The molecular landscape of well differentiated retroperitoneal liposarcoma

Tyler, Robert; Dilworth, Mark P; James, Jonathan; Blakeway, Daniel; Stockton, Joanne D; Morton, Dion G; Taniere, Phillipe; Gourevitch, David; Desai, Anant; Beggs, Andrew D

DOI:
10.1002/path.5749

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Tyler, R, Dilworth, MP, James, J, Blakeway, D, Stockton, JD, Morton, DG, Taniere, P, Gourevitch, D, Desai, A & Beggs, AD 2021, 'The molecular landscape of well differentiated retroperitoneal liposarcoma', *Journal of Pathology*, vol. 255, no. 2, pp. 132-140. https://doi.org/10.1002/path.5749

Journal of Pathology
J Pathol October 2021; **255:** 132–140
Published online 4 August 2021 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.5749

ORIGINAL PAPER

The molecular landscape of well differentiated retroperitoneal liposarcoma

Robert Tyler[1], Mark P Dilworth[1], Jonathan James[1], Daniel Blakeway[1], Joanne D Stockton[1], Dion G Morton[1], Phillipe Taniere[2], David Gourevitch[2], Anant Desai[2] and Andrew D Beggs[1]*

[1] *Institute of Cancer and Genomic Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK*
[2] *Midland Abdominal Retroperitoneal Sarcoma Unit (MARSU), University Hospital Birmingham, Birmingham, UK*

**Correspondence to: AD Beggs, Institute of Cancer and Genomic Science, University of Birmingham, Vincent Drive, Birmingham, B15 2TT, UK. E-mail: a.beggs@bham.ac.uk*

Abstract

Well differentiated liposarcoma (WD-LPS) is a relatively rare tumour, with fewer than 50 cases occurring per year in the UK. These tumours are both chemotherapy- and radiotherapy-resistant and present a significant treatment challenge requiring radical surgery. Little is known of the molecular landscape of these tumours and no current targets for molecular therapy exist. We aimed to carry out a comprehensive molecular characterisation of WD-LPS via whole genome sequencing, RNA sequencing, and methylation array analysis. A recurrent mutation within exon 1 of *FOXD4L3* was observed (chr9:70,918,189A>T; c.322A>T; p.Lys108Ter). Recurrent mutations were also observed in Wnt signalling, immunity, DNA repair, and hypoxia-associated genes. Recurrent amplification of *HGMA2* was observed, although this was in fact part of a general amplification of the region around this gene. Recurrent gene fusions in *HGMA2*, *SDHA*, *TSPAN31*, and *MDM2* were also observed as well as consistent rearrangements between chromosome 6 and chromosome 12. Our study has demonstrated a recurrent mutation within *FOXD4L3*, which shows evidence of interaction with the PAX pathway to promote tumourigenesis.

Keywords: liposarcoma; whole genome sequencing; RNA sequencing; methylation

Received 22 May 2021; Revised 14 June 2021; Accepted 18 June 2021

No conflicts of interest were declared.

Introduction

Retroperitoneal liposarcoma (RPLS) is the most common type of retroperitoneal sarcoma (RPS), making up around half of all RPSs [1]. RPLS typically presents incidentally on cross-sectional imaging as a large homogeneous retroperitoneal mass [2]. As these tumours become larger, the risk of their transformation from well differentiated (WD) to dedifferentiated (DD) increases. The molecular events that initiate WD-LPS and lead to dedifferentiation of well differentiated liposarcoma are currently unknown. Molecular changes seen in the various types of RPLS include coordinated amplification of *MDM2* and *CDK4* [3]; giant ring chromosomes [4–6]; amplification of *CDK4*, *HMGA2*, and *TSPAN31*; and the *FUS–DDIT3* (solely in myxoid liposarcoma) fusion oncogene, seen in a subset of liposarcomas caused by t(12;16) or t(12;22) translocation [7,8]. To date, there have been few comprehensive studies on the molecular genetics of well differentiated RPLS [9]. The key feature is amplification of the *MDM2* gene, which has been seen almost ubiquitously in WD-RPLS and DD-RPLS at rates approaching 100% [10]. In an exome sequencing study, Kanojia *et al* [11] found consistent amplification, as previously described, in *MDM2* and *HMGA2*. The most frequently mutated gene in their cohort was in *PLEC* (27% of samples), which codes for plectin, a cytoskeleton protein. Amin-Mansour *et al* [12] studied exome sequencing trios of normal, well differentiated, and dedifferentiated tissue regions in patients with liposarcoma, finding no common driver but verifying the recurrent amplification of *CDK4*.

However, the lack of a comprehensive integrated analysis means that therapeutic options for treatment of this tumour are limited as the molecular pathways that are potentially dysregulated in liposarcoma have not been identified with certainty. We therefore carried out a comprehensive pilot multi-platform molecular characterisation, utilising whole genome sequencing, RNA sequencing, and methylation arrays of well differentiated retroperitoneal liposarcoma.

Materials and methods

Sample collection

Retroperitoneal sarcoma samples were obtained from patients undergoing laparotomy for *en bloc* excision of

a retroperitoneal mass. Immediately after the specimen had been removed from the patient, it was conveyed to the Queen Elizabeth Hospital Birmingham (QEHB) Histopathology laboratories, where a representative sample of liposarcoma and associated normal fat was taken as far away as possible from the tumour, and immediately flash frozen in liquid nitrogen and then stored at −80 °C. Where no normal fat was available, normal adjacent organ (i.e. kidney/colon/pancreas, etc.) was sampled. Where no frozen material was available to match with the specimen, formalin-fixed, paraffin-embedded (FFPE) tissue blocks were retrieved from the histopathology archives of the QEHB.

Ethical approval was obtained via the University of Birmingham Human Biomaterials Resource Centre Biobanking ethics (Ref 09/H1010/75). All experiments were performed in accordance with relevant guidelines and regulations, and informed consent was obtained from all participants.

DNA/RNA extraction

DNA extraction was performed by proteinase K digestion. A 5-mm^3 piece of tissue was immersed in buffer ATL overnight with 20 mg of proteinase K. The resulting lysate was processed using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Manchester, UK). Extracted DNA was eluted into buffer AL and was quantified using Nanodrop spectrophotometry for purity and Qubit (Thermo Fisher Scientific, Loughborough, UK) fluorimetry for DNA concentration.

Because of limited volumes of tissue, RNA extraction was performed on formalin-fixed, paraffin-embedded sections. Three 10 μm FFPE sections were cut using a microtome, then processed using the Purelink FFPE RNA extraction kit (Invitrogen, Renfrew, UK) according to the manufacturer's instructions. Extracted RNA was quantified for purity using Nanodrop spectrophotometry and for quantity using Qubit fluorimetry (Thermo Fisher Scientific, Loughborough, UK).

Whole genome sequencing (WGS) and targeted sequencing

Shotgun WGS was performed using 500 ng of paired tumour–normal DNA by the Illumina R&D group at the Illumina Great Chesterford facility. DNA was library-prepared for sequencing using the Truseq PCR free library preparation kit according to the manufacturer's guidelines. For six samples (associated normal), no fresh frozen material was available and therefore a PCR-based WGS library preparation was used (Illumina FFPE) using 1 μg of DNA. Sequencing was performed on an Illumina HiSeq 2500, pooling one sample across four lanes using the v4 chemistry. Sequenced reads were aligned to the GrCh37 reference genome using the Isaac aligner [13]. Tumour–normal single nucleotide variant (SNV) and indel calling was performed using the Strelka variant caller [14]. Copy number variant calling was performed in Canvas in tumour–normal mode, and where no control normal was available, a panel of normals was used. Structural variant calling was performed using Manta.

In order to resolve complex variants and paralogues, targeted nanopore sequencing of the *FOXD4L3* region was carried out. Custom primers were designed with Primer3 to target the entire coding sequence of FOXD4L3 in one amplicon. Long-range PCR (conditions available on request) was performed and PCR amplicons were cleaned and ligated using an Oxford Nanopore ligation sequencing kit (LSK-109; Oxford Nanopore, Oxford, UK). Library was then loaded onto an ONT MinIon 9.5.1 flow cell and run for 36 h. Data were basecalled using Guppy and variant calling was performed by MiniMap2. No polishing was performed.

Whole genome sequencing data for the well differentiated liposarcoma line 93T449 were downloaded from the NCBI SRA archive. Downloaded data were aligned to the GRCh38 reference genome using MiniMap2; then variant calling was performed with FreeBayes and copy number calling with Canvas as previously described.

RNA sequencing

Whole transcriptome sequencing using 1 μg of total RNA was performed by the Shenzen Research Laboratories of Beijing Genomics Inc using the Illumina TruSeq RNA Sample preparation kit v2. RNA was DNAase-digested, fragmented, and cDNA was synthesised using SuperScript III (Invitrogen). cDNA was then end-repaired, A-tailed, adapter-ligated, PCR-amplified, and purified. After library quantification, this was treated with duplex specific nuclease (Evrogen, Cambridge, UK) and purified. The final libraries were sequenced on an Illumina HiSeq 2000 to an average of 35 million reads with a read length of 90 bp.

Sequenced reads were quality-trimmed using Trimmomatic and aligned with the hg38 reference genome using STAR [15]. Absolute gene expression was determined using *featureCounts* [16] and differential expression was determined using the GSA module of Partek Flow [17,18]. Log normalised FPKM values were submitted to the CIBERSORT immune infiltrate pipeline [19] for deconvolution of immune cell types and were used to calculate the coordinate immune response cluster (CIRC) score [20].

RNA fusion analysis was carried out using the fusion analysis pipeline of STAR-FUSION [21].

DNA methylation arrays

Analysis of the methylome was achieved using Illumina HumanMethylation450 arrays at the University of Birmingham. One microgram of DNA was bisulphite-converted using the Zymo EZ-DNA methylation kit using the custom Illumina microarray protocol. This was then hybridised to Illumina HumanMethylation450 microarrays according to the manufacturer's protocol. Hybridised arrays were washed and stained, and then immediately scanned using an Illumina iScan microarray scanner. Extracted microarray intensities were normalised using

the *Bioconductor/ChAMP* pipeline [22] and individual beta values were exported. Differential methylation was quantified using *eBayes/limma* and differentially methylated regions were ascertained using *DMRhunter*.

Further bioinformatics analysis

Identification of recurrent somatic mutations was carried out as follows. Firstly, exported subtracted tumour–normal variant calls from Strelka were transformed from Variant Call Format (VCF) files to MAF (Mutation Annotation Format) files using vcf2maf. In patients where no matched normal was available, tumour VCF files were filtered using dbSNP 1000 Genomes Project and GnoMAD such that all SNPs with a minor allele frequency of greater than 0.01 were discarded. MAF files were then concatenated together in order for analysis in MutSigCV [23]. Analysis of recurrent mutations in the cohort was carried out using MutSigCV using standard settings. All recurrent mutations were ranked and sorted by *P* value.

Identification of recurrent structural variations was carried out using the Manta [24] structural variant caller (https://github.com/Illumina/manta). The package was run with standard parameters on tumour BAM files. Subsequent output was then filtered to produce Circos (RCircos) plots by taking a subset of the data from Manta and limiting it to interchromosomal inversions and translocations.

Copy number alterations were called using the Illumina Canvas [25] (https://github.com/Illumina/canvas) copy number variant caller. The package was run with standard parameters on tumour BAM files. Regions were output as standard VCF files and then filtered using a custom script to preserve high-quality copy number calls (quality score > 30) on regions greater than 5 kilobases in length (as this was thought to be more functionally relevant). Copy number calls were then imported into GREVE [26] (http://www.well.ox.ac.uk/GREVE/) with genome-wide plots and chromosome level significance plots output.

Validation of findings

Sanger sequencing to validate mutational findings was carried out using standard methodology. Primers flanking the region of interest were designed using ExonPrimer with a T_m of 60 °C. Primer specificity was verified using UCSC *in silico* PCR and primers were optimised with gradient PCR before use. 10 ng of tumour DNA underwent PCR using the Qiagen Multiplex kit using primers against *FOXD4L3* (supplementary material, Table S1). PCR products were cleaned with ExoSAP-IT and underwent a BigDye sequencing reaction under standard conditions. Sequencing products were run on an ABI 3700 sequencer.

Nanopore sequencing was carried out via long-range PCR of the entire *FOXD4L3* gene in one amplicon. A master mix was prepared containing template DNA, 2 µl of forward and reverse primers at 0.4 µM, 25 µl of LongAmp *Taq* 2X Master Mix (New England Biolabs, Ipswich, MA, USA), and nuclease-free water up to a total of 50 µl. Thirty cycles of long-range PCR (with a 2.5 min extension time at 65 °C) were performed and the product underwent repair and end prep using the Oxford Nanopore LSK-109 sequencing protocol. Samples were indexed with the Rapid Barcoding Kit and then loaded and sequenced on an R9.4.1 MinIon flow cell for 24 h. Raw sequence FAST5 files were demultiplexed and converted to FASTQ using the Guppy Basecaller. Called FASTQ files underwent two rounds of polishing with Canu, followed by a round of polishing with Medaka. Sequence data were aligned using minimap2 to the GRCh38 genome followed by variant calling using FreeBayes (command line `freebayes -C 2 -0 -O -q 20 -z 0.10 -E 0 -X -u -p 2 -F 0.6`). Raw allele counts at the desired sites were output with the samtools command.

Results

A total of eight WD-LPS samples underwent whole genome sequencing of three tumour–normal pairs and five tumour-only samples (due to lack of available paired normal tissue) to an average of 85x read depth, RNA transcriptomic sequencing of tumours only (average 35 million reads per sample), and analysis on the HumanMethylation450 array system. The characteristics of the patients from which tumour samples were obtained are shown in Table 1. Six out of eight patients had died at the time of publication: two from recurrence of a dedifferentiated liposarcoma and four from other causes.

Whole genome sequencing reveals recurrent mutations in *FOXD4L3*

The mutational rate in all tumours was between 1.65 and 1.8 mutations per megabase, making WD-LPS a relatively infrequently mutated tumour to others studied in the TCGA (supplementary material, Tables S2 and S3). Analysis of recurrent mutations using MutSigCV2 within the samples demonstrated significant mutations (Table 2 and Figure 1) in *SPRN*, *FOXD4L3*, *ADH1C*, and *KRTAP2-2*. Pathway analysis of these gene mutations determined that only *FOXD4L3* was likely to be relevant in a cancer context, due to it being a Forkhead transcription factor-associated gene. *FOXD4L3* was predicted by the STRINGS database (supplementary material, Table S4) to interact with *PAX3*, *PAX7*, *T*, and *TBX19* transcription factors, all involved in cell differentiation and development, as well as being implicated in alveolar rhabdomyosarcoma [27], biphenotypic sinonasal sarcoma, and others. We also analysed recurrently mutated genes within the Cancer Gene Census and those previously reported in liposarcoma [11] using the Illumina Encore pipeline (supplementary material, Figure S1) including immune evasion (*HLA-A*, 5/8 samples), Wnt signalling (*FAT1*;4/8 samples), double strand

Table 1. Table of patient characteristics.

Patient ID	Sample ID	Type	Gender	Age (years)	Tumour location	Tumour histological type	Patient alive?
P000111	S000532	Tumour	Male	59	Retroperitoneum	Well differentiated liposarcoma	Alive
P000214	S000565	Tumour	Male	76	Retroperitoneum	Well differentiated liposarcoma	Deceased
P001190	S038759	Tumour	Female	63	Retroperitoneum	Well differentiated liposarcoma	Alive
P001193	S038757	Tumour	Female	81	Retroperitoneum	Well differentiated liposarcoma	Deceased
	S038758	Normal	Female	81	Retroperitoneum	Normal fat	Deceased
P001207	S010732	Tumour	Female	60	Retroperitoneum	Well differentiated liposarcoma	Deceased
	S044139	Normal	Female	60	Retroperitoneum	Normal fat	Deceased
P001212	S038764	Tumour	Male	58	Retroperitoneum	Well differentiated liposarcoma	Alive
P001224	S010732	Tumour	Female	63	Retroperitoneum	Well differentiated liposarcoma	Deceased
	S038770	Normal	Female	63	Retroperitoneum	Normal fat	Deceased
P001216	S038761	Tumour	Male	79	Retroperitoneum	Well differentiated liposarcoma	Deceased

Table 2. List of mutations ranked by significance.

Rank	Gene	Gene name	*P* value	*q* value	Significant
1	*SPRN*	Shadow of prion protein homolog	8.21E−09	1.55E−04	Yes
2	*ADH1C*	Alcohol dehydrogenase 1C	1.52E−07	1.43E−03	Yes
3	*FOXD4L3*	Forkhead box D3 like 3	2.57E−07	1.62E−03	Yes
4	*KRTAP2-2*	Keratin-associated protein 2-2	2.04E−06	9.63E−03	Yes
5	*LILRB3*	Leukocyte immunoglobulin-like receptor, subfamily B	3.37E−06	1.27E−02	No
6	*GYPE*	Glycophorin E	8.26E−05	2.60E−01	No
7	*GOLGA8B*	Golgin A8 family, member B	1.22E−04	3.28E−01	No
8	*OR2A42*	Olfactory receptor family 2 subfamily A member 42	2.05E−04	4.82E−01	No
9	*IL7R*	Interleukin-7 receptor	2.52E−04	4.82E−01	No
10	*C12orf77*	Chromosome 12 open reading frame 77	2.68E−04	4.82E−01	No

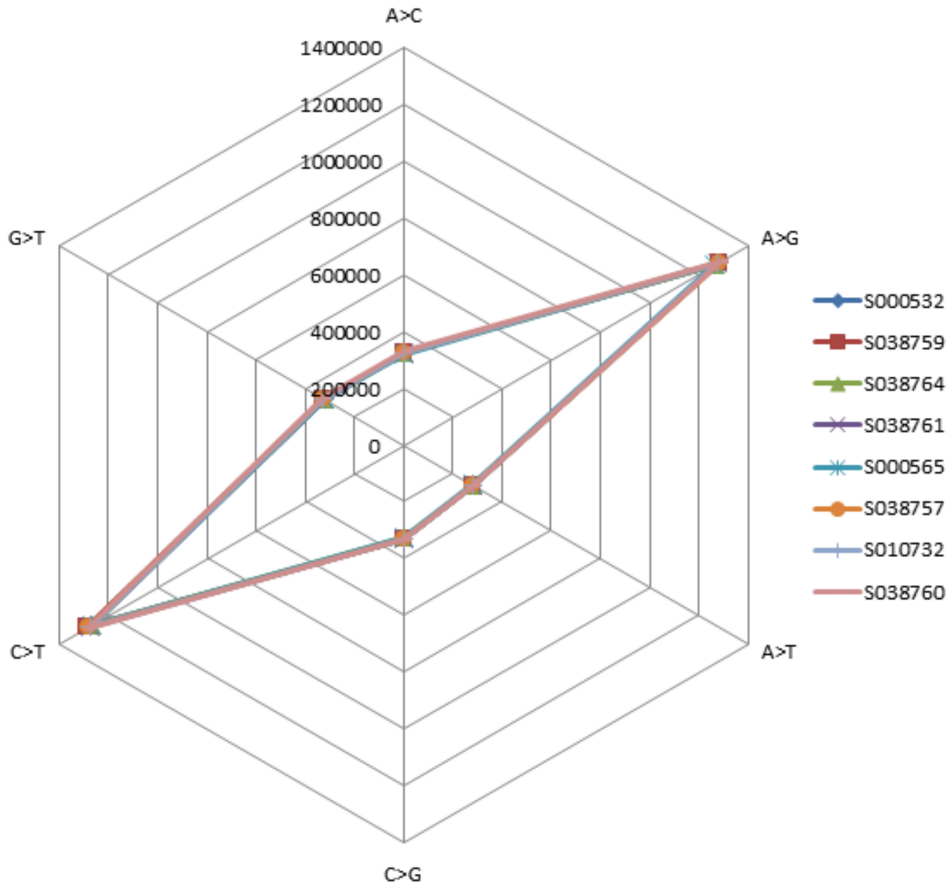

Figure 1. Spider chart of mutation type in whole genome sequencing data.

break repair (*ATM*, 2/8 samples; *BRCA1*, 2/8 samples), PAX transcription factor (*PAX5*, 2/8 samples), and hypoxia tumour suppressors (*SDHA*, 2/8 samples).

We therefore studied *FOXD4L3* in more detail and found a recurrent mutation within exon 1 of *FOXD4L3* (chr9:70,918,189A>T; c.322A>T; p.Lys108Ter) which resulted in a stop gain. This mutation was present in 7/8 samples and was verified by bidirectional Sanger sequencing of tumour samples; however, this mutation was also present in two normal samples (patients P001224 and P001207) but not in a further normal sample (P001193). The mutation seen in the normal tissue as determined by NGS was not seen on Sanger sequencing. Further investigation revealed that the 'normal' tissue was in fact 'normal' fat immediately adjacent to the tumour and probably represented inadvertent sampling of the leading edge of the WD-LPS sample. The mutation was just upstream of the Forkhead box domain of the protein (supplementary material, Figure S2), suggesting that this mutation may have functional consequences due to loss of this protein binding domain and is expressed in tissue (supplementary material, Figure S3).

Because of the incongruity between the whole genome sequencing and the Sanger sequencing results, we decided to carry out further investigation. *FOXD4L3* is a paralogue of *FOXD4* [which has multiple (six) paralogues], with minimal sequence variation between them. We therefore designed long-range PCR primers to flank the entirety of *FOXD4L3* and performed long-range nanopore amplicon sequencing on a cohort of 24 FFPE samples (consisting of 12 tumour samples plus 12 adjacent normal tissues) plus the 3TL3 liposarcoma cell line. All samples had a mutation within *FOXD4L3*. We found that the variant allele frequency (VAF) of the A>T transversion was significantly different between tumour and normal tissue, with the tumour having a median VAF of the T allele of 0.43 [interquartile range (IQR) 0.39–0.45] and the normal tissue having a median VAF of 0.21 (IQR 0.16–0.22, Wilcoxon rank $p < 0.001$). Investigation of the 93T449 well differentiated liposarcoma cell line revealed a copy number (CN) = 3 of the regions surrounding the *FOXD4L3* gene which was replicated in the whole genome-sequenced samples, suggesting a complex mutation and copy number event within *FOXD4L3* giving the differences in allele frequency.

Copy number analysis of WGS confirms recurrent gains in *MDM2* as part of a larger amplification

Copy number variant analysis of the tumour set (Table 3) demonstrated the recurrent copy number gain in *HGMA2*, the binding partner of *MDM2* [28], in six out of eight cases, with all cases having amplification within the 12q15 region to some extent. Recurrent amplification was seen (Figure 2) throughout the long arm of chr12q13.13–q24.33, demonstrating that it is not necessarily restricted to the *MDM2/HGMA2* amplification previously described. Analysis of structural variation (SV) revealed a complex pattern (supplementary material, Figure S4) of small-scale chromosomal translocation across the tumour types. Rearrangement of both between and within chromosomes 6 and 12 was seen (supplementary material, Figure S5), consistent which previous cytogenetic observation of WD-LPS.

RNA sequencing demonstrates changes in fat-associated genes and downregulation of Wnt and PI3K signalling

Differential expression analysis between liposarcoma and normal fat was carried out in Partek flow (supplementary material, Table S5). The most differentially expressed gene was *TNS1* (tensin 1, normalised counts 20.54 versus 110.9, $p = 2 \times 10^{-6}$, $Q = 0.03$), which is part of the PTEN signalling pathway. Pathway analysis, however, demonstrated dysregulated pathways in 'Pathways in Cancer' (hsa05200, enrichment score (ES) = 11.61, $p = 9.1 \times 10^{-6}$), the MAPK signalling pathway (hsa04010, ES = 6.49, $p = 1.52 \times 10^{-3}$), and PI3K–Akt signalling (hs04151, ES = 5.26, $p = 5.21 \times 10^{-3}$).

Analysis of alternative splicing was carried out using STAR (supplementary material, Table S6), with *FN1* (fibronectin 1) being the top ranked alternatively spliced gene (normalised counts tumour = 10.28, normal = 0.41; $p = 4.52 \times 10^{-9}$, $Q = 8.25 \times 10^{-6}$). Pathway analysis demonstrated that the focal adhesion pathway was

Table 3. Recurrent copy number events.

Chr	Type	Start	End	Size (bp)	Position (hg19 coordinates)	Samples present in	Score	Genome-wide *P* value	Chromosome-wise *P* value	Gene
chr1	Loss	152555501	152586848	31347	chr1:152555501-152586848	6	0.75	7.44E−15	2.22E−16	*LCE3C, LCE3B*
chr1	Loss	210078101	210085706	7605	chr1:210078101-210085706	6	0.75	7.44E−15	2.22E−16	NA
chr2	Loss	98153590	98162215	8625	chr2:98153590-98162215	6	0.75	7.44E−15	2.22E−16	*ANKRD36B*
chr2	Loss	146862601	146876770	14169	chr2:146862601-146876770	6	0.75	7.44E−15	2.22E−16	NA
chr3	Loss	68739701	68747751	8050	chr3:68739701-68747751	6	0.75	7.44E−15	2.22E−16	NA
chr8	Loss	8073170	8086655	13485	chr8:8073170-8086655	6	0.75	7.44E−15	2.22E−16	NA
chr12	Gain	66225201	66235890	10689	chr12:66225201-66235890	6	0.75	1.78E−13	3.25E−09	*HGMA2*
chr15	Gain	20832601	20844600	11999	chr15:20832601-20844600	6	0.75	1.78E−13	2.12E−14	NA
chr15	Gain	22344037	22344161	124	chr15:22344037-22344161	6	0.75	1.78E−13	2.12E−14	Antibody heavy chains
chrX	Loss	115138001	115152874	14873	chrX:115138001-115152874	6	0.75	7.44E−15	5.03E−11	NA

NA, not available.

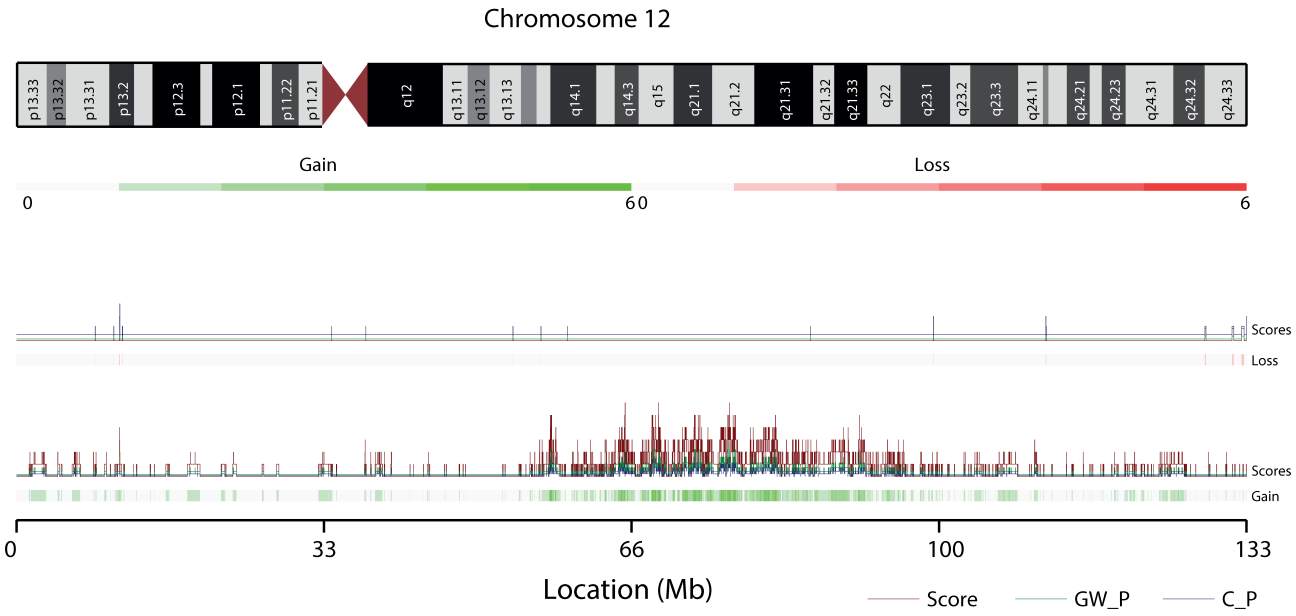

Figure 2. Chromosomal ideogram showing chromosomal gains/losses in chromosome 12.

significantly alternatively spliced (hsa04510, ES = 18.22, $p = 1.22 \times 10^{-8}$).

CIBERSORT analysis of genes associated with immune cell infiltration was carried out and showed that liposarcoma has variable immune cell infiltration, with all samples having some degree of T-cell infiltration, suggesting some degree of recognition of the tumour by the immune system (supplementary material, Table S7).

Gene fusion analysis demonstrates a recurrent fusion in SDHA which causes loss of activity

Gene fusion prediction (Table 4) via analysis of WGS data demonstrated recurrent fusion events in *HGMA2*, *SDHA*, *TSPAN31*, *MDM2*, and *WIF1*. *HGMA2* [28], *TSPAN31*, and *MDM2* [9] amplification/fusion have been well described in liposarcoma; however, *SDHA* (succinate dehydrogenase complex flavoprotein subunit A) has not previously been identified, with 11 fusions in four samples in this dataset. In order to validate this, we then carried out RNA fusion analysis using a targeted RNA panel analysis on a separate set of 24 FFPE samples (12 tumour/12 associated normal) as previously described. This validated (supplementary material, Table S8) a recurrent fusion in *SDHA* in 8/12 samples, between exon 12 of *SDHA* and exon 8 of *SLC9A3*.

Methylation analysis shows dysregulation of metastasis suppression

We then carried out methylome analysis using the Illumina HumanMethylation450 array. The top differentially methylated gene (supplementary material, Figure S6 and Table S9) was *ANKAR* (cg06479433, $\%\ \text{meth}_{\text{tumour}} = 95.0\%$, $\%\ \text{meth}_{\text{normal}} = 74.1\%$, $p_{\text{adj}} = 2.59 \times 10^{-10}$, BF = 19.01). Analysis of differentially methylated regions demonstrated six differentially methylated regions (DMRs) that met a genome-wide significance level of 1×10^{-6} (Table 5). The most significant DMR was *TIMP2*

Table 4. List of gene fusions as predicted by WGS.

Gene	Chromosome	Gene start	Gene end	No. of fusions	No. of samples	Sample ID
HMGA2	12	66217911	66360075	4	4	S000565, S038759, S038764, S038761
HLA-A	6	29909037	29913661	7	4	S000532, S000565, S038764, S038761
SDHA	5	218356	256815	11	4	S010732, S000565, S038759, S038764
TSPAN31	12	58131796	58143994	3	2	S010732, S038761
ESR1	6	151977826	152450754	4	2	S038758, S010732
CDC73	1	193091147	193223031	2	1	S038761
ERC1	12	1099675	1605099	2	1	S038761
NCOR2	12	124808961	125052135	2	1	S000565
SUFU	10	104263744	104393292	2	1	S038761
MYH11	16	15797029	15950890	3	1	S000565
NOTCH1	9	139388896	139440314	3	1	S038761
TPR	1	186280784	186344825	3	1	S038761
MDM2	12	69201956	69239214	5	1	S000565
WIF1	12	65444406	65515346	5	1	S000565
AFF3	2	100162323	100759201	17	1	S038761

Table 5. List of significantly differentially methylated regions (DMRs) in WD-LPS.

Rank	Gene ID	Gene feature	Chr	Start of DMR	End of DMR	Size of DMR (bp)	Change in per cent methylation	*P* value for DMR	Name of gene	Function
1	*TIMP2*	3' UTR_shore	17	76850012	76850463	452	−35%	1.71E−08	Tissue inhibitor of metalloprotineases 2	Metastasis suppressor
2	*C22orf9*	Body_none	22	45607945	45609196	1252	39%	2.42E−08	Chromosome 22 open reading frame 9	Lysine acetylation
3	*SLC44A4*	TSS200_none	6	31846943	31847023	81	−25%	1.13E−07	Solute carrier family 44 member 4	Overexpressed in prostate cancer
4	NA	IGR_shore	6	1601197	1601541	345	−34%	1.29E−07	CpG island 5 kb upstream of FOXC1	
5	NA	IGR_shelf	2	177020321	177024020	3700	−29%	6.73E−07	CpG 4 kb upstream of near HOXD3	
6	*BCL2L15*	Body_none	1	114429332	114430298	967	−40%	7.11E−07	BCL2-like 15	Pro-apoptotic gene

NA, not available.

(tissue metalloproteinase 2), which acts as a metastasis suppressor by suppression of the Wnt/β-catenin pathway [29]. We observed an average increase in methylation across the region of 35%, which would suggest that *TIMP2* becomes inactivated, thus dysregulating Wnt. Two DMRs were in CpG islands approximately 5 kb upstream of *FOXC1* (a Forkhead box gene of unknown function, implicated in both breast [30] and small cell lung cancer [31]) and *HOXD3* (a homeobox gene responsible for growth and differentiation [32]).

Gene set enrichment analysis of the differentially methylated positions (DMPs) was carried out, which demonstrated significant enrichment of gene sets associated with cancer, with the top ranked gene set being 'FARMER_BREAST_CANCER_CLUSTER_5', AUC = 0.84, $p = 2.94 \times 10^{-7}$.

Discussion

We have demonstrated a recurrent loss-of-function mutation in the *FOXD4L3* gene, a Forkhead box gene. This is in contrast to other studies [11], which suffered from poor coverage of *FOXD4L3* due to a mixed sample set. A significant weakness of our study is the lack of paired normal tissues for all samples, which may cause low frequency or rare mutant alleles to not be detected due to germline variation; however, we were limited by available tumour material because of the site and size of the tumour.

The *FOXD4L3* gene is within a complex region of an ancestral duplication site [33] in chromosome 9 that is postulated to have occurred early in hominid evolution. Although the sequence similarity is high [34] between the different paralogues of *FOXD4* (approximately >95%) the principal difference between paralogues is the presence of an indel leading to a frameshift, either a 1-bp deletion at codon 292 or a 53-bp deletion at codon 363, causing alternate protein products to be produced. We observed a stop gain mutation at codon 260, well upstream of the previously described deletions, within the Forkhead binding domain, supported by Sanger sequencing primers specific for *FOXD4L3*. This mutation has been observed in dbSNP 147 (rs7021123) from ExAC data, but at extremely low frequency (MAF = 0.00825) in one subject only. We did not observe any RNA sequencing reads that aligned to *FOXD4L3*, despite it being expressed in normal fat on the Illumina Human BodyMap 2.0 database. Typically, a premature stop codon would cause nonsense mediated decay (NMD) of the product after the initial round of translation but as *FOXD4L3* is a single exon gene, we cannot with certainty point to NMD as the mechanism here.

We successfully demonstrated the previously identified [11] recurrent copy number change in *HGMA2*, the binding partner of MDM2, as part of amplification of a larger region on the long arm of chromosome 12 rather than an effect specific to *HGMA2*.

Our matched RNAseq analysis showed dysregulation in both the PI3K–AKT and the MAPK pathways, as well as genes associated with the transition away from fat to a 'tumour' phenotype. Methylation analysis of these tumours demonstrated activation of pro-apoptotic genes and genes associated with metastasis. The recurrent fusion within *SDHA* is also of interest, as similar genes (*SDHB* in paraganglioma [35] and *SDHC* in gastrointestinal stromal tumours [36]) are implicated in tumourigenesis.

Further study and functional models are required to understand the effect of changes in these genes and their relevance in WD-LPS as well as the consequences and effects of *FOXD4L3* mutation in association with WD-LPS. However, multiple potential target pathways for therapy have now been identified.

Acknowledgements

We thank Illumina UK (Miao He, Sean Humphray, Mahesh Vasipalli, Zoya Kingsbury, and David Bentley) for sequencing and bioinformatics support for this project. This work was partially funded by a pump priming grant from the Cancer Research UK Birmingham Experimental Cancer Medicine Centre. AB acknowledges funding from the Wellcome Trust (102732/Z/13/Z),

Cancer Research UK (C31641/A23923), and the Medical Research Council (MR/M016587/1).

Author contribution statement

ADB, AD and DG recruited patients and collected samples. PT and ADB carried out histopathological analysis and immunohistochemistry. ADB, MPD, JJ, DB and JDS processed samples and carried out DNA/RNA extraction. ADB, MPD, JJ, DB and JDS performed validation sequencing. ADB carried out data analysis. ADB, MPD, JJ, DB, JDS, DGM, PT, DG and AD wrote and reviewed the manuscript.

Data availability statement

Sequencing and microarray data are available for access in the NIH SRA Archive with Accession ID PRJNA751390 (http://www.ncbi.nlm.nih.gov/bioproject/751390).

A preprint related to this article has been deposited at: https://www.preprints.org/manuscript/202101.0008/v1 [Not peer-reviewed].

References

1. Crago AM, Singer S. Clinical and molecular approaches to well differentiated and dedifferentiated liposarcoma. *Curr Opin Oncol* 2011; **23:** 373–378.
2. Pasquali S, Vohra R, Tsimopoulou I, *et al*. Outcomes following extended surgery for retroperitoneal sarcomas: results from a UK Referral Centre. *Ann Surg Oncol* 2015; **22:** 3550–3556.
3. Dei Tos AP, Doglioni C, Piccinin S, *et al*. Coordinated expression and amplification of the *MDM2*, *CDK4*, and *HMGI-C* genes in atypical lipomatous tumours. *J Pathol* 2000; **190:** 531–536.
4. Micci F, Teixeira MR, Bjerkehagen B, *et al*. Characterization of supernumerary rings and giant marker chromosomes in well-differentiated lipomatous tumors by a combination of G-banding, CGH, M-FISH, and chromosome- and locus-specific FISH. *Cytogenet Genome Res* 2002; **97:** 13–19.
5. Pedeutour F, Suijkerbuijk RF, Van Gaal J, *et al*. Chromosome 12 origin in rings and giant markers in well-differentiated liposarcoma. *Cancer Genet Cytogenet* 1993; **66:** 133–134.
6. Pedeutour F, Forus A, Coindre JM, *et al*. Structure of the supernumerary ring and giant rod chromosomes in adipose tissue tumors. *Genes Chromosomes Cancer* 1999; **24:** 30–41.
7. Riggi N, Cironi L, Provero P, *et al*. Expression of the FUS–CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma. *Cancer Res* 2006; **66:** 7016–7023.
8. Kåbjörn Gustafsson C, Ståhlberg A, Engtröm K, *et al*. Cell senescence in myxoid/round cell liposarcoma. *Sarcoma* 2014; **2014:** 208786.
9. Conyers R, Young S, Thomas DM. Liposarcoma: molecular genetics and therapeutics. *Sarcoma* 2011; **2011:** 483154.
10. Thway K, Flora R, Shah C, *et al*. Diagnostic utility of p16, CDK4, and MDM2 as an immunohistochemical panel in distinguishing well-differentiated and dedifferentiated liposarcomas from other adipocytic tumors. *Am J Surg Pathol* 2012; **36:** 462–469.
11. Kanojia D, Nagata Y, Garg M, *et al*. Genomic landscape of liposarcoma. *Oncotarget* 2015; **6:** 42429–42444.
12. Amin-Mansour A, George S, Sioletic S, *et al*. Genomic evolutionary patterns of leiomyosarcoma and liposarcoma. *Clin Cancer Res* 2019; **25:** 5135–5142.
13. Raczy C, Petrovski R, Saunders CT, *et al*. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* 2013; **29:** 2041–2043.
14. Saunders CT, Wong WS, Swamy S, *et al*. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* 2012; **28:** 1811–1817.
15. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 2013; **41:** e108.
16. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30:** 923–930.
17. Law CW, Chen Y, Shi W, *et al*. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; **15:** R29.
18. Ritchie ME, Phipson B, Wu D, *et al*. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43:** e47.
19. Chen B, Khodadoust MS, Liu CL, *et al*. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* 2018; **1711:** 243–259.
20. Lal N, Beggs AD, Willcox BE, *et al*. An immunogenomic stratification of colorectal cancer: implications for development of targeted immunotherapy. *Oncoimmunology* 2015; **4:** e976052.
21. Haas BJ, Dobin A, Li B, *et al*. Accuracy assessment of fusion transcript detection via read-mapping and *de novo* fusion transcript assembly-based methods. *Genome Biol* 2019; **20:** 213.
22. Morris TJ, Butcher LM, Feber A, *et al*. ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 2014; **30:** 428–430.
23. 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.
24. Chen X, Schulz-Trieglaff O, Shaw R, *et al*. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016; **32:** 1220–1222.
25. Roller E, Ivakhno S, Lee S, *et al*. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* 2016; **32:** 2375–2377.
26. Cazier JB, Holmes CC, Broxholme J. GREVE: Genomic Recurrent Event ViEwer to assist the identification of patterns across individual cancer samples. *Bioinformatics* 2012; **28:** 2981–2982.
27. Begum S, Emami N, Cheung A, *et al*. Cell-type-specific regulation of distinct sets of gene targets by Pax3 and Pax3/FKHR. *Oncogene* 2005; **24:** 1860–1872.
28. Italiano A, Bianchini L, Keslair F, *et al*. *HMGA2* is the partner of *MDM2* in well-differentiated and dedifferentiated liposarcomas whereas *CDK4* belongs to a distinct inconsistent amplicon. *Int J Cancer* 2008; **122:** 2233–2241.
29. Xia Y, Wu S. Tissue inhibitor of metalloproteinase 2 inhibits activation of the beta-catenin signaling in melanoma cells. *Cell Cycle* 2015; **14:** 1666–1674.
30. Jensen TW, Ray T, Wang J, *et al*. Diagnosis of basal-like breast cancer using a FOXC1-based assay. *J Natl Cancer Inst* 2015; **107:** djv148.
31. Wei LX, Zhou RS, Xu HF, *et al*. High expression of FOXC1 is associated with poor clinical outcome in non-small cell lung cancer patients. *Tumour Biol* 2013; **34:** 941–946.
32. Chen LN, Rubin RS, Othepa E, *et al*. Correlation of HOXD3 promoter hypermethylation with clinical and pathologic features in screening prostate biopsies. *Prostate* 2014; **74:** 714–721.
33. Fan Y, Newman T, Linardopoulou E, *et al*. Gene content and function of the ancestral chromosome fusion site in human

chromosome 2q13–2q14.1 and paralogous regions. *Genome Res* 2002; **12:** 1663–1672.
34. Fan Y, Linardopoulou E, Friedman C, *et al*. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Res* 2002; **12:** 1651–1662.
35. Neumann HP, Pawlu C, Peczkowska M, *et al*. Distinct clinical features of paraganglioma syndromes associated with *SDHB* and *SDHD* gene mutations. *JAMA* 2004; **292:** 943–951.
36. Pantaleo MA, Astolfi A, Urbini M, *et al*. Analysis of all subunits, SDHA, SDHB, SDHC, SDHD, of the succinate dehydrogenase complex in KIT/PDGFRA wild-type GIST. *Eur J Hum Genet* 2014; **22:** 32–39.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. Bar chart of the frequency of mutations from whole genome sequencing data for genes taken from the Cancer Gene Census and from previous literature on liposarcoma

Figure S2. Diagram of *FOXD4L3* showing the Forkhead binding domain

Figure S3. Ensemble reference diagram showing the expression of *FOXD4L3* in normal adipose tissue via RNAseq and cDNA, and in pooled RNAseq data

Figure S4. Circos plots of structural variation in four samples

Figure S5. Structural rearrangement plots from whole genome sequencing data

Figure S6. Methylation volcano plot of differentially expressed genes between liposarcoma and normal fat

Table S1. Table of primers for targeting of exon 1 of *FOXD4L3*

Table S2. Table of mutation metrics for liposarcoma samples

Table S3. Mutation spectra

Table S4. Table of associations between *FOXD4L3* and other genes from the STRINGS database

Table S5. RNAseq data from tumour versus normal

Table S6. Alternate splicing events from RNAseq data of tumour versus normal

Table S7. CIBERSORT analysis for tumour RNAseq

Table S8. Fusion calling from RNAseq

Table S9. List of methylation variable positions comparing liposarcoma with normal