Comprehensive molecular characterization of adamantinoma and OFD-like adamantinoma bone tumours

Naser M Ali (MSc)[1], Stefania Niada (PhD)[2], Mark R Morris (PhD)[3], Anna T Brini (PhD)[2,4], David Huen (PhD)[3], Vaiyapuri Sumathi (MD)[5#], Farida Latif (PhD)[1#].

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

2. IRCCS Istituto Ortopedico Galeazzi.

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

4. Department of Biomedical, Surgical and Dental Sciences, Università degli Studi di Milano, Milan, Italy.

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

[#]Corresponding authors

Corresponding authors:

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

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

Abstract

Adamantinoma and osteofibrous dysplasia (OFD)-like adamantinoma are rare primary bone tumours that are predominantly confined to the tibia. These two entities show similarities in location, histology and radiological appearance; however, adamantinoma is malignant and therefore differentiating between these bone tumours is essential for optimal patient care. To elucidate their genomic and transcriptomic alteration profiles and expand their aetiological mechanisms, whole exome sequencing (WES) and RNA Sequencing (RNA-Seq) were conducted on adamantinoma and OFD-like adamantinoma tumours. CNV analysis using WES data revealed distinct chromosomal alteration profiles for adamantinoma tumours compared to OFD-like adamantinomas, allowing molecular differentiation between the two tumour subtypes. Combining WES and CNV analyses, the chromatin remodelling-related gene *KMT2D* was recurrently altered in 3/8 adamantinoma tumours (38%), highlighting the potential involvement of deregulated chromatin structure and integrity in adamantinoma tumourigenesis. RNA-Seq analysis revealed a novel somatic gene fusion (*EBHB4-MARCH10*) in an adamantinoma, the gene fusion was fully characterized. Hierarchical clustering analysis of RNA-Seq data distinctly clustered adamantinoma tumours from OFD-like adamantinomas, allowing to molecularly distinguish between the two entities. David Gene Ontology analysis of differentially expressed genes

identified distinct altered pathways in adamantinoma and OFD-like adamantinoma tumours, highlighting the different histopathological characteristics of these bone tumour subtypes. Moreover, RNA-Seq expression profiling analysis identified elevated expression of *DLK1* gene in adamantinomas, serving as a potential molecular biomarker. The present study revealed novel genetic and transcriptomic insights for adamantinoma and OFD-like adamantinoma tumours, allowing to differentiate genetically and transcriptomically between the two lesions and identifying a potential diagnostic marker for adamantinomas.

Introduction

Classic adamantinoma (henceforth referred to as adamantinoma) is a low-grade and slow-growing primary malignant bone tumour that typically arises in long bones, predominantly in the mid-portion of the tibia with or without the involvement of the fibula (1, 2). The tumour can also occur in the fibula without the involvement of the tibia (3). Adamantinoma is a very rare neoplasm that represents less than 1% of all primary bone tumours (4). The tumour is characterised histologically by a variety of morphological patterns, consisting of both dominant epithelial component and areas of osteofibrous dysplasia-like fibro-osseous tissue (5). Adamantinoma commonly occurs in the second to fifth decade of life (median age 25–35) and a slight male predominance is noted (3). The initial symptoms are often indolent and nonspecific; however, the majority of patients present with swelling, symptoms may last >30 years before diagnosis (6). The tumour is locally aggressive and potential local recurrence or metastasis may develop years after the diagnosis of the primary lesion (1).

Adamantinoma of long bone can share similarities in tumour location, radiological features and microscopic characteristics with two other bone lesions, osteofibrous dysplasia (OFD)

and OFD-like adamantinoma. OFD is a rare benign lesion involving the mid-shaft of the tibia with or without involvement of the fibula (2) and usually occurs at a younger age than adamantinoma (20 years or younger) (7). OFD-like adamantinoma tumours also arise in the mid-shaft of the tibia with or without the involvement of the fibula (2), characterised by a uniform predominance of osteofibrous dysplasia-like fibro-osseous tissue and contains scattered clusters of epithelial cells that are highlighted with keratin immunohistochemistry (6).

In view of their overlapping clinical features, a potential histogenetic relationship between adamantinoma, OFD-like adamantinoma and OFD has been speculated. Of these three lesions, adamantinoma is considered to be on the malignant end, OFD on the benign end and OFD-like adamantinoma in the middle of the spectrum (3, 7). In addition, there has been controversial theories about the progression or regression between OFD, OFD-like adamantinoma and adamantinoma (7, 8). However, these relationships remain unresolved with little guidance from the literature.

Recurrent genomic abnormalities involving extra copies of chromosomes 7, 8, 12, 19 or 21 have been reported in adamantinoma and OFD-like adamantinoma, supporting a potential genetic relationship between these entities (3, 6). However, because of their rarity, comprehensive genomic studies of these two bone tumours are few and, therefore, the aetiology and genetic profiles of these tumours remain poorly understood. To date, no stringent molecular diagnostic testing is available for these bone lesions. Using whole exome sequencing (WES) and RNA sequencing (RNA-Seq), we performed comprehensive genetic and transcriptomic analyses on adamantinoma and OFD-like adamantinoma tumours to elucidate the genetic and transcriptomic alteration landscapes of these tumours. To our knowledge, the current study represents the first integrative genomic study using high-throughput sequencing technologies that can provide insights into the molecular and pathogenic mechanisms of adamantinoma and OFD-like

adamantinoma tumours, genetically differentiate between the two entities and guide personalised therapeutic development.

Materials and methods

Tumour samples

A retrospective search of the pathology database at the Royal Orthopaedic Hospital (Birmingham, UK) for resected samples of adamantinoma and OFD-like adamantinoma was conducted. In total, eight cases with the diagnosis of adamantinoma and four OFD-like adamantinoma samples were identified. All cases were thoroughly examined by a specialist bone tumour pathologist (VPS) to confirm their pathological diagnoses. Fresh frozen tumour sample was obtained from all cases. Six of eight adamantinoma and three of four OFD-like adamantinoma tumours had matched-normal samples, obtained from tumour-free muscle tissue adjacent to the tumour site. All samples were retrieved 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 accordance with the ethical rules and regulations presented in the Declaration of Helsinki.

For adamantinoma tumours, the age of the patients ranged between 7–57 (mean 34) at the time of diagnosis. Of these patients, five were male and three were female. The size of adamantinoma tumours ranged from 3–12.5 cm, majority were low grade and occurring in the tibia. For OFD-like adamantinoma tumours, the age of the patients ranged between 6–15 (mean 11). All of these tumours were of low-grade, occurred in the tibia and ranged from 2–8.5 cm in size.

Whole exome sequencing and copy number variation analysis

DNA from fresh frozen was extracted and purified using the DNA Isolation kit (Roche Diagnostic Ltd, UK) following the manufacturer's instructions. A total of 1-3 µg of DNA from 11 tumours (8 adamantinoma; 3 OFD-like adamantinoma) and 9 corresponding normal samples (6 adamantinoma; 3 OFD-like adamantinoma) 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 Illumina HiSeq 2000 (100-bp paired-end) or NextSeq500 (150-bp paired-end) platforms (Illumina, San Diego, CA, USA). One normal-paired tumour was 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.

Using WES data, copy number variation (CNV) analysis was performed on adamantinoma and OFD-like adamantinoma tumours using CNVkit (San Francisco, CA, USA), applying the tool's default settings (https://cnvkit.readthedocs.io/en/stable/).

WES analysis pipeline: reads mapping and variant calling

Sequenced reads were aligned to the human genome reference sequence GRCh37 using BWA 0.7.x bioinformatic tool (9). Optical and PCR duplicates were removed using Picard 1.x (http://picard.sourceforge.net). Somatic SNVs and indels were identified using VarScan2 (10) and MuTects (11) tools, applying the default settings. Variant allele frequency (VAF) of ≥ 10% was selected for all tumours. A series of filtering and prioritisation steps were followed to identify genuine 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 prioritised. Variants were annotated with a modified version of Ensembl Variant Effect Predictor (12).

SIFT and PolyPhen2 bioinformatic tools were utilised to predict the potential pathogenic effect of missense substitutions and only variants that were predicted "deleterious/damaging" by at least one tool were retained. Variants were manually visualised and investigated using Integrative Genomics Viewer (IGV) and artefactual changes were discarded. To eliminate potential germline calls, variants were retained if they were novel or reported in dbSNP build 137 or the Exome Aggregation Consortium (ExAc) datasets with a minor allele frequency (MAF) of <0.1%. Additional stringent steps were applied to three tumours for which matched-normal DNA samples were not available. In these tumours, variants reported in dbSNP or ExAc (regardless of MAF) were discarded, except for variants with <0.01% MAF and described somatic in COSMIC database. The Catalogue Of Somatic Mutations (COSMIC) database (http://cancer.sanger.ac.uk/cosmic) was used to identify previously reported cancer mutations. COSMIC Cancer Gene Census (CCGC) consortium and Integrative Onco Genomics (InTOgen) database (https://www.intogen.org/search) were utilised to identify potential cancer driver genes. To identify potential drug-gene interactions, the Drug Gene Interaction Database (DGIdb) was used (http://www.dgidb.org).

Pathway analysis

Pathway analysis were performed using Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com/) utilising the complete mutated gene list of each tumour subtype separately. IPA was run with the experimentally observed confidence filter to obtain significant analysis output (p-value ≤ 0.05; right tailed Fisher's exact test).

WES variant confirmation by PCR and Sanger sequencing

DNA of the tumour and adjacent normal tissue from the same patient (when available) was amplified by targeting the genomic regions flanking the mutation. microCLEAN (Microzone,

Stourbridge, UK) was used to purify PCR products. Sanger sequencing reactions were performed using the BigDye v3.1 cycle sequencing kit. Prepared sequencing reactions were sequenced on an ABI 3730 DNA analyser (Applied Biosystems, California, USA), following the manufacture's protocol.

RNA Sequencing

Total RNA from five adamantinoma and three OFD-like adamantinoma fresh frozen tissues was extracted using either standard TRIzol-chloroform method or Qiagen RNeasy Mini kit (Qiagen, Manchester, UK) following the manufacturer's instructions. RNA-Seq was performed using a total of 50-100 ng of DNase-treated RNA from each tumour. The tumours were massively sequenced at the GenomicsBirmingham facility (Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK). Briefly, cDNA libraries were constructed using the Neoprep stranded mRNA library prep (Illumina, NP-202-1001) (Illumina, San Diego, CA, USA), according to manufacturer's instructions. Prepared libraries were sequenced on NextSeq 500 (Illumina), producing paired-end reads (75bp in length). The Tuxedo Suite (13) was used for mapping and alignment of the sequenced reads (human genome reference sequence GRCh37). To identify gene fusions, TopHat2 (14) and STAR-Fusion (15) were used and a series of filtering steps were followed to identify genuine gene fusions as described in (16). The *EPHB4-MARCH10* gene fusion was validated by RT-PCR and genomically characterised using long range PCR (LR-PCR) as detailed in (16).

RNA-Seq differential gene expression and hierarchical clustering analyses

The alignment was performed using STAR, version 2.5.3a, on the reference genome GRCh38, Gencode version 28. Differential Gene Expression analysis was performed using the R package DESeq2, version 1.6.3, without beta. Prior assumption and applying

independent filtering on results. Wald test was applied to expressed genes, where expressed genes were defined as genes with at least 1 cpm (count per million reads) on at least three different samples. Significance levels were adjusted using the Benjamini-Hochberg correction, and significant genes were defined as those genes showing an FDR less than 0.05.

Principal Component Analysis was performed on zero-centred and scaled data, using the R package prcomp. Unsupervised clustering of the samples was performed through hierarchical clustering, as implemented in the R package pheatmap. Both algorithms were applied to the top 500 most variable expressed genes. Gene ontology analyses were performed with DAVID v6.8 (https://david.ncifcrf.gov/home.jsp). Functional enrichments were also identified with STRING v10 (17).

Results

The number of WES somatic mutations identified in adamantinoma and OFD-like adamantinoma tumours

To demarcate their genomic landscape, WES was conducted on eight adamantinoma tumours (six normal-paired tumours and two tumours only) and four OFD-like adamantinoma tumours (three normal-paired tumours and one tumour-only) at an average coverage depth of at least 50X in each sequenced sample. A series of stringent filtering and prioritisation steps were followed to identify genuine somatic mutations (details in Materials and Methods). Following this filtering scheme, in eight adamantinoma tumours, a total of 21 insertions and deletions (indels) and 182 single nucleotide variants (SNVs) were identified, corresponding to a median of 0.84/megabase (range, 0–2.57) per coding megabase (Figure 1A & Supplemental Digital Content, Table S1). Somatic SNVs consisted of: 156 missense (85.7%), 19 nonsense (10.4%), and 7 splice site (3.8%) alterations. The low overall somatic mutation burden identified in adamantinomas is similar

to that recently documented in undifferentiated pleomorphic sarcoma of bone (2.2/megabase) (16), 206 soft tissue sarcomas (1.06/megabase) (18) and in Ewing sarcoma (0.42/megabase) (19).

In four OFD-like adamantinoma tumours, a total of 7 indels and 59 SNVs were detected (Figure 1B & Supplemental Digital Content, Table S1). These SNVs comprised 54 missense (91.5%), 3 nonsense (5.1%) and 2 splice site (3.4%) changes. Two of these tumours lacked any WES variants, one tumour harboured one indel only, whereas the fourth sample (OFD-like adamantinoma T4) harboured the remaining of the SNV and indel alterations. The average mutation rate in the first three OFD-like adamantinoma tumours is almost zero; however, taken together, in the four OFD-like adamantinoma tumours, the average coding mutation rate is 2.20/megabase; (Supplemental Digital Content, Table S1).

Somatic copy number variation analysis using WES data

Somatic CNV analyses were conducted on eight adamantinoma and four OFD-like adamantinoma tumours utilising WES data and using CNVkit tool. The overall somatic CNV heatmaps for adamantinoma and OFD-like adamantinoma tumours are shown in Supplemental Digital Content, Figure S1 & Figure S2, respectively. In adamantinoma tumours, recurrent chromosomal gains in chromosomes 7, 8, 12, and 19, consistent with previously published data (6). In addition, we found recurrent chromosomal gains in chromosomes 1q, 2, and 10 as well as recurrent chromosomal losses in chromosomes 1p, 3, 5, 6, 15, and 17. In comparison to adamantinoma tumours, we observed no prominent chromosomal gains or losses in the four OFD-like adamantinoma tumours, distinguishing the genetic CNV profile of OFD-adamantinoma tumours from adamantinomas.

We then assessed the CNV status of the most frequently mutated genes in all cancer types as reported in the COSMIC database. In adamantinoma tumours, recurrent CNV

gains and losses were found in *BRAF* (4/8; 50%), *JAK2* (3/8; 38%), *KRAS* (3/8, 38%) *TP53* (2/8; 25%) and *EGFR* (4/8; 50%) (Figure 1C). In contrast, in OFD-like adamantinomas, these genes were not altered, except for non-recurrent CNV losses in *JAK2* and *KRAS* that were identified in OFD-like adamantinoma T4 (Figure 1C).

KMT2D is recurrently mutated in adamantinoma tumours

Point mutations in *KMT2D* (*MLL2*) were identified in 2/8 adamantinoma tumours (25%) (Figure 1C and Figure 2). The first *KMT2D* mutation detected, c.16154C>A; p.Ser5385Ter, is a nonsense change described as somatic (COSMIC database) in lung carcinoma and squamous cell lung carcinoma tumours (ID: COSM3954819). Using Sanger sequencing, this variant was confirmed somatic in sample ADA-T2 (present in tumour DNA but absent in the corresponding normal DNA) (Supplemental Digital Content, Figure S3). The second *KMT2D* mutation identified in ADA-T8, c.10930C>T; p.Pro3644Ser, is a missense substitution predicted deleterious by SIFT and PolyPhen2 *in silico* tools. This mutation was confirmed in the tumour DNA using Sanger sequencing (Supplemental Digital Content, Figure S3); however, no corresponding normal sample was available to confirm the somatic status of the mutation. Nevertheless, this mutation has not been reported previously in the Exome Aggregation Consortium (ExAc) or dbSNP databases (novel), thus increasing the likelihood of the mutation being a true somatic change.

To assess genomic CNVs in *KMT2D*, the CNVkit data outputs of adamantinoma and OFD-like adamantinoma tumours were analysed. In adamantinoma tumours, CNV gains were identified in two adamantinoma samples, ADA-T2 and ADA-T3. Combining SNVs and CNVs, *KMT2D* is mutated in 3/8 adamantinoma tumours (38%). By contrast, no CNV alterations in *KMT2D* were identified in OFD-like adamantinoma tumours.

In OFD-like adamantinoma tumours, no gene was mutated in more than one sample. However, in OFD-like adamantinoma tumour T4, a missense substitution (c.6437C>T; p.Pro2146Leu) in *KMT2D* was identified (Figure 1C and Figure 2). This variant is

described as somatic (COSMIC database) in acute myeloid leukaemia and malignant melanoma tumours (COSM3461578). Sanger sequencing confirmed this mutation in tumour DNA but no DNA from corresponding normal sample was available for somatic status testing (Supplemental Digital Content, Figure S3).

To infer its role in tumourigenesis and potential druggability, *KMT2D* was assessed against COSMIC, InTOgen and DGIdb databases. From the COSMIC Cancer Gene Census (CCGC) list, *KMT2D* is classified as a 'causally implicated gene in cancer'. In addition, the inTOgen database categorises *KMT2D* as a mutational cancer driver in the following cancer types: prostate adenocarcinoma, bladder carcinoma, breast carcinoma, medulloblastoma, lung squamous cell carcinoma, and head and neck squamous cell carcinoma. In DGIdb, a potential drug-gene interaction for *KMT2D* is identified, classifying this gene as potentially druggable.

Pathway analyses

To identify putatively enriched pathways in adamantinoma and OFD-like adamantinoma tumours, the complete list of mutated genes of each tumour subtype was analysed using IPA tool. A number of significantly enriched pathways were identified in each tumour subtype (Supplemental Digital Content, Table S2). Calcium signalling and cell morphology pathways were commonly altered in both tumour subtypes.

RNA-Seq identifies *EPHB4-MARCH10* somatic gene fusion in an adamantinoma tumour

RNA-Seq was performed on five adamantinoma and three OFD-like adamantinoma tumours, achieving a mean of 69,785,777 reads with 93.88% successful alignment rate to reference genome. To identify genuine somatic gene fusions, two fusion caller tools were used, followed by a series of filtering and prioritisation steps (see Materials and Methods). No somatic gene fusions were detected in OFD-like adamantinoma tumours. However, one somatic gene fusion, *EPHB4-MARCH10*, was identified in an adamantinoma tumour

(ADA-T4) (Figure 1C & Supplemental Digital Content, Figure S4). *EPHB4-MARCH10* is a novel gene fusion and occurs by joining the first eight exons of the 5′-gene partner *EPHB4* (Transcript ID: ENST00000358173.3) to the last exon (exon 2) of the 3′-gene partner *MARCH10* (Transcript ID: ENST00000582568.1) (Supplemental Digital Content, Figure S4 A). Using RT-PCR and Sanger sequencing, this gene fusion was detected in tumour cDNA but not in the corresponding normal material, suggesting that the fusion is highly tumour-specific (Supplemental Digital Content, Figure S4 B). Genomic characterisation analysis using LR-PCR and Sanger sequencing for *EPHB4-MARCH10* was performed (Supplemental Digital Content, Figure S4 C & D). The gene fusion was confirmed in the tumour DNA but was not detected in corresponding normal material, confirming the somatic status of the fusion. Genomic breakpoint analysis indicated that this gene fusion occurs as a result of a genomic translocation rearrangement between chromosomes 7 and 17, fusing the fusion gene partners. The *EPHB4-MARCH10* chimeric transcript is in-frame and the predicted translated chimeric product is 640 amino acid in length, consisting of 529 amino acids from *EPHB4* and 111 amino acids belonging to *MARCH10* (Figure 3). In this chimeric transcript, the organisation of the protein domains of EPHB4 is as follows: the ephrin receptor ligand binding, putative ephrin-receptor like and fibronectin type III (repeats) functional domains are retained, while ephrin type-A receptor 2 transmembrane, catalytic protein tyrosine kinase and sterile alpha motif domains are completely lost (Figure 3). Using DGIdb, multiple drug-gene interactions were identified for *EPHB4*, classifying this gene as a potential drug target.

Hierarchical clustering and differential expression analyses using RNA-Seq data

Using RNA-Seq data, unsupervised hierarchical clustering analysis was performed on five adamantinoma and three OFD-like adamantinoma tumours. The tumours of each lesion subtype distinctly clustered together, thus molecularly distinguishing adamantinoma from OFD-like adamantinoma tumours (Figure 4A).

Supervised differential expression analysis of adamantinoma versus OFD like adamantinoma highlighted *DLK1*, Delta Like Non-Canonical Notch Ligand 1, as a specifically expressed gene in adamantinomas. The complete list of differentially expressed genes is provided in Supplemental Digital Content, Table S3. David Gene Ontology analyses of specific genes expressed in adamantinoma and OFD-like adamantinoma tumours identified strong enrichment of genes involved in epidermis/epithelium development and collagen biosynthesis and modifying enzymes, respectively (Figure 4 B, C). The tumor-specific enrichments of genes involved in the development/organization of epithelium and bone were confirmed with STRING tool. This last analysis also suggested an increased kinase activity for Adamantinomas (Supplemental Digital Content, Table 4).

Discussion

In the current study, we investigated the genetic and transcriptomic alteration landscapes in adamantinoma and OFD-like adamantinoma bone tumours using WES and RNA-Seq technologies We compared the transcriptomic profiles of adamantinoma and OFD-like adamantinoma tumours and identified a potential biomarker for adamantinoma tumours.

WES was conducted on eight adamantinoma and four OFD-like adamantinoma tumours. In OFD-like adamantinoma tumours, no recurrent gene mutations or somatic gene fusions were identified. Moreover, no prominent chromosomal gains or losses were detected in these tumours in comparison to adamantinomas. By contrast, in adamantinoma tumours, *KMT2D* was mutated in 38% of the samples. In addition, multiple recurrent chromosomal gains and losses (novel and previously reported) were identified in adamantinoma samples these included gains and losses in well-known cancer genes (*BRAF, JAK2, KRAS, TP53*, and *EGFR*).

In adamantinoma tumours, *KMT2D* was recurrently mutated (3/8 tumours; 38%), for which two tumours harboured two SNVs and two tumours exhibited CNV gains. *KMT2D*, encoding histone lysine N-methyltransferase 2D, is a histone-modifier gene that plays an important role in regulating chromatin structures and DNA accessibility. Many somatic mutations in *KMT2D* have been reported in multiple cancers, including non-Hodgkin lymphomas [20], acute myeloid leukaemia, lung carcinomas [21] and histiocytic sarcoma [22]. Moreover, *KMT2D* is one of the most frequently mutated genes in multiple paediatric malignancies [23]. Here, we report for the first time the association of *KMT2D* alterations in adamantinoma bone tumours, linking the potential role of epigenetic regulation in the pathogenesis of this tumour subtype.

The majority of *KMT2D* reported mutations are nonsense or frameshift changes that are predicted to produce truncated products [24]. In the ADA-T2 sample of the current study, we report a nonsense (truncating) mutation (c.16154C>A; p.Ser5385Ter) occurring just upstream of the SET domain, thus producing a protein product that lacks the entire SET domain at the C-terminus (Figure 2). Frameshift and nonsense changes that affect the SET domain account for 37% of total *KMT2D* mutations [24]. The enzymatically active SET domain has an important role in H3K4 methyltransferase activity [25]. Set1-containing lysine methyltransferases (KMTs), including KMT2D, are part of large highly conserved multicomponent complexes known as SET/COMPASS (COMplex of Proteins ASsociated with Set) [26]. The SET/COMPASS complexes are key regulators for the majority of global H3K4 dimethylation and trimethylation, which play a role in gene expression regulation [27].

To infer the potential pathological role of SET domain in mice, a study by [28] inactivated the methyltransferase activity of Kmt2C (*Kmt2C$^{\Delta SET/\Delta SET}$*), a closely related paralogue to *Kmt2D* with extended SET domain similarity. The *Kmt2C$^{\Delta SET/\Delta SET}$* mice showed incomplete embryonic lethality and exhibited cellular hyperproliferation and kidney ureter urothelium

tumours, accompanied with increased levels of DNA damage. Therefore, the truncating *KMT2D* mutation reported in the current study, which leads to complete loss of the SET domain, is likely tumourigenesis-related and is possibly characterised by loss or reduced tumour suppressing activity.

The second recurrent *KMT2D* missense mutation identified in ADA-T8 (c.10930C>T; p.Pro3644Ser) does not reside in, or in close proximity of annotated functional domains of *KMT2D* (Figure 2). Further investigations are required to elucidate the pathological role of this mutation.

In an OFD-like adamantinoma tumour (T4), the missense mutation (c.6437C>T; p.Pro2146Leu) in *KMT2D* is located in close proximity of the high mobility group (HMG-I) domain (Figure 2). HMG-I domain is comprised of short motifs that interact with the minor groove of DNA sequences, thus regulating DNA replication and transcription (24).

In the second part of the study, we aimed to identify somatic gene fusions arising from chromosomal rearrangements in five adamantinoma and three OFD-like adamantinoma tumours. As mentioned earlier, no genuine somatic gene fusions were identified in OFD-like adamantinoma samples. In contrast, in adamantinoma tumours, a novel somatic gene fusion (*EPHB4-MARCH10*) was detected, validated as highly tumour-specific by RT-PCR, characterised genomically and confirmed somatic at the DNA level using LR-PCR.

EPHB4-MARCH10 gene fusion was identified in ADA-T4 sample that lacked *KMT2D* SNV or CNV alterations. The *EPHB4* gene encodes a receptor tyrosine kinase (RTK) that belongs to the erythropoietin-producing hepatocellular carcinoma (Eph) family, the largest class of RTKs (29). RTKs are transmembrane proteins that regulate key biological processes, including cell differentiation and proliferation. Impaired Eph/ephrin cellular interactions have been implicated in carcinogenesis, contributing to tumour growth, metastasis, chemoresistance and tumour angiogenesis (30). Mutations in *EPHB4* have been reported in several lung cancers including lung adenocarcinoma and small cell lung

cancer (31). *MARCH10* belongs to the MARCH family membrane-bound E3 ubiquitin ligases that are involved in endocytosis, protein sorting and lysosomal degradation (32).

Bidirectional signalling is a hallmark of the Eph receptor and ephrin ligand signalling system which consists of forward signalling in receptor-expressing cells and reverse signalling in ligand-expressing cells (33). The Eph/ephrin signalling network regulates cell morphology, adhesion, migration and, in the context of tumour growth, this signalling network promotes tumour growth, invasiveness and tumour angiogenesis (30). In cancer cells, the role of EPHB4 is complex; however, both tumour-suppressing and tumour-promoting (oncogenic) activities have been described (34). In the *EPHB4-MARCH10* fusion reported in the current study, three of EPHB4 functional domains are completely lost, thus possibly interfering with the Eph/ephrin bidirectional system. Nevertheless, further investigations are needed to determine the pathological role of this gene fusion.

Unsupervised clustering analysis using RNA-Seq data distinctly clustered the tumours adamantinoma and OFD-like adamantinoma subtypes, confirming their pathological diagnoses and allowing for molecular differentiation between these tumour subtypes. The clustering analysis can be beneficial for tumours such as OFD-like adamantinoma T4 for which a mutational profile is similar to those in adamantinomas.

Supervised differential expression profiling of adamantinoma and OFD-like adamantinoma tumours revealed an elevated expression of *DLK1* in the former bone tumour subtype. *DLK1* is a member of the Epidermal Growth Factor (EGF)-like repeat-containing family of proteins. DLK1 protein plays a role in modulating adipogenesis (35) and regulating osteoblast differentiation (36). Abberant expression of *DLK1* has been observed in multiple human cancers, including neuroblastoma (37) and human prostate cancer (38). A study by (39) assessed the expression of Dlk1 in tumours derived from skeletal muscle, adipose tissue and bone. The authors identified Dlk1 expression in skeletal muscle derived tumours (18/21 cases) and liposarcomas (adipose tissue tumours; 5/19); however, no

expression of Dlk1 was observed in nine bone-derived tumours (osteogenic sarcoma, osteoblastoma, osteochondroma). In the current study, the elevated expression of *DLK1* in adamantinomas can serve as a molecular biomarker specific to this tumour subtype that can be diagnostically utilised in clinics; nevertheless, additional confirmatory experiments are required.

The results of David and STRING analyses correlated with the different histopathological characteristics of the two tumor types (40). In fact, enrichments of genes involved in the development/organization of epithelium and bone were identified for Adamantinomas and OFD-like adamantinomas, respectively. Moreover, the identification of increased kinase activity in adamantinomas can suggest therapeutic targets for these tumors. Of note, cases of metastatic adamantinomas responding to kinases inhibitors have already been described (41, 42).

In conclusion, the current study represents the first comprehensive genetic and transcriptomic alteration landscapes of adamantinoma and OFD-like adamantinoma tumours using NGS technologies, indicating a genetic differentiation between these two entities and expanding our understanding of their aetiology and pathology. We reported *KMT2D* as a novel recurrently mutated gene in adamantinoma tumours, highlighting the role of deregulated chromatin remodelling mechanisms in pathogenesis of adamantinomas and providing insights into developing *KMT2D* targeted therapeutics. In addition, we identified a novel *EPHB4-MARCH10* somatic gene fusion that may be a tumour driver. Overall, we showed different genetic and transcriptomic profiles for adamantinoma and OFD-like adamantinoma tumours, therefore molecularly differentiating between these two bone tumour subtypes. The elevated expression of *DLK1* identified in adamantinoma tumours may serve as a potential diagnostic biomarker. However, the small cohort size is a limitation of this study and finding additional cases for these ultra-rare bone tumours should be pursed in further studies.

Reference list

1.   Horvai A, Dashti NK, Rubin BP, et al. Genetic and molecular reappraisal of spindle cell adamantinoma of bone reveals a small subset of misclassified intraosseous synovial sarcoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2018:1-11.

2.   Kahn LB. Adamantinoma, osteofibrous dysplasia and differentiated adamantinoma. *Skeletal radiology.* 2003;32:245-258.

3.   Jain D, Jain VK, Vasishta RK, et al. Adamantinoma: a clinicopathological review and update. *Diagnostic pathology.* 2008;3:8.

4.   Puchner SE, Varga R, Hobusch GM, et al. Long-term outcome following treatment of Adamantinoma and Osteofibrous dysplasia of long bones. *Orthopaedics & traumatology, surgery & research : OTSR.* 2016;102:925-932.

5.   Taylor RM, Kashima TG, Ferguson DJ, et al. Analysis of stromal cells in osteofibrous dysplasia and adamantinoma of long bones. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2012;25:56-64.

6.   Christopher D, Fletcher JA, Bridge P. WHO classification of tumours of soft tissue and bone. *International agency for research on cancer 4th edition Lyon.* 2013.

7.   Camp MD, Tompkins RK, Spanier SS, et al. Best cases from the AFIP: Adamantinoma of the tibia and fibula with cytogenetic analysis. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2008;28:1215-1220.

8.   Scholfield DW, Sadozai Z, Ghali C, et al. Does osteofibrous dysplasia progress to adamantinoma and how should they be treated? *The bone & joint journal.* 2017;99-b:409-416.

9.   Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754-1760.

10.   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.

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

12.   McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26:2069-2070.

13.   Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols.* 2012;7:562-578.

14.   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.

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

16.   Ali NM, Niada S, Brini AT, et al. Genomic and transcriptomic characterisation of undifferentiated pleomorphic sarcoma of bone. *J Pathol.* 2018.

17.   Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447-452.

18.   The Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell.* 2017;171:950-965.e928.

19.   Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499:214-218.

20. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature.* 2011;476:298-303.

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

22. Hung YP, Lovitch SB, Qian X. Histiocytic sarcoma: New insights into FNA cytomorphology and molecular characteristics. *Cancer cytopathology.* 2017;125:604-614.

23. Huether R, Dong L, Chen X, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun.* 2014;5:3630.

24. Rao RC, Dou Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nature reviews Cancer.* 2015;15:334-346.

25. Dorighi KM, Swigut T, Henriques T, et al. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular cell.* 2017;66:568-576.e564.

26. Ford DJ, Dingwall AK. The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer genetics.* 2015;208:178-191.

27. Froimchuk E, Jang Y, Ge K. Histone H3 lysine 4 methyltransferase KMT2D. *Gene.* 2017;627:337-342.

28. Lee J, Kim DH, Lee S, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proc Natl Acad Sci U S A.* 2009;106:8513-8518.

29. Noren NK, Pasquale EB. Paradoxes of the EphB4 receptor in cancer. *Cancer Res.* 2007;67:3994-3997.

30. Merchant AA, Jorapur A, McManus A, et al. EPHB4 is a therapeutic target in AML and promotes leukemia cell survival via AKT. *Blood advances.* 2017;1:1635-1644.

31.   Ferguson BD, Tan YH, Kanteti RS, et al. Novel EPHB4 Receptor Tyrosine Kinase Mutations and Kinomic Pathway Analysis in Lung Cancer. *Scientific reports.* 2015;5:10641.

32.   Morokuma Y, Nakamura N, Kato A, et al. MARCH-XI, a novel transmembrane ubiquitin ligase implicated in ubiquitin-dependent protein sorting in developing spermatids. *The Journal of biological chemistry.* 2007;282:24806-24815.

33.   Zhang G, Brady J, Liang WC, et al. EphB4 forward signalling regulates lymphatic valve development. *Nat Commun.* 2015;6:6625.

34.   Rutkowski R, Mertens-Walker I, Lisle JE, et al. Evidence for a dual function of EphB4 as tumor promoter and suppressor regulated by the absence or presence of the ephrin-B2 ligand. *International journal of cancer.* 2012;131:E614-624.

35.   Nueda ML, Garcia-Ramirez JJ, Laborda J, et al. dlk1 specifically interacts with insulin-like growth factor binding protein 1 to modulate adipogenesis of 3T3-L1 cells. *J Mol Biol.* 2008;379:428-442.

36.   Abdallah BM, Boissy P, Tan Q, et al. dlk1/FA1 regulates the function of human bone marrow mesenchymal stem cells by modulating gene expression of pro-inflammatory cytokines and immune response-related factors. *The Journal of biological chemistry.* 2007;282:7339-7351.

37.   Begum A, Kim Y, Lin Q, et al. DLK1, delta-like 1 homolog (Drosophila), regulates tumor cell differentiation in vivo. *Cancer letters.* 2012;318:26-33.

38.   Ceder JA, Jansson L, Helczynski L, et al. Delta-like 1 (Dlk-1), a novel marker of prostate basal and candidate epithelial stem cells, is downregulated by notch signalling in intermediate/transit amplifying cells of the human prostate. *European urology.* 2008;54:1344-1353.

39.   Jorgensen LH, Sellathurai J, Davis EE, et al. Delta-like 1 homolog (dlk1): a marker for rhabdomyosarcomas implicated in skeletal muscle regeneration. *PloS one.* 2013;8:e60692.

40.   Gleason BC, Liegl-Atzwanger B, Kozakewich HP, et al. Osteofibrous dysplasia and adamantinoma in children and adolescents: a clinicopathologic reappraisal. *Am J Surg Pathol.* 2008;32:363-376.

41.   Dudek AZ, Murthaiah PK, Franklin M, et al. Metastatic adamantinoma responds to treatment with receptor tyrosine kinase inhibitor. *Acta Oncol.* 2010;49:101-104.

42.   Cohen Y, Cohen JE, Zick A, et al. A case of metastatic adamantinoma responding to treatment with pazopanib. *Acta Oncol.* 2013;52:1229-1230.

43.   Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun.* 2015;6:5897.

Figure legends

Figure 1. The total number of somatic alterations and WES and RNA Sequencing alterations landscape in adamantinoma and OFD-like adamantinoma tumours. The total somatic alterations in (A) adamantinomas and (B) OFD-like adamantinoma tumours, single nucleotide variants (SNVs) consist of missense alterations that are predicted deleterious by at least one *in silico* tool (SIFT or PolyPhen2), nonsense and splice site changes. INDELs consist of small in-frame or out-of-frame insertions or deletions. (C) The WES and RNA alterations landscape of OFD-like and adamantinoma tumours. WES was conducted on eight adamantinoma and four OFD-like adamantinoma tumours. *KMT2D* was mutated in two adamantinoma samples. RNA-sequenced samples are underlined. ADA: adamantinoma; OFD-like: OFD-like adamantinoma. ADA: adamantinoma; OFD-like: OFD-like adamantinoma.

Figure 2. Somatic *KMT2D* mutations and their relative protein positions identified in adamantinoma and OFD-like adamantinoma tumours. The black balls resemble mutations and their amino acid (AA) locations. Two *KMT2D* mutations were identified in two adamantinoma tumours (S5385Ter, nonsense; P3644S, missense). One *KMT2D* mutation was detected in an OFD-like adamantinoma tumour (P2146L; missense). Zf-HC: PHD-like zinc-binding domain; PHD: PHD zinc finger; HMG-I: high-mobility-group (HMG)-box domain; SET: SET domain.

Figure 3. The protein domain organization of the gene partners of *EPHB4-MARCH10* gene fusion. The grey shaded area resembles the retained protein domains of the fused exons of *EPHB4* and *MARCH10* gene partners. The gene fusion breakpoints are denoted by a black double slash. The gene fusion is in-frame, resulting from joining the first 529 amino acids of *EPHB4* to the last 111 amino acids of *MARCH10*.

Figure 4. Unsupervised and supervised analysis of adamantinoma and OFD-like adamantinoma tumours using RNA-Seq data. (A) Un supervised clustering analysis of

adamantinoma and OFD-like adamantinoma tumours clearly clustered the tumours of each subtype in two groups. David Gene Ontology analyses of differentially expressed genes for (B) adamantinoma and (C) OFD-like adamantinoma tumours indicated strong enrichment for genes involved in epidermis development and collagen biosynthesis and modifying enzymes, respectively.

Supplemental table and figure legends.

Table S1. Coding mutation rate of adamantinoma and OFD-like adamantinoma tumours. The mutational rate is calculated by dividing the number of identified mutations by the total length of the exome (30 megabase). ADA: adamantinoma; OFD-like: OFD-like adamantinoma.

Table S2. Canonical and bio function pathways affected in adamantinoma and OFD-like adamantinoma tumours. Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) software, utilising the complete mutated gene list of each tumour subtype separately.

Table S3. List of differentially expressed genes of adamantinoma versus OFD-like adamantinma. Genes overexpressed in adamantinoma tumous have '1' significance, whereas genes overexpressed in OFD-like adamantinoma tumours have '-1' significance.

Table S4. Functional enrichment of genes in the Adamantinoma and OFD-like Adamantinoma networks identified by STRING v10. Only the top 10 results in the Biological Process, Cellular Component and Molecular function GO categories are shown.

Figure S1. Somatic copy number variation heatmap of adamantinoma tumours. Genomic gains and losses regions were identified by CNVkit tool using WES data. ADA: adamantinoma.

Figure S2. Somatic copy number variation heatmap of OFD-like adamantinoma tumours. Genomic gains and losses regions were identified by CNVkit tool using WES data. OFD-like: OFD-like adamantinoma.

Figure S3. Sanger sequencing confirmation of *KMT2D* mutations in adamantinoma and OFD-like adamantinoma tumours. ADA: adamantinoma; OFD-like: OFD-like adamantinoma.

Figure S4. Schematic representation and validation of *EPHB4-MARCH10* gene fusion. (A) The genomic structure of gene fusions partners, *EPHB4* (orange) and *MARCH10* (blue). (B) Sanger sequencing validation of the gene fusion, confirming the fusion breakpoint at the cDNA level (by RT-PCR) and showing an in-frame chimeric product. (C) Diagrammatic representation of the gene fusion at the genomic (43) level and mapping of the gene fusion breakpoints. (D) LR-PCR analysis using Sanger sequencing confirming the gene fusion breakpoints at the DNA level, resulting from a translocation between chromosomes 7 and 17.

List of suplemental digital content

Supplemental table and figure legends. Word

Table S1. Excel

Table S2. Excel

Table S3. Excel

Table S4. Excel

Figure S1. TIFF

Figure S2. TIFF

Figure S3. TIFF

Figure S4. TIFF