Modern Pathology (2020) 33:1331–1340
https://doi.org/10.1038/s41379-020-0457-8

ARTICLE

PHF1 fusions cause distinct gene expression and chromatin accessibility profiles in ossifying fibromyxoid tumors and mesenchymal cells

Jakob Hofvander[1] · Vickie Y. Jo[2] · Christopher D. M. Fletcher[2] · Florian Puls[3] · Uta Flucke[4] · Jenny Nilsson[1] · Linda Magnusson[1] · Fredrik Mertens[1,5]

Received: 28 October 2019 / Revised: 3 January 2020 / Accepted: 3 January 2020 / Published online: 13 January 2020

Abstract

Ossifying fibromyxoid tumor (OFMT) is a soft tissue tumor frequently displaying gene fusions, most of which affect the *PHF1* gene. *PHF1* encodes plant homeodomain finger protein 1, which is involved in various processes regulating gene transcription, including those orchestrated by the polycomb repressor complex 2. Here, a series of 37 OFMTs, including 18 typical, 9 atypical, and 10 malignant variants, was analyzed with regard to transcriptomic features, gene fusion and copy number status, and/or single-nucleotide variants. The effects on gene expression and chromatin accessibility of three detected fusions (*EP400–PHF1*, *MEAF6–PHF1*, and *PHF1–TFE3*) were further evaluated in fibroblasts. Genomic imbalances showed a progression-related pattern, with more extensive copy number changes among atypical/malignant lesions than among typical OFMTs; loss of the *RB1* gene was restricted to atypical/malignant OFMTs, occurring in one-third of the cases. RNA sequencing identified fusion transcripts in >80% of the cases analyzed, including a novel *CSMD1–MEAF6*. The gene-expression profile of OFMT was distinct from that of other soft tissue tumors, with extensive transcriptional upregulation of genes in OFMT. These findings were largely recapitulated in gene fusion-expressing fibroblast lines, suggesting that genes involved in, e.g., Wnt signaling and/or being regulated through trimethylation of lysine 27 in histone 3 (H3K27me3) are pivotal for OFMT development. The genes showing differentially higher expression in fusion-expressing cells paralleled increased chromatin accessibility, as revealed by ATAC sequencing. Thus, the present study suggests that OFMT develops through gene fusions that have extensive epigenetic consequences.

Introduction

Ossifying fibromyxoid tumor (OFMT) is a soft tissue neoplasm with highly variable clinical features; although it typically affects adults (median age 50 years) it may occur also in children and tumors may arise anywhere in the body, either subcutaneously or within skeletal muscle [1, 2]. OFMT derives its name from its morphological characteristics, namely nests and cords of round to ovoid cells in a fibromyxoid stroma usually surrounded by an incomplete layer of bone [3]. On the basis of nuclear grade, cellularity and mitotic rate, OFMT can be subdivided into typical, atypical, and malignant tumors [1]. Local recurrences and metastases can occur in all three subtypes, but more often among the malignant lesions [1, 4].

Gene fusions constitute important pathogenetic events in OFMT development. More than 85% of the cases, including typical, atypical, as well as malignant lesions, harbor a gene fusion [5, 6]. The most common ones (accounting for ~50%

Supplementary information The online version of this article (https://doi.org/10.1038/s41379-020-0457-8) contains supplementary material, which is available to authorized users.

✉ Fredrik Mertens
Fredrik.mertens@med.lu.se

1 Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden

2 Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

3 Department of Clinical Pathology and Genetics, Sahlgrenska University Hospital, Gothenburg, Sweden

4 Department of Pathology, Radboud University Medical Centre, Nijmegen, The Netherlands

5 Department of Clinical Genetics and Pathology, Skåne University Hospital, SE-221 85 Lund, Sweden

of all OFMTs) all involve the *PHF1* gene—*EP400–PHF1*, *MEAF6–PHF1*, *EPC1–PHF1*, and *PHF1–TFE3*—but rare cases with *ZC3H7B–BCOR*, *CREBBP–BCORL1*, and *KDM2A–WWTR1* fusions have also been described [6–9]. A common feature among these fusions is that one of the proteins (PHF1, BCOR, BCORL1, KDM2A) in the fusion chimeras is involved in histone modification. Apart from the gene fusions, which presumably constitute the primary genetic events, little is known about the genetics of OFMT or the impact of secondary mutations on tumor progression.

Here, we have characterized the genetic features of a cohort of typical, atypical, and malignant OFMTs using single-nucleotide polymorphism (SNP) arrays for copy number analysis and, in selected cases, RNA sequencing (RNA-seq) for gene fusion assessment and global gene-expression profiling, and whole-exome sequencing (WES) for the detection of sequence variants. In addition, the cellular effects of three of the recurrent fusions, *EP400–PHF1*, *MEAF6–PHF1*, and *PHF1–TFE3*, were investigated in fibroblasts using the Tet-On 3G inducible gene-expression system. This allowed us to compare changes in gene expression in vivo with those in vitro, identifying several downstream targets of the fusion proteins. Furthermore, we were able compare the effects on gene expression with changes in chromatin accessibility by performing an assay for transposase-accessible chromatin (ATAC-seq).

Methods

Patients and tumors

The study included tumors from 37 patients with OFMT (Supplementary Table 1). Tissue samples for genetic analyses were derived from sarcoma centers in Boston, Lund, Nijmegen, Birmingham, and Stockholm. Tumors were classified into typical ($n = 18$), atypical ($n = 9$), or malignant ($n = 10$) OFMT (by VYJ and CDMF), as described [7]. The morphology, cytogenetics, FISH, and SNP-array results of cases 2, 19, and 28 have been reported before [7].

The global gene-expression profiles of seven fresh-frozen OFMT were compared with those of 50 other soft tissue tumors, including dermatofibrosarcoma protuberans ($n = 10$), myxoinflammatory fibroblastic sarcoma ($n = 8$), inflammatory leiomyosarcoma ($n = 5$), myxoid liposarcoma (MLS; $n = 7$), pleomorphic liposarcoma ($n = 5$), PRDM10-rearranged tumor ($n = 6$), and sclerosing epithelioid fibrosarcoma (SEF; $n = 9$). In addition, RNA-seq data on three formalin-fixed paraffin-embedded (FFPE) OFMT were used to study the expression levels of selected genes and were compared with corresponding data from 31 other soft tissue tumors, including angiofibroma ($n = 5$), calcifying aponeurotic fibroma ($n = 6$), SEF ($n = 15$), and MLS ($n = 5$). All tumors used for comparison with OFMT were diagnosed according to established criteria [10].

Immunohistochemical (IHC) analysis with regard to trimethylation of lysine 27 on histone H3 (H3K27me3) and expression of β catenin (CTNNB1) was performed on five tumors (Supplementary Table 1) using 4 µM FFPE sections after deparaffinization at 42 °C. Antigen retrieval was performed using citrate buffer (pH 6.0) and pressure cook. For β catenin, a mouse monoclonal antibody (BD Bioscience, #610154) was used at a 1:1000 dilution; H3K27me3 antibody (Cell Signaling Technology, Danvers, MA) was used at a 1:500 dilution. Appropriate positive and negative controls were used throughout. β-catenin immunohistochemistry was scored according to nuclear staining, and H3K27me3 was interpreted as "lost" only if nuclear staining was absent in lesional cells with intact staining in nonneoplastic cells (i.e., endothelial cells and inflammatory cells).

All samples were obtained after informed consent and the study was approved by the institutional review boards of the participating sarcoma centers.

Cell lines

The TERT-immortalized Bj5ta human fibroblast cell line had previously been transduced with the pLVX-Tet3G vector encoding the regulator Tet-On 3G protein [11]. Cells were further transduced with one of four different response plasmids; E–P encoding the *EP400–PHF1* fusion found in case 20, M–P encoding the *MEAF6–PHF1* fusion found in case 29, P–T encoding the *PHF1–TFE3* fusion found in case 2, or EV containing an empty vector. Transductions were performed with the RetroNectin-bound virus infection method according to the manufacturer's instructions (Takara/Clontech, Gothenburg, Sweden).

Response plasmids were selected for by adding puromycin at a final concentration of 0.5 µg/ml. The transcription of the inserted gene constructs was turned on by adding doxycycline (dox) to the culture medium at a final concentration of 1 µg/ml for 48 h. Cells were harvested and pelleted in RLT buffer (Qiagen, Valencia, CA) with 1% 2-Mercaptoethanol.

RNA-seq

RNA was extracted from fresh-frozen samples and from the Bj5ta fibroblast cell lines as described [11]. cDNA libraries were prepared from poly-A selected RNA using the TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, USA) according to the manufacturer's instructions. Paired-end 151 bp reads were generated from the cDNA libraries on a NextSeq 500 (Illumina).

Unstained 4 µM FFPE sections mounted on slides from 20 cases and unstained 10 µM FFPE scrolls from two

cases were available for RNA extraction. RNA of sufficient quantity and quality, i.e., with DV_{200} values ≥ 30%, could be obtained in 10 cases, using Qiagen's RNeasy FFPE Kit (Qiagen, Valencia, California). cDNA libraries were prepared from 20 to 400 ng of RNA, depending on the DV_{200} value, using the capturing chemistry of the TruSeq RNA Access Library Prep Kit (Illumina) or using the TruSight RNA Fusion Panel (Illumina) according to the manufacturer's instructions. The gene fusion panel targets a total of 507 genes and for all previously reported gene fusions in OFMT at least one of partner genes was included in the list of target genes. Paired-end 85 nt reads were generated from the cDNA libraries on a NextSeq 500 (Illumina).

ChimeraScan 0.4.5 and FusionCatcher 1.0 with default settings and STAR-Fusion 1.4.0 with parameters --min_-junction_reads 0 --FusionInspector validate were used to identify candidate fusion transcripts from the sequence data [12–14]. The GRCh37/hg19 build was used as the human reference genome.

The raw unfiltered RNA-seq reads were aligned to human reference genome hg19 using STAR 2.5.0a [15]. Gene-expression values were calculated as fragments per kilobase of transcript per million reads (FPKM) using Cufflinks 2.2.1 [16].

SNP array

SNP-array analysis on DNA from fresh-frozen tumor samples was performed using the Illumina HumanCNV370 or Illumina HumanOmni-Quad version 1.0 arrays (Illumina, San Diego, CA, USA), and on DNA from 4 µM FFPE sections scraped-off from slides using Affymetrix Oncoscan CNV arrays (Affymetrix, Santa Clara, CA, USA), as described [7, 11]; the platform used in individual cases is shown in Supplementary Table 1. Tumor Aberration Prediction Suite and Rawcopy were used, in combination with visual inspection, for segmentation of copy number shifts and for copy number evaluation [17, 18]. The position of the chromosomal imbalances was based on the GRCh37/hg19 sequence assembly.

RT-PCR

RT-PCR was performed to verify the novel fusion *CSMD1-MEAF6* identified in case 1. Reverse transcription and PCR amplification was performed using the following primers: CSMD1-Ex41F: 5′-TTGGACAATTTAGCGGCACG-3′ and MEAF6-Ex7R: 5′-CTGGTGCCAGCCAGTTTTG-3′. Amplified fragments were purified from agarose gels and Sanger sequenced using the Big Dye v1.1 cycle sequencing kit (Applied Biosystems, CA, USA) on an ABI-3130 genetic analyzer (Applied Biosystems).

WES

DNA from seven fresh-frozen tumor samples was used for WES and accompanying peripheral blood samples were available for three tumors. Whole exome capture, template preparation, and sequencing on the Ion Chef and Ion Proton systems, using the Ion PI IC 200 kit (Thermo Fisher Scientific, Waltham, MA, USA), were performed as described [19].

Variant calling for tumors with accompanying blood samples was performed with the AmpliSeq exome tumor-normal pair work-flow in Ion Reporter (Thermo Fisher Scientific), as well as with the Torrent Variant Caller plugin in the Torrent Suite Software (Thermo Fisher Scientific), using standard parameters for somatic variant detection. Annotation of detected variants was performed in Ion Reporter, using the GRCh37/hg19 assembly.

Only missense mutations with ≥20× coverage and ≥10% alternate allele frequency were retained. For tumors lacking corresponding peripheral blood, variants reported in dbSNP (v138) were removed and variants reported in cosmic (v67) were kept. All remaining variants were manually inspected using IGV.

From the four samples lacking constitutional DNA, also targeted sequencing of 50 neoplasia-associated genes was performed, using the Ion AmpliSeq Cancer Hot Spot Panel v2 (Thermo Fisher Scientific, Waltham, MA, USA) as described [20].

ATAC-seq

ATAC-seq was performed on biological duplicates or triplicates of the Bj5ta cell line with E–P, M–P, P–T, or EV constructs after 48 h of dox treatment. Libraries were prepared from 50,000–75,000 cells (per replicate) according to the Omni-ATAC protocol [20] with minor adjustments. In order to remove primer-dimers and large fragments, the final spin column purification step was replaced with double-sided (0.5× and 1.3×) AMPure XP bead purification. Libraries were sequenced on a NextSeq 500 (Illumina) with paired-end reads of 80 bp.

Remaining adapter sequences were removed from the FASTQ files using Trim-galore (v0.4.1). The trimmed reads were aligned to the human reference genome hg19 using BWA-MEM (v0.7.10) and duplicate reads were removed using Picard (v2.2.4). Bam files were filtered using SAMTools (v1.3). Reads from mitochondrial DNA and the Y chromosome were removed and only properly paired reads with high mapping quality (-q 30) were used for further analysis. An average of 108 million reads per sample was retained after filtering and the fragment-size distribution for the individual bam files was used for quality control (Supplementary Fig. 1). Coverage tracks for visualization were generated using deepTools

bamCoverage with parameters -bs = 1 -normalizeUsingRPKM. Peak calling was performed with Genrich using the parameter -j (ATAC-seq mode) and encode blacklisted regions were excluded. Peak files from all samples were combined and overlapping peaks were merged using bedtools. The number of reads per peak was counted separately for each sample using featureCounts and library size normalization was performed using DEseq2 [21, 22].

De novo motif discovery and annotation of peak regions was performed with HOMER [23].

Evaluation of RNA-seq and ATAC-seq data

Correlation-based principal component analysis (PCA) and hierarchical clustering analysis were performed using the Qlucore Omics Explorer version 3.3 (Qlucore AB, Lund, Sweden). FPKM and DEseq2 values were log2 transformed and the data were normalized to a mean of 0 and a variance of 1. Variables were filtered based on variance and the projection score was used to determine the optimal filtering threshold [24]. Hierarchical clustering of both samples and variables was performed using Euclidean distance and average linkage.

Two-sided *t* test, corrected for multiple testing by the Benjamini–Hochberg method, was used to identify statistically significant differences between groups. Fold change ≥ 2 and FDR ≤ 0.05 were used as cutoffs unless otherwise specified.

Gene set enrichment analysis (GSEA) was performed on gene-expression data using Qlucore Omics Explorer (v3.3). Signal-to-noise ratio was used as ranking metrics for analyzing curated gene sets (C2: KEGG and H3K27me3) acquired from the Molecular Signatures Database. Gene sets with <15 or >500 genes were excluded. Enriched gene sets after 1000 permutations at a false-discovery rate of ≤0.05 were considered significant.

Results

SNP array

All 37 OFMTs (18 typical, 9 atypical, 10 malignant) were analyzed with SNP arrays. Excluding copy number changes <100 Kb and copy-neutral LOH <1 Mb, a total of 30 cases showed imbalances (Supplementary Table 2; Fig. 1). Typical OFMTs had a median of 2 imbalances per case (range 0–7, mean 2.2), compared with 6 (range 2–12, mean 6.1) in atypical and 7 (range 0–16, mean 8.0) in malignant lesions. The most commonly involved chromosome band was 6p21, to which *PHF1* maps, showing deletions in 16/37 (43%) of the samples across all subtypes. Imbalances affecting other chromosome bands harboring genes involved in fusions were deletions in 12q24 (*EP400*) in 2/7 cases with *EP400–PHF1* fusions and duplications in both 16p13 (*CREBBP*) and Xq26 (*BCORL1*) in the single OFMT with a *CREBBP–BCORL1* fusion. Monosomy 13 or deletion of parts of 13q were seen only in atypical (4/9) and malignant (3/10) OFMTs, with a minimal shared region including three genes: *ITM2B*, *RB1*, and *RCBTB2*.

SNVs

WES achieved an average coverage of 82× in the targeted regions. From the three cases with corresponding peripheral blood, 8, 10, and 21 non-synonymous exonic SNVs were identified. For the four cases lacking peripheral blood, WES reported an average of 6 SNVs while no pathogenic mutations were detected at gene panel analysis. No gene was found to be recurrently mutated among the seven OFMT samples (Supplementary Table 3).

Gene fusion status in tumors

The gene fusion status was investigated in a total of 17 cases by either RNA-seq (nine cases) or a gene fusion panel (eight cases). Gene fusions were identified in all samples analyzed by RNA-seq and in 5/8 samples analyzed by the gene fusion panel (Supplementary Table 4). The most commonly identified gene fusion was *EP400–PHF1* found in seven cases followed by *PHF1–TFE3* in three. The remaining fusions *EPC1–PHF1, MEAF6–PHF1, CREBBP–BCORL1*, and *CSMD1–MEAF6* were found in one sample each. All fusions except *CSMD1–MEAF6* have been reported in OFMT before [25]. The novel *CSMD1–MEAF6* fusion was verified by RT-PCR, confirming that *CSMD1* exon 41 (NM_033225) was fused in-frame with *MEAF6* exon 7 (NM_022756; Supplementary Fig. 2).

Gene-expression profile of OFMT

The gene-expression profiles of seven OFMT were compared with those of 50 other soft tissue tumors. By unsupervised hierarchical clustering of the 1568 genes retained after variance filtration, OFMTs formed a distinct expression cluster (Fig. 2), with 845 coding genes being significantly differentially expressed between OFMT and the other tumor types (FDR < 0.1; Supplementary Table 5). The vast majority (95%) of the differentially expressed genes were upregulated in OFMTs.

Gene-expression profile and ATAC-seq findings in cell lines

Extensive differences in gene expression were observed between the cells expressing fusions and those with an

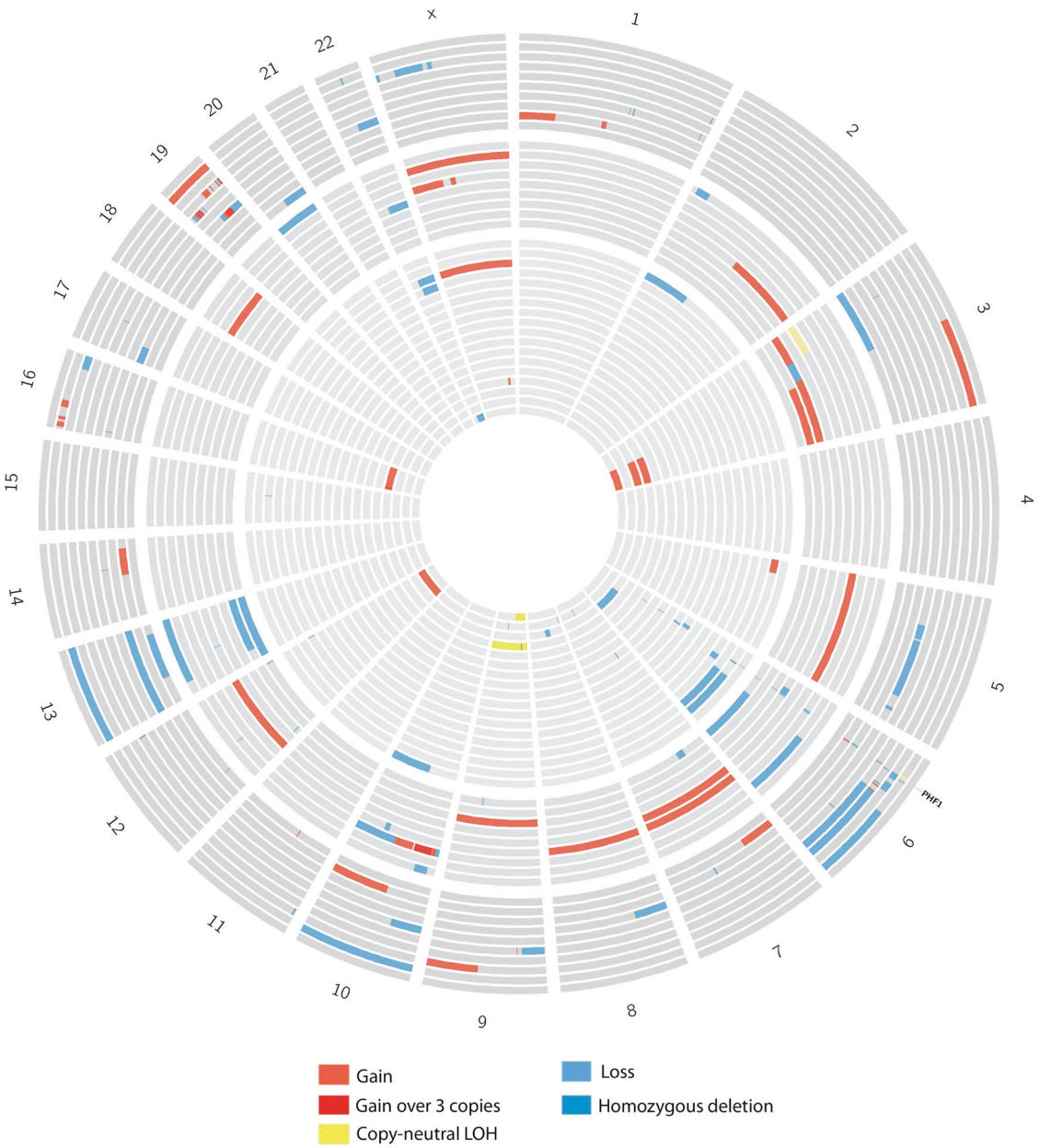

Fig. 1 Circos plot showing SNP-array results in OFMT. Samples are ordered by case number with the inner most circle representing case 1 and outer most circle representing case 37. Blue = loss of one copy; dark blue = homozygous loss; red = gain of one copy; dark red = gain of two or more copies; yellow = copy-neutral loss of heterozygosity.

empty vector. In agreement with gene-expression data from tumors, the majority of the genes found to be differentially expressed between cells with EV or a fusion construct were upregulated (mean 81%). Among the 794 genes found to be upregulated in OFMTs, 33% were upregulated also by all three fusion constructs. In addition, by GSEA, several gene sets were found to be enriched in both tumors and cell lines, including the Wnt, Hedgehog, and calcium signaling pathways (Supplementary Table 6). However, the most strongly enriched gene set was defined as genes with high-CpG-density promoters bearing histone mark H3K27me3 in neural progenitor cells [26]. Many of the top upregulated genes in OFMTs, including *GCGR, SFRP5, WNT7B,* and *KREMEN2*, were highly expressed also in the cell lines, as well as in the three FFPE OFMT samples (Supplementary Fig. 3). IHC for H3K27me3 and CTNNB1 in five OFMTs showed retained expression of the former and no nuclear staining for the latter.

The gene-expression profiles in the cell lines were also investigated by unsupervised PCA by which variation between the different fusion constructs was observed. While E–P and M–P seemed to have similar effects on the cells,

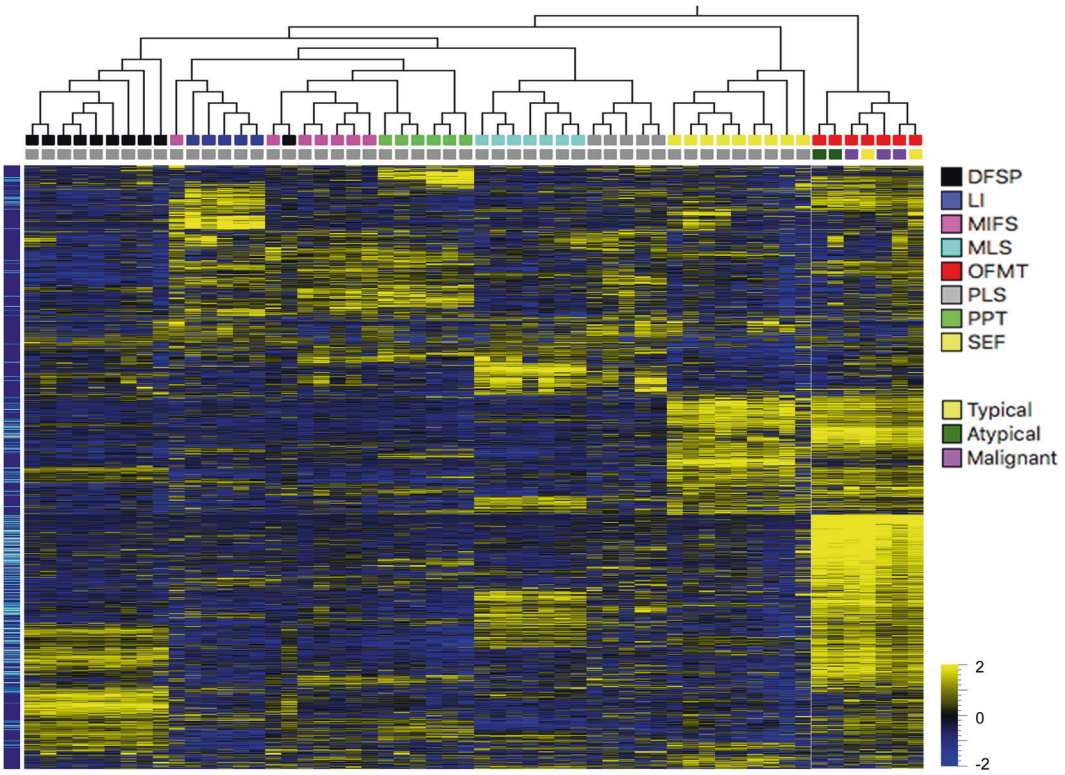

Fig. 2 OFMT forms a distinct expression cluster by unsupervised hierarchical clustering of 1568 genes. The upper panel of colored boxes below the dendrogram indicate the tumor type of each sample and the lower colored boxes indicate OFMT classification; DFSP dermatofibrosarcoma protuberans; IL inflammatory leiomyosarcoma; MLS myxoid liposarcomas; MIFS myxoinflammatory fibroblastic sarcoma; PLS pleomorphic liposarcoma; PRT PRDM10-rearanged tumors; SEF sclerosing epithelioid fibrosarcoma. Turquoise color in the dark blue panel to the left highlights genes found to be upregulated by all three fusion constructs.

P–T cells formed a distinct group and resulted in more extensive changes in gene expression (Supplementary Fig. 4A). These results were in line with the ATAC-seq data; unsupervised PCA of differentially accessible peaks demonstrated that the M–P and E–P cells clustered together while P–T formed a separate group and the largest variation was observed between EV and P–T (Supplementary Fig. 4B). Although there was considerable overlap among the different fusion constructs with regard to upregulated genes and differentially accessible peaks in promoter regions (Fig. 3a, b), the effect on these shared targets was stronger in P–T cells (Fig. 3c, d).

A comparison of the peaks found to be differentially open in the fusion-expressing cells compared with those in EV using de novo motif discovery identified several enriched transcription factor (TF) binding motifs. Enriched TF-motifs shared among the fusion-expressing cells included JunB, Egr2, and BORIS (Supplementary Table 7).

Discussion

The central role played by gene fusions in the development of OFMT was confirmed in the present study. Fusions were found in 14 of 17 cases analyzed by RNA-seq or a gene fusion panel, a frequency that is similar to the 85% reported before using various methodologies [5, 6]. However, there is reason to expect that the true frequency is even higher; the starting material (RNA extracted from archival FFPE slides that were mounted on slides) was clearly suboptimal for fusion panel analysis, not only limiting the number of cases that could be analyzed but also possibly creating false negative results; two of the three cases that were negative for gene fusions showed deletions in 6p21, where *PHF1* is located, on SNP-array analysis. Indeed, 6p21, which harbors the *PHF1* gene, was the most commonly involved chromosome band in chromosomal imbalances, showing deletions in 16 cases. *PHF1* is transcribed in the opposite direction in relation to its fusion partners (*EP400*, *EPC1*, *MEAF6*, and *TFE3*), meaning that in-frame fusions with these genes cannot be achieved through simple translocations. Hence, the frequent finding of copy number shifts in 6p21 could reflect that the *PHF1* fusions require more than two double strand breaks. Furthermore, 6p21 deletions were roughly as common in benign OFMTs (7/18) as in atypical/malignant cases (9/19), suggesting that the deletions are not progression related.

The spectrum of gene fusions in OFMT was further expanded by the identification of a novel *CSMD1–MEAF6*

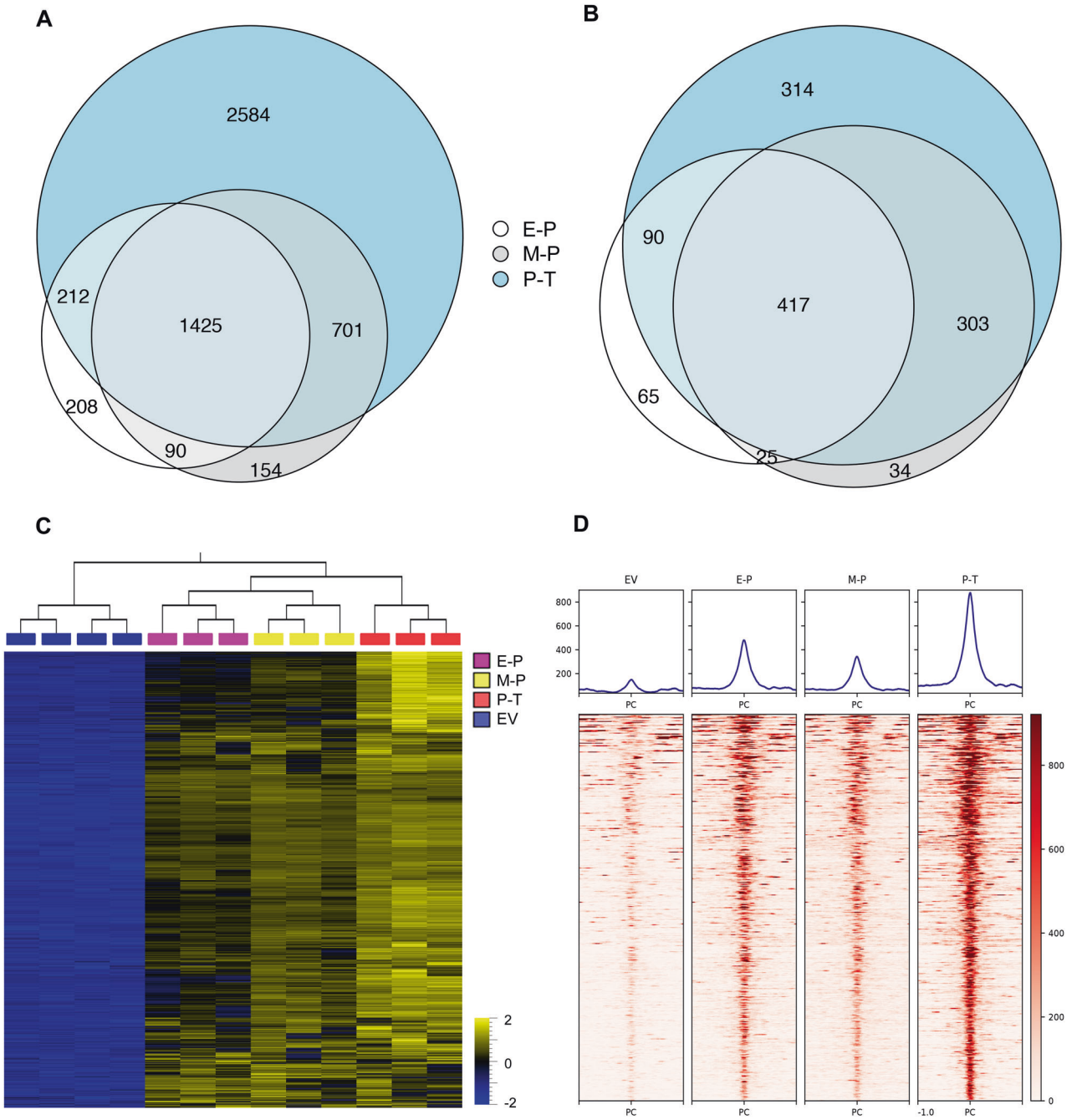

Fig. 3 Analysis of overexpressed genes and differentially open promoter regions in the TERT-immortalized Bj5ta human fibroblast cell line with the *EP400–PPHF1* (E–P), *MEAF6–PHF1* (M–P), *PHF1–TFE3* (P–T), or empty vector (EV) plasmids. a Venn diagram of upregulated genes in E–P, M–P, and T–P compared with EV. **b** Venn diagram of differentially accessible promoter regions E–P, M–P, and T–P compared with EV. **c** Heatmap of the 1425 upregulated genes found to be shared between E–P, M–P, and P–T. **d** visualization of coverage (RPKM) in the 417 differentially open promoter regions shared between E–P, M–P, and P–T. Plots display a region of ±1 kb from the peak centers (PC). The heatmap shows the coverage around shared open regions and the line plot shows the mean coverage around the same regions.

fusion. *MEAF6* encodes a member of the acetyl transferase complexes and has previously been reported as the 5′partner in fusions together with *PHF1*. *CSMD1* encodes a membrane-bound complement inhibitor [27], which has been suggested to act as a tumor suppressor gene in various malignancies [28]. The remaining gene fusions were in line with previous data, with *EP400–PHF1* being the most common variant, found in all types of OFMT. The second most common fusion was *PHF1–TFE3*, detected in two typical and one malignant OFMT. Thus, we could not

corroborate the suggestion by Suurmeijer et al. that the *PHF1–TFE3* fusion is associated with malignant OFMT [6]. This notwithstanding, it was of interest to note that the P–T construct had the strongest effect on the cells, resulting in a higher number of differentially expressed genes and differentially accessible regions. In addition, although the three different fusions shared several targets with regard to upregulated genes and differentially accessible promoter regions, the effect on these targets was stronger for the *PHF1–TFE3* fusion.

As our data suggest that all types of OFMT share a similar spectrum of gene fusions, atypical and malignant OFMTs likely represent progression-related variants of typical OFMTs, rather than distinct entities. The SNP-array data support this interpretation. Genomic imbalances were more common among atypical/malignant lesions than among typical OFMTs, and some specific genomic imbalances were more common among the former lesions (Supplementary Table 2 and Fig. 1). Notably, losses involving 13q were exclusively found (in 7/19 cases) in atypical/malignant OFMTs, and the minimal shared region of loss included only three genes, one of which was the tumor suppressor *RB1*. Thus, it seems likely that at least some of the morphological differences between typical and atypical/malignant OFMTs are due to clonal evolution. By contrast, SNVs and indels seem to have little impact on OFMT development or progression, as no recurrent variants were detected.

Despite the fact that five different fusions were found among the seven OFMTs available for gene-expression analysis, they formed a uniform group, easily distinguishable from the seven other soft tissue sarcoma subtypes by unsupervised hierarchical clustering. The expression pattern of OFMT was characterized by global upregulation, with 95% of the differentially expressed genes being expressed at higher levels than in the other sarcoma subtypes. Similar observations were made in cell lines expressing the E–P, M–P, and P–T fusion constructs; when compared with Bj5ta fibroblasts harboring an empty vector all three fusion constructs resulted in a general upregulation of gene expression. Furthermore, many of the top upregulated genes in OFMT tumor samples overlapped with genes upregulated by all three *PHF1* fusion constructs in the cell lines, including *GCGR*, *WNT7B*, *KREMEN2*, and *SFRP5*. The impact of the gene fusions on the gene-expression profile was further supported by GSEA as several gene sets were found to be highly enriched in both tumors and cell lines. These included signaling pathways associated with differentiation and development, such as Wnt and Hedgehog, as well as polycomb repressor complex 2 (PRC2) target genes, i.e., genes associated with the H3K27me3 mark.

Bearing in mind that OFMT is one of few soft tissue tumors frequently displaying bone formation, the high expression of the Wnt signaling pathway in tumors and Bj5ta cells is intriguing. Wnt signaling in general is well known to be critical for osteogenic differentiation, as shown in both human diseases and stem cells and in experimental animal models [29, 30]. In tumor tissues, several Wnt genes (*WNT3A*, *WNT4*, *WNT6*, *WNT7A*, *WNT7B*, *WNT10A*, and *WNT11*) were significantly upregulated in OFMTs. The highest fold change was seen for *WNT7B*, the induced expression of which has been shown to increase the number of osteoblasts and stimulate bone formation in mice, presumably through activation of the mammalian target of rapamycin complex 1 (mTORC1) rather than by activating β-catenin [31, 32]. In line with these findings, we here showed by IHC that OFMTs do not display nuclear expression of CTNNB1. However, known targets of mTORC1 signaling, such as EIF4EBP1, RPS6KB1, and RPS6KB2, were not among the differentially expressed genes in OFMTs (Supplementary Table 5). Furthermore, known antagonists of canonical Wnt signaling, including *KREMEN2* and *SFRP5*, were also highly upregulated in OFMTs; whether this is a physiological response to the upregulation of Wnt genes remains to elucidated.

The PRC2 target genes are also of particular interest in OFMT. PRC2 is a complex of proteins with histone methyltransferase activity and mainly mediates transcriptional repression through trimethylation of histone H3 at lysine 27 (H3K27me3) [33, 34]. PRC2 has essential roles in the maintenance of cell identity, differentiation, and proliferation, and its components appear to have complex roles in tumorigenesis, with both gain-of-function and loss-of-function mutations occurring in several types of cancer [35]. The *PHF1* gene, which is involved in the majority of OFMT-associated fusions, encodes plant homeodomain finger protein 1, a 456 aa protein involved in the repair of DNA double strand breaks, as well as in transcriptional regulation [34, 36, 37]. It is an accessory component of PRC2, and contains a Tudor domain (aa 28–87, encoded by exons 2–4) that specifically binds trimethylated lysine 36 on histone H3 (H3K36me3). It preferentially binds to partially unwrapped nucleosomes, thereby facilitating transcription through enhanced nucleosome accessibility, by decreasing the dissociation rate of DNA-binding proteins, and by protecting chromatin from transcriptional repression by the PRC2 complex [38–41]. However, not only the Tudor domain and its interaction with PRC2, but also the two PHD fingers and interactions with a variety of methylation marks and proteins are important for the epigenetic impact of PHF1 [37]. In all reported gene fusions involving *PHF1*, irrespective of whether it occurs as the 5′- or 3′-partner, the Tudor domain, as well as the zinc finger domains encoded by exons 4–8, are included. Furthermore, almost all of them will contain the entire amino-terminal domain (aa 2–28), which has been shown to strongly enhance the ability of the

Tudor domain to increase DNA accessibility in nucleosomes [41].

Driver mutations affecting the function of the PRC2 complex have previously been reported in other sarcomas, including malignant peripheral nerve sheath tumor (MPNST) and endometrial stromal sarcoma (ESS). MPNST harbors recurrent loss-of-function alterations in the PRC2 core components SUZ12 and EED. These tumors display global loss of the H3K27me3 mark, resulting in a general upregulation of gene expression when compared with MPNST with wt *SUZ12* and *EED* [42, 43]. ESS, like OFMT, harbors chromosomal translocations affecting proteins associated with PRC2, including SUZ12, PHF1, and EPC1 [25, 44]. The fusion proteins in ESS have been suggested to initiate activation of the Wnt signaling pathway by impairment of PRC2 [45, 46]. Similar to our findings in OFMT, gene-expression analysis of ESS showed that 88% of differentially expressed genes were upregulated, and there was strong enrichment of genes associated with the Wnt signaling pathway among the upregulated genes [46]. Most likely, the fusion of PHF1 with ectopic sequences will give rise to a chimeric protein that cannot properly function in one or more of the many protein–protein interactions in which wt PHF1 participates [37]. Our results strongly suggest that the PHF1 fusions deregulate gene expression through alterations in chromatin accessibility. However, this effect is unlikely to be achieved through blocking of PRC2, as both we and others [43] have found retained H3K27me3 staining by IHC in OFMTs. Thus, further studies are needed to understand how the chimeric proteins affect chromatin regulation.

Acknowledgements The kits for the TruSight Fusion RNA Fusion Panel library preparation were kindly provided by Illumina, San Diego, USA.

Funding This study was supported by the Swedish Cancer Society and Region Skåne.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Folpe AL, Weiss SW. Ossifying fibromyxoid tumor of soft parts. A clinicopathologic study of 70 cases with emphasis on atypical and malignant variants. Am J Surg Pathol. 2003;27:421–31.
2. Miettinen M, Finnell V, Fetsch JF. Ossifying fibromyxoid tumor of soft parts – a clinicopathologic and immunohistochemical study of 104 cases with long-term follow-up and a critical review of the literature. Am J Surg Pathol. 2008;32:996–1005.
3. Enzinger FM, Weiss SW, Liang CY. Ossifying fibromyxoid tumor of soft parts. A clinicopathological analysis of 59 cases. Am J Surg Pathol. 1989;13:817–27.
4. Graham RP, Dry S, Li X, Binder S, Bahrami A, Raimondi SC, et al. Ossifying fibromyxoid tumor of soft parts: a clinicopathologic, proteomic, and genomic study. Am J Surg Pathol. 2011;35:1615–25.
5. Mertens F, Antonescu CR, Mitelman F. Gene fusions in soft tissue tumors: recurrent and overlapping pathogenetic themes. Genes Chromosom Cancer. 2016;55:291–310.
6. Suurmeijer AJH, Song W, Sung YS, Zhang L, Swanson D, Fletcher CDM, et al. Novel recurrent PHF1-TFE3 fusions in ossifying fibromyxoid tumors. Genes Chromosom Cancer. 2019;58:643–9.
7. Gebre-Medhin S, Nord KH, Möller E, Mandahl N, Magnusson L, Nilsson J, et al. Recurrent rearrangement of the PHF1 gene in ossifying fibromyxoid tumors. Am J Pathol. 2012;181:1069–77.
8. Antonescu CR, Sung Y-S, Chen C-L, Zhang L, Chen H-W, Singer S, et al. Novel ZC3H7B-BCOR, MEAF6-PHF1, and EPC1-PHF1 fusions in ossifying fibromyxoid tumors - molecular characterization shows genetic overlap with endometrial stromal sarcoma. Genes Chromosom Cancer. 2014;53:183–93.
9. Kao Y-C, Sung Y-S, Zhang L, Chen C-L, Huang S-C, Antonescu CR. Expanding the molecular signature of ossifying fibromyxoid tumors with two novel gene fusions: CREBBP-BCORL1 and KDM2A-WWTR1. Genes Chromosom Cancer. 2017;56:42–50.
10. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F (editors). WHO classification of tumours of soft tissue and bone. Lyon: IARC; 2013. p. 468.
11. Arbajian E, Puls F, Antonescu CR, Amary F, Sciot R, Debiec-Rychter M, et al. In-depth genetic analysis of sclerosing epithelioid fibrosarcoma reveals recurrent genomic alterations and potential treatment targets. Clin Cancer Res. 2017;23:7426–34.
12. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics. 2011;27:2903–4.
13. Nicorici D, Satalan M, Edgren H, Kangaspeska S, Murumägi A, Kallioniemi O, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. 2014. https://doi.org/10.1101/011650.
14. Haas B, Dobin A, Stransky N, Li B, Yang X, Tickle T, et al. STAR-Fusion: fast and accurate fusion transcript detection from RNA-Seq. 2017. https://doi.org/10.1101/120295.
15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
16. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
17. Rasmussen M, Sundström M, Kultima H, Botling J, Micke P, Birgisson H, et al. Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. Genome Biol. 2011;12:R108.
18. Mayrhofer M, Viklund B, Isaksson A. Rawcopy: improved copy number analysis with affymetrix arrays. Sci Rep. 2016;6:36158.
19. Hofvander J, Jo VY, Ghanei I, Gisselsson D, Mårtensson E, Mertens F. Comprehensive genetic analysis of a paediatric pleomorphic myxoid liposarcoma reveals near-haploidization and loss of the RB1 gene. Histopathology. 2016;69:141–7.
20. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017;14:959–62.
21. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
23. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.
24. Fontes M, Soneson C. The projection score – an evaluation criterion for variable subset selection in PCA visualization. BMC Bioinform. 2011;12:307.
25. Mitelman F, Johansson B, Mertens F. Mitelman database of chromosome aberrations and gene fusions in cancer. 2019. https://mitelmandatabase.isb-cgc.org/.
26. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.
27. Escudero-Esparza A, Kalchishkova N, Kurbasic E, Jiang WG, Blom AM. The novel complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex assembly. FASEB J. 2013;12:5083–93.
28. Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, et al. Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. Cancer Biol Ther. 2009;10:907–16.
29. Liu G, Vijayakumar S, Grumulato L, Arroyave R, Qiao H, Akiri G, et al. Canonical Wnts function as potent regulators of osteogenesis by human mesenchymal stem cells. J Cell Biol. 2009;185:67–75.
30. Houschyar KS, Tapking C, Borrelli MR, Popp D, Duscher D, Maan ZN, et al. Wnt pathway in bone repair and regeneration – what do we know so far? Front Cell Dev Biol. 2019;6:170.
31. Chen J, Tu X, Esen E, Joeng KS, Lin C, Arbeit JM, et al. WNT7B promotes bone formation in part through mTORC1. PLOS Genet. 2014;10:e1004145.
32. Sun J, Ermann J, Niu N, Yan G, Yang Y, Shi Y, et al. Histone demethylase LSD1 regulates bone mass by controlling WNT7B and BMP2 signaling in osteoblasts. Bone Res. 2018;6:14.
33. Margueron R, Reinberg D. The polycomb complex PRC2 and its mark in life. Nature. 2011;469:343–9.
34. Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol Cell Biol. 2008;28:2718–31.
35. Comet I, Riising EM, Leblanc B, Helin K. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nat Rev Cancer. 2016;16:803–10.
36. Cao R, Wang H, He J, Erdjument-Bromage H, Tempst P, Zhang Y. Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol Cell Biol. 2008;28:1862–72.
37. Liu R, Gao J, Yang Y, Qiu R, Zheng Y, Huang W, et al. PHD finger protein (PHF1) is a novel reader for H4R3 symmetric dimethylation and coordinates with PRMT5-WDR77/CRL4B complex to promote tumorigenesis. Nucl Acids Res. 2018;46:6608–26.
38. Musselman CA, Gibson MD, Hartwick EW, North JA, Gatchalian J, Poirier MG, et al. Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. Nat Commun. 2013;4:2969.
39. Lu R, Wang GG. Tudor: a versatile family of histone methylation ”readers”. Trends Biochem Sci. 2013;38:546–55.
40. Choi J, Bachmann AL, Tauscher K, Benda C, Beat Fierz B, Müller J. DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat Struct Mol Biol. 2017;24:1039–47.
41. Gibson MD, Gatchalian J, Slater A, Kutateladze TG, Poirier MG. PHF1 Tudor and n-terminal domains synergistically target partially unwrapped nucleosomes to increase DNA accessibility. Nucl Acids Res. 2017;45:3767–76.
42. Lee W, Teckie S, Wiesner T, Ran L, Prieto Granada CN, Lin M, et al. PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nat Genet. 2014;46:1227–32.
43. Prieto-Granada CN, Wiesner T, Messina JL, Jungbluth AA, Chi P, Antonescu CR. Loss of H3K27me3 expression is a highly sensitive marker for sporadic and radiation-induced MPNST. Am J Surg Pathol. 2016;40:479–89.
44. Panagopoulos I, Micci F, Thorsen J, Gorunova L, Eibak AM, Bjerkehagen B, et al. Novel fusion of MYST/Esa1-associated factor 6 and PHF1 in endometrial stromal sarcoma. PLoS One. 2012;7:e39354.
45. Ma X, Wang J, Wang J, Ma CX, Gao X, Patriub V, et al. The JAZF1-SUZ12 fusion protein disrupts PRC2 complexes and impairs chromatin repression during human endometrial stromal tumorogenesis. Oncotarget. 2017;8:4062–78.
46. Przybyl J, Kidzinski L, Hastie T, Debiec-Rychter M, Nusse R, van de Rijn M. Gene expression profiling of low-grade endometrial stromal sarcoma indicates fusion protein-mediated activation of the Wnt signaling pathway. Gynecol Oncol. 2018;149:388–93.