple cancers including gliomas [29], small cell lung cancers [32] and six adult soft tissue sarcomas (including UPS, leiomyosarcoma, dedifferentiated liposarcoma) [12]. Studies have demonstrated a regulational role of ATRX, along with DAXX (death domain-associated protein) and other histone chaperone complex proteins, in the enrichment of histone H3.3 in telomeres and heterochromatin regions [33,34]. Dysfunction ATRX/DAXX is associated with the alternative lengthening of telomeres (ALT), a phenomenon observed in 10-15% of cancers of mesenchymal origin (e.g. UPSb) resulting in widespread genomic destabilisation [35,36]. In this study, we identified two missense mutations in *ATRX:* S2109I and E351V (COSMIC ID: COSM6608613) occurring in the Helicase/ adenosine triphosphatase (ATPase) conserved C-terminus and Enhancer zeste homologue 2 (EZH2) interacting region of the ATRX protein, respectively (Figure 2C). The helicase/ATPase subunit is the catalytic core of ATRX protein [37] which, along with other SNF2 proteins, is involved in ATP hydrolysis responsible for chromatin structural conformations [38]. The ATRX EZH2 interaction region is involved in the interaction of ATRX with (polycomb repressive complex 2) PRC2 protein complex, including EZH2 catalytic subunit [39,40]. Although further studies are required, mutations in these two functionally important domains disrupt ATRX protein function, affecting the integrity of the chromatin structure. PRC2/EZH2 inhibitors are currently in clinical trials [39,41]; however, the complex and dual oncogenic and tumour-suppressing role of PRC2 in cancers

requires detailed mechanistic insights before establishing the efficacy of these inhibitors [39].

In the second part of this study, we aimed to identify genome-wide gene fusions arising from chromosomal aberrations in eight UPSb tumour using RNA-Seq. A total of eight highly tumour-specific gene fusions were validated by RT-PCR, five of which were genomically characterised by LR-PCR. We identified two gene fusions previously reported in other cancers, *CLTC-VMP1* and *FARP1-STK24*.

The *CLTC-VMP1* gene fusion is identified in T13, lacking any WES alterations in *TP53* or, *H3F3A*, *ATRX, DOT1L* chromatin remodelling genes. The cDNA breakpoint of *CLTC* identified here is different to those previously reported; however, the breakpoint position in *VMP1* in UPSb is the same as reported for other tumours. The genomic interstitial deletion at 17q23.1 locus, revealed by genomic breakpoint analysis, leads to a complete deletion of *PTRH2* gene and a loss of six repeats of Clathrin heavy chain/VPS and four transmembrane helices of *CLTC* and *VMP1*, respectively (Figure 3A). The Clathrin protein is involved in chromosome segregation and Golgi reassembly during mitosis and protein-protein interactions [45]. Rearrangements involving *CLTC-PTRH2-VMP1* locus have been observed in multiple tumour types, including glioblastoma, lung cancer, breast cancer and leukaemias [7]. VMP1 encodes an autophagy-related protein that promotes apoptosis in pancreatic cancer cells [46]. PTRH2 is a mitochondrial protein that induces apoptosis by regulation of the function of Groucho family transcriptional regulators [47]. Knockdown of *PTRH2/BIT1* in adherent cells decreased cell survival and promoted staurosporine and serum-deprivation apoptosis of cells, consistent with tumour suppressive role [48]. Altogether and as Giacomini et al. [7] suggested, the rearrangement involving *CLTC-PTRH2-VMP1* and the *CLTC-VMP1* gene fusion being out-of-frame are indicative of a disruption of the tumour suppressor activity.

The *FARP1-STK24* gene fusion is formed by joining the two gene partners in opposing (sense-to-antisense) orientations, known as 5'-to-5' gene fusions Figure 3B). 5'-5' gene fusions have previously been reported in breast cancers [49]. Since the 5' transcription regulatory apparatus of both *FARP1* and *STK24* is retained, theoretically, both sense (*FARP1*) and antisense (*STK24*) genes can start transcription that extends into the other gene partner, or vice versa. Gene fusions involving *FARP1* have been reported in multiple cancers, including lung adenocarcinoma, breast adenocarcinoma and lower grade glioma (http://www.tumorfusions.org). *STK24* is serine/threonine protein kinase belonging to the mammalian Sterile20-like (MST) kinase family, key signalling molecules that regulate cell division cycle, cell morphogenesis, apoptosis and oncogenic transformation [50,51]. A caspase-dependent apoptotic role has been identified for STK24 protein [52]. The STK24 protein is cleaved and activated by caspase-3 protein, through the STK24 regulatory domain, and translocated into the nucleus to promote apoptotic responses [50]. The loss of STK24 regulatory domain can subsequently interfere with STK24 activation and nuclear localisation. The *FARP1-STK24* forms an in-frame fusion protein with a potential constitutive activation of a kinase, a phenomenon observed in gene fusions exerting oncogenic functionality [53]. Altogether, *FARP1-STK24* may be associated with oncogenic properties but further investigations are required.

Gene fusions involving protein kinases are considered potential therapeutic targets, the use of kinase inhibitors in tumours harbouring kinase-related gene fusions can improve tumour prognosis and patient outcome [54]. For example, the efficacy of using ALK inhibitors in advanced non–small-cell lung cancer with *ALK* rearrangement is evident in clinical trials [55]. A study by Olsen et al. [56] identified 14 inhibitors, eight of which are in clinical trials or are FDA approved, that inhibited the enzymatic activity of STK24. Although further detailed investigations are required, STK24-selective inhibitors are potential cancer therapeutics in tumours harbouring *STK24* rearrangements.

Unsupervised clustering analysis of the RNA-Seq data clearly distinguished the UPSb samples from classical UPS as well as synovial sarcomas and osteosarcomas. Supervised expression profiling of UPSb versus other tumour subtypes revealed elevated expression of FGF23, which was confirmed in four UPSb tumours using quantitative RT-PCR. A study by Shiba et al. (Shiba, 2016 #341) showed a highly specific immunohistochemical expression of FGF23 in phosphaturic mesenchymal tumours, whereas FGF23 expression was negative in 46 tumours, including osteosarcoma, chondrosarcoma and synovial sarcoma. Elevated expression of FGF23 can serve as a molecular marker specific to UPSb which can be diagnostically utilised in clinics. However, confirmation of elevated FGF23 expression in a larger UPSb cohort is recommended. FGF family, comprised of signalling proteins, has a role in tissue repair and tumourigenesis, by regulating cell proliferation, migration and angiogenesis (Feng, 2015 #340).   Promising curative results of a FGF23 monoclonal antibody drug, KRN23, was observed in patients with tumour-induced osteomalacia, a rare paraneoplastic syndrome clinically described by bone pain fractures and muscle weakness (Florenzano, 2017 #401). Further investigations are necessary to confirm FGF23 targeted therapeutic opportunity in UPSb.

An enrichment of immune response genes was identified in UPSb, which has also been documented in other sarcomas, including UPSst (Cancer Genome Atlas Research Network. Electronic address, 2017 #271). Therapeutic benefit of pembrolizumab, an immune checkpoint inhibitor, has been documented in 40% of UPSst cases (Tawbi, 2017 #342). Although further investigations are needed, immune checkpoint inhibitors can be considered as a potential therapeutic option in UPSb patients.

In summary, this study provides a first detailed genetic and transcriptomic alterations landscape of UPSb tumours, thus providing useful insights into tumourigenesis and broadening the understanding of this tumour subtype. We identified novel recurrent gene mutations in multiple cancer-related genes, including chromatin remodelling genes, that

are reported in UPSb tumours for the first time. We also identified novel and previously reported gene fusions. Several of the recurrent mutated genes and gene fusions represent potential druggable targets that can translate into clinics and improve patient prognosis. Elevated expression of FGF23 was identified, which may be a potential molecular biomarker for UPSb.

Authors' Contributions

NMA, ATB, VS and FL designed the research and analysed the data; VS provided tumour samples; reviewed slides and the clinical information; NMA, SN, MRM, SK, AA, DH, CRA, FT performed experiments and/or analysed data; NMA, VS and FL wrote the paper. All authors approved the final version.

References

1. Fletcher CDM. WHO classification of tumours of soft tissue and bone. In. (ed)^(eds). International Agency for Research on Cancer, 2013.

2. Niini T, Lahti L, Michelacci F, *et al.* Array comparative genomic hybridization reveals frequent alterations of G1/S checkpoint genes in undifferentiated pleomorphic sarcoma of bone. *Genes Chromosomes Cancer* 2011; **50**: 291-306.

3. Doyle LA. Sarcoma classification: an update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. *Cancer* 2014; **120**: 1763-1774.

4. Chen S, Fritchie K, Wei S, *et al.* Diagnostic utility of IDH1/2 mutations to distinguish dedifferentiated chondrosarcoma from undifferentiated pleomorphic sarcoma of bone. *Hum Pathol* 2017; **65**: 239-246.

5. Gerrand C, Athanasou N, Brennan B, *et al.* UK guidelines for the management of bone sarcomas. *Clin Sarcoma Res* 2016; **6**: 7.

6. Kawaguchi K, Oda Y, Sakamoto A, *et al.* Molecular analysis of p53, MDM2, and H-ras genes in osteosarcoma and malignant fibrous histiocytoma of bone in patients older than 40 years. *Mod Pathol* 2002; **15**: 878-888.

7. Giacomini CP, Sun S, Varma S, *et al.* Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. *PLoS Genet* 2013; **9**: e1003464.

8. Koboldt DC, Zhang Q, Larson DE, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22**: 568-576.

9. Cibulskis K, Lawrence MS, Carter SL, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213-219.

10. Kim D, Pertea G, Trapnell C, *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013; **14**: R36.

11. Haas B, Dobin A, Stransky N*, et al.* STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv* 2017.

12. Cancer Genome Atlas Research Network. Electronic address edsc, Cancer Genome Atlas Research N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* 2017; **171**: 950-965 e928.

13. Nair SS, Kumar R. Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Molecular oncology* 2012; **6**: 611-619.

14. Schwartzentruber J, Korshunov A, Liu XY*, et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 2012; **482**: 226-231.

15. Veeraraghavan J, Tan Y, Cao XX*, et al.* Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun* 2014; **5**: 4577.

16. Babiceanu M, Qin F, Xie Z*, et al.* Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res* 2016; **44**: 2859-2872.

17. Nair J, Jain P, Chandola U*, et al.* Gene and miRNA expression changes in squamous cell carcinoma of larynx and hypopharynx. *Genes & cancer* 2015; **6**: 328-340.

18. Liu J, Lee W, Jiang Z*, et al.* Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res* 2012; **22**: 2315-2327.

19. Kerr DA, Lopez HU, Deshpande V*, et al.* Molecular distinction of chondrosarcoma from chondroblastic osteosarcoma through IDH1/2 mutations. *The American journal of surgical pathology* 2013; **37**: 787-795.

20. Amary MF, Bacsi K, Maggiani 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* 2011; **224**: 334-343.

21. Gianferante DM, Mirabello L, Savage SA. Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy. *Nat Rev Endocrinol* 2017; **13**: 480-491.

22. Baugh EH, Ke H, Levine AJ*, et al.* Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death Differ* 2018; **25**: 154-160.

23. Kandoth C, McLellan MD, Vandin F*, et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502**: 333-339.

24. Behjati S, Tarpey PS, Presneau N*, et al.* Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet* 2013; **45**: 1479-1482.

25. Lewis PW, Muller MM, Koletsky MS*, et al.* Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* 2013; **340**: 857-861.

26. Cleven AH, Hocker S, Briaire-de Bruijn I*, et al.* Mutation Analysis of H3F3A and H3F3B as a Diagnostic Tool for Giant Cell Tumor of Bone and Chondroblastoma. *The American journal of surgical pathology* 2015; **39**: 1576-1583.

27. Amary F, Berisha F, Ye H*, et al.* H3F3A (Histone 3.3) G34W Immunohistochemistry: A Reliable Marker Defining Benign and Malignant Giant Cell Tumor of Bone. *The American journal of surgical pathology* 2017; **41**: 1059-1068.

28. Righi A, Mancini I, Gambarotti M*, et al.* Histone 3.3 mutations in giant cell tumor and giant cell-rich sarcomas of bone. *Hum Pathol* 2017; **68**: 128-135.

29. Nandakumar P, Mansouri A, Das S. The Role of ATRX in Glioma Biology. *Front Oncol* 2017; **7**: 236.

30. Prasad P, Lennartsson A, Ekwall K. The roles of SNF2/SWI2 nucleosome remodeling enzymes in blood cell differentiation and leukemia. *Biomed Res Int* 2015; **2015**: 347571.

31. Tang L, Nogales E, Ciferri C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Progress in biophysics and molecular biology* 2010; **102**: 122-128.

32. Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science advances* 2015; **1**: e1500447.

33. Drane P, Ouararhni K, Depaux A*, et al.* The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* 2010; **24**: 1253-1265.

34. Wong LH, McGhie JD, Sim M*, et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* 2010; **20**: 351-360.

35. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* 2010; **11**: 319-330.

36. Heaphy CM, de Wilde RF, Jiao Y*, et al.* Altered telomeres in tumors with ATRX and DAXX mutations. *Science* 2011; **333**: 425.

37. Gibbons RJ, Wada T, Fisher CA*, et al.* Mutations in the chromatin-associated protein ATRX. *Hum Mutat* 2008; **29**: 796-802.

38. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 2013; **154**: 490-503.

39. Comet I, Riising EM, Leblanc B*, et al.* Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer* 2016; **16**: 803-810.

40. Ratnakumar K, Bernstein E. ATRX: the case of a peculiar chromatin remodeler. *Epigenetics* 2013; **8**: 3-9.

41. McCabe MT, Ott HM, Ganji G*, et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 2012; **492**: 108-112.

42. Wong M, Tee AEL, Milazzo G, *et al.* The Histone Methyltransferase DOT1L Promotes Neuroblastoma by Regulating Gene Transcription. *Cancer Res* 2017; **77**: 2522-2533.

43. Wong M, Polly P, Liu T. The histone methyltransferase DOT1L: regulatory functions and a cancer therapy target. *American journal of cancer research* 2015; **5**: 2823-2837.

44. Zhu B, Chen S, Wang H, *et al.* The protective role of DOT1L in UV-induced melanomagenesis. *Nat Commun* 2018; **9**: 259.

45. Young A. Structural insights into the clathrin coat. *Semin Cell Dev Biol* 2007; **18**: 448-458.

46. Pardo R, Lo Re A, Archange C, *et al.* Gemcitabine induces the VMP1-mediated autophagy pathway to promote apoptotic death in human pancreatic cancer cells. *Pancreatology* 2010; **10**: 19-26.

47. Jan Y, Matter M, Pai JT, *et al.* A mitochondrial protein, Bit1, mediates apoptosis regulated by integrins and Groucho/TLE corepressors. *Cell* 2004; **116**: 751-762.

48. Griffiths GS, Grundl M, Leychenko A, *et al.* Bit-1 mediates integrin-dependent cell survival through activation of the NFkappaB pathway. *The Journal of biological chemistry* 2011; **286**: 14713-14723.

49. Shlien A, Raine K, Fuligni F, *et al.* Direct Transcriptional Consequences of Somatic Mutation in Breast Cancer. *Cell Rep* 2016; **16**: 2032-2046.

50. Cho CY, Lee KT, Chen WC, *et al.* MST3 promotes proliferation and tumorigenicity through the VAV2/Rac1 signal axis in breast cancer. *Oncotarget* 2016; **7**: 14586-14604.

51. Thompson BJ, Sahai E. MST kinases in development and disease. *J Cell Biol* 2015; **210**: 871-882.

52. Fuller SJ, McGuffin LJ, Marshall AK, *et al.* A novel non-canonical mechanism of regulation of MST3 (mammalian Sterile20-related kinase 3). *Biochem J* 2012; **442**: 595-610.

53. Stransky N, Cerami E, Schalm S, *et al.* The landscape of kinase fusions in cancer. *Nat Commun* 2014; **5**: 4846.

54. Tamura R, Yoshihara K, Yamawaki K, *et al.* Novel kinase fusion transcripts found in endometrial cancer. *Sci Rep* 2015; **5**: 18657.

55. Shaw AT, Kim DW, Nakagawa K, *et al.* Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *The New England journal of medicine* 2013; **368**: 2385-2394.

56. Olesen SH, Zhu JY, Martin MP, *et al.* Discovery of Diverse Small-Molecule Inhibitors of Mammalian Sterile20-like Kinase 3 (MST3). *ChemMedChem* 2016; **11**: 1137-1144.

Table and Figures Legends

Figure 1. The mutational and gene fusion landscape of UPSb. (A) Recurrent genes identified by whole exome sequencing in 14 tumours, grouped according to their biological pathway and function. **(B)** Gene fusions identified by RNA sequencing in eight tumours. **CCGC**: COSMIC Cancer Census Gene; **DGIdb**: The Drug Gene Interaction database. Gene described 'cancer driving gene' in CCGC or druggable in DGIdb are shaded in grey. †: *TRIO* is classified 'cancer driver gene' by IntOGen.

Figure 2. Recurrent somatic mutations in *TP53*, *H3F3A*, *ATRX* and *DOT1L*. All the mutations (black balls) are missenses except for one an in-frame insertion in H3F3A. The relative positions of mutations are shown in the predicted protein sequence of (A) TP53, (B) H3F3A, (C) ATRX and (D) DOT1L. T.A.: p53 transactivation domain; T.M.: p53 tertramerization domain; ADD: ATRX-DNMT3-DNMT3L domain; EZH2: enhancer zeste

homologue 2 protein (EZH2) interacting region; DAXX: death domain-associated protein interacting domain; ATPas: Helicase adenosine triphosphatase domain; Catalytic DOT1 domain: Histone-lysine N-methyltransferase DOT1 domain; STAT1: signal transducer and activator of transcription 1 binding motifs.

Figure 3. The impact of the *CLTC-VMP1* and *FARP1-STK24* gene fusions on protein domain organization. The grey shaded area represents the retained protein domains of the fused exons of **(A)** *CLTC-VMP1* and **(B)** *FARP1-STK24*. The gene fusion breakpoints are denoted by a black double slash. The *CLTC-VMP1* is out of frame, reaching a premature stop (denoted by *) at codon 399 of *VMP1* gene. The *FARP1-STK24* is in-frame, resulting from joining the first 88 and 211 amino acids of *FARP1* and *STK24*, respectively. ATP B.S: ATP binding site.

Figure 4. Unsupervised and supervised analyses of UPSb, other sarcomas and SRP090849 datasets using RNA-Seq data. (A) Unsupervised clustering analysis using Ward's distance and 1-pearson correlation highlighted two groups of UPSb tumors (dark grey). UPSb_G1 samples are clearly separated from classical UPS (light grey) and synovial sarcomas from the SRP057793 dataset and from osteosarcomas from the SRP090849 dataset. UPSb_G2 samples are closer to classical UPS but remain clearly distinct. **(B)** Supervised analysis showing a violin plot of the *FGF23* gene expression across UPSb and the different tumor types present in the SRP057793 dataset and in the osteosarcoma SRP090849 dataset. Individual samples are shown in circle and group median is represented as a black bar. **(C)** David Gene Ontology analyses of specific UPSb genes indicating a strong enrichment of immune response genes. Fold enrichment (bar chart) and the $-\log_{10}$ of the hypergeometric test *P*-value corrected by Bonferroni (black line with closed circles) are shown.

List of supplementary material online Information

Supplementary table and figure legends.

Table S1. Coding mutational rate of UPSb.

Table S2. Summary of the individual recurrent mutations identified in *TP53* and chromatin remodelling genes

Table S3. Somatic copy number alterations status of genes for which copy number alterations have been previously reported in UPS of bone and UPS soft tissue.

Table S4. Summary of the eight somatic gene fusions identified by RNA-Seq.

Figure S1. The number of somatic alterations identified in each UPSb tumour.

Figure S2. Examples of Sanger sequencing confirmation of eight recurrent mutations identified by WES.

Figure S3. Somatic copy number alteration heatmap of ten normal-tumour paired UPSb samples.

Figure S4. Diagrammatic representation and validation of *CLTC-VMP1* gene fusion.

Figure S5. Schematic representation and validation of *FARP1-STK24* fusion.

Figure S6. Microscopy images of T1 and T10.

Figure S7. Expression of *FGF23* in UPSb tumours versus normal tissue samples using quantiative RT-PCR.

Supplementary methods

Tumour sample additional information

All cases were thoroughly examined by a specialist bone tumour pathologist (VPS) to exclude osteosarcoma and dedifferentiated chondrosarcoma. Fresh frozen tumour sample was obtained from thirteen out of the fourteen cases and formalin-fixed paraffin-embedded

(FFPE) tissue was used in one case (T12). All tumours had matched-normal samples obtained from tumour-free muscle tissues adjacent to the tumour site. All samples were obtained from the Royal Orthopaedic Hospital NHS Foundation Trust Tumour Bank with informed consent from the patient and ethical approval from institutional and local research committee boards. Prior to the study, all patient samples were anonymised and used in alliance with the ethical rules and regulations presented in the Declaration of Helsinki. The age of the patients ranged between 24-88 years (median 64) at the time of diagnosis. Five patients were male and nine were female. The size of the tumours ranged from 3 to 12.5 cms and anatomical sites included femur, tibia, humerus and fibula. The morphology of tumours consisted of spindle (n=10); pleomorphic (n=1); spindle and epithelioid (n=2); spindle and pleomorphic (n=1) cells.

WES analysis pipeline: reads mapping and variant calling

Sequenced reads from all samples were aligned to the human genome reference sequence GRCh37 using BWA 0.7.x bioinformatic tool (Li, 2009 #337). Optical and PCR duplicates were marked with Picard 1.x (http://picard.sourceforge.net). Somatic SNVs and indels were called using VarScan2 [8] and MuTect [9] tools, applying the default settings. Variant allele frequency (VAF) of ≥ 10% was selected for all tumours, except for one FFPE tumour where the VAF was adjusted to ≥20% to reduce false positive artefacts. A step-wise filtering scheme was followed to identify likely somatic and pathogenic variants. In short, only nonsynonymous SNVs (missense, splice site, nonsense) and indels with ≥3 or ≥5 supporting sequencing reads and ≥20% and ≥10% VAF, respectively, were selected. Variants were annotated with a modified version of Ensembl Variant Effect Predictor (McLaren, 2010 #339). Variants were manually visualised using Integrative Genomics Viewer (IGV) and artefactual calls were excluded. To discard potential germline polymorphisms, variants reported in dbSNP build 137 or the Exome Aggregation

Consortium (ExAc) datasets with a minor allele frequency (MAF) of >0.1% were eliminated. Additional rigorous steps were applied to four tumours for which matched-normal DNA samples were not WES but were available for Sanger sequencing validation. In these four tumours, reported variants in dbSNP or ExAc (regardless of MAF) were discarded except for variants with <0.01% MAF and described somatic in COSMIC. The COSMIC database (http://cancer.sanger.ac.uk/cosmic) was used to highlight previously reported cancer mutations. SIFT and PolyPhen-2 tools were used to predict the pathogenic impact of missense substitutions and only variants that were predicted "deleterious" by at least one tool were retained. Recurrent genes were investigated in COSMIC Cancer Gene Census (CCGC) consortium and InTOgen database (https://www.intogen.org/search) to highlight potential cancer driver genes. The Drug Gene Interaction Database (DGIdb) was used to identify potential druggable genes (http://www.dgidb.org).

WES variant validation by standard PCR and Sanger sequencing

DNA of the tumour and adjacent normal tissue from the same patient was amplified by PCR (primer sequences are available upon request) targeting the genomic regions flanking the variant. PCR products were purified using microCLEAN (Microzone, Stourbridge, UK). Sequencing reactions were performed using the BigDye v3.1 cycle sequencing kit and sequenced on an ABI 3730 DNA analyser (Applied Biosystems, California, USA), following manufacture's guidelines.

RNA Sequencing analysis pipeline: candidate gene fusion identification

Briefly, cDNA libraries were constructed using the Neoprep stranded mRNA library prep (Illumina, NP-202-1001), according to manufacturer's protocol. Prepared libraries were

sequenced on NextSeq 500/550 High Output (Illumina) to produce paired-end reads (75bp in length). Mapping and alignment of sequenced reads to the human genome reference sequence GRCh37 was achieved using the Tuxedo Suite (Trapnell, 2012 #338). Candidate gene fusions were called using TopHat2 [10] and STAR-Fusion fusion-junction mappers [11]. A series of filtering and prioritisation steps were followed to identify genuine fusion calls. A 'putative' gene fusion required a minimum of three reads supporting fusion junction (flanking right and left sides of the fusion). Gene fusions where both fusion partners are intronic or in intergenic regions were excluded as they can be DNA contaminants or unspliced mRNA precursors. Gene fusions were manually inspected and candidate gene fusions with uniform distribution of fusion junction supporting reads were prioritised.

Gene fusion validation by RT-PCR, LR-PCR and Sanger Sequencing

The total RNA was used to generate cDNA using SensiFAST cDNA Synthesis Kit (Bioline, UK), following the manufacturer's instructions. The reverse transcription PCR (RT-PCR) reactions were assembled as following: PCR reaction buffer (50 mM Tris/HCl, 10 mM KCl, 5 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$), 2.5mM of dNTPs, 0.5U of Fast Start DNA polymerase (Roche, Burgess Hill, UK) and 20 pmole of each forward and reverse primers flanking fusion breakpoints for 35 cycles. Genomic breakpoints analysis, through genome walking, was carried out by long range PCR (LR-PCR) using PrimeSTAR GXL DNA polymerase kit (Takara, USA). LR-PCR reaction was assembled as following: 10µl of 5X PrimeSTAR GXL buffer, 200 µM of dNTPs, 0.2 µM of each forward and reverse primers, 1.25 U PrimeSTAR GXL DNA polymerase enzyme and 50ng DNA template. LR-PCR products were amplified using a two-step PCR programme: 10 seconds at 98°C and 10 minutes at 60°C for 30 cycles. All RT-PCR and LR-PCR primers are available upon request. Somatic status assessment of gene fusions by RT-PCR and LR-PCR was conducted on both tumour and match-normal nucleic acids. PCR products were visualized using 1-2% (w/v) agarose gel

(Bioline, UK). Purified PCR products were cleaned and Sanger sequenced as mentioned above.

Quantitative RT-PCR analysis of FGF23

qRT-PCR analysis was performed using Taqman gene expression assays for FGF23 (Hs00221003_m1) and GAPDH (Hs02786624_g1) (Thermofisher Scientific, UK) according to manufacturer's instructions. cDNA was generated using a High Capacity cDNA Reverse Transcription Kit (Multiscribe Reverse Transcriptase, Applied Biosystems, California, US) following the manufacturer's protocol. The mRNA level of FGF23 in the sample was determined following a real-time PCR reaction on a QuantoStudio 6 Flex Real-Time PCR system (Applied Biosystems, California, US) for a 20 µL reaction using TaqMan gene expression assays (20X primers and probe) and Gene Expression Master Mix (2X TaqMan Universal mix) (all by Applied Biosystems). Gene expression was normalized for control gene expression (GAPDH) and calculated according to the ΔΔCT method.

A

Recurrent gene	1	2	3	4	5	6	7	8	9	10	11	12	13	14	CCCG	DGId
TP53			M*		M*	M*			M*						▓	▓
ATRX						M								M*	▓	▓
H3F3A	I*									M*					▓	
DOT1L			M			M										▓
COL4A2		M										M*				▓
PKLR			M			M										▓
PEG3			M*			M*										
MCAM						M						M				
PCDH15			M									M				
SYNE2			M			M*										
PTPRT	M		M												▓	
TRIO†			M										M			
GCGR		N										M				▓
ZFHX3										I			I		▓	
ARAP2			M									M				
LOXHD1			M*									M				
KCNQ3						M				M						▓
KRTAP5-5													I*	I*		
PTF1A			F											M		
KIAA1109			F*											M		
PHF3			F*			M										
TVP23A			F*									F*				
SLC12A1			S			M										▓
MYO7B			M								M*					
DNAH3			M									M*				
TMEM150B			M							M						
MAMDC4	M		M													
ZBTB34						M						M				
SACS																
TTBK1			M									M				
CSMD3			M									M			▓	

Row groups (left margin labels):
- Chromatin remodelling genes: *ATRX*, *H3F3A*, *DOT1L*
- Cell cycle; Cell death; DNA replication & recombination: *COL4A2*, *PKLR*, *PEG3*, *MCAM*, *PCDH15*, *SYNE2*
- Cellular growth & proliferation; Cellular signalling: *PTPRT*, *TRIO†*
- Signal transduction: *GCGR*, *ZFHX3*, *ARAP2*, *LOXHD1*
- Developmental biology: *KCNQ3*, *KRTAP5-5*, *PTF1A*

Legend:

M	Missense	F	Frameshift
N	Nonsense	I	In-frame INDEL
S	Splice site	*	Reported in COSMIC

B

Gene fusion	1	2	5	6	9	10	13	14	CCCG	DGIdb
CLTC-VMP1							▨		CLTC	
FARP1-STK24				▓						░
APOL1-MYH9							▓		MYH9	
ASAP2-ADAM17				▨						░
PKNOX2-MMP20		▨								░
OSBPL2-CABLES2					▨					
MICAL3-UFD1L					▨					

▨ **Out-of-frame fusion**
▓ **In-frame fusion**

CMAS-PYROXD1		▨								

Figure 1: The mutational and gene fusion landscape of UPSb. (A) Recurrent genes identified by whole exome sequencing in 14 tumours, grouped according to their biological pathway and function. **(B)** Gene fusions identified by RNA sequencing in eight tumours. **CCGC**: COSMIC Cancer Census Gene; **DGIdb**: The Drug Gene Interaction database. Gene described 'cancer driving gene' in CCGC or druggable in DGIdb are shaded in grey. † *TRIO* is classified as 'cancer driver gene' by IntOGen.