The molecular landscape of well differentiated retroperitoneal liposarcoma

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

DOI:
10.20944/preprints202101.0008.v1

License:
Creative Commons: Attribution (CC BY)

Document Version
Early version, also known as pre-print

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 26. Oct. 2023
The molecular landscape of well differentiated retroperitoneal liposarcoma

Robert Tyler¹, Mark P Dilworth¹, Jonathan James¹, Daniel Blakeway¹, Joanne D Stockton¹, Dion G. Morton¹, Phillipe Taniere², David Gourevitch², Anant Desai², Andrew D Beggs¹

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

Address for correspondence:

Andrew Beggs
Institute of Cancer & Genomic Science
University of Birmingham
Vincent Drive
Birmingham
B15 2TT

Email: a.beggs@bham.ac.uk

Keywords: Liposarcoma, whole genome sequencing, RNA sequencing, methylation

Additional information:

Funding: The 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).

Data statement: Sequencing and microarray data is available for access in the European Genome Phenome Archive (ref: XXX)

Acknowledgments: We thank Illumina UK (Miao He, Sean Humphray, Mahesh Vasipalli, Zoya Kingsbury and David Bentley) for sequencing and bioinformatics support for this project.
Competing financial interests: The authors declare they have no competing financial interests

Author contribution statement:

Sample collection and patient recruitment: ADB, AD, DG
Histopathological analysis and IHC: PT, ADB
Sample processing & DNA/RNA extraction: ADB, MPD, JJ, DB, JDS
Validation sequencing: ADB, MPD, JJ, DB, JDS
Data analysis: ADB
Manuscript: ADB, MPD, JJ, DB, JDS, DGM, PT, DG, AD.
ABSTRACT

Well differentiated liposarcoma (WD-LPS) is a relatively rare tumour, with less 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 exon1 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 tumorigenesis.
INTRODUCTION

Retroperitoneal liposarcoma (RPLS) is the most common type of retroperitoneal sarcoma (RPS), making up around a half of all RPS (1). RPLS typically presents incidentally on cross-sectional imaging as a large homogenous retroperitoneal mass (2). As these tumours become larger, the risk of their transformation from well-differentiated to de-differentiated increases. The molecular events that initiate WD-LPS and lead to de-differentiation of well differentiated liposarcoma are currently unknown. Molecular changes seen in the various types of RPLS include co-ordinated amplification of MDM2 and CDK4 (3), giant ring chromosomes (4-6), amplification of CDK4, HMG2 and TSPAN31 and the FUS-DDIT3 fusion oncogene, seen in a subset of liposarcomas caused by t(12;16) or (t12;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 HGMA2. 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
METHODS

Sample collection

Retroperitoneal sarcoma samples were obtained from patients undergoing laparotomy for en-bloc excision of retroperitoneal mass. Immediately after the specimen was removed from the patient, it was conveyed to the Queen Elizabeth Hospital Birmingham Histopathology laboratories where a representative sample of liposarcoma and associated normal fat was taken, 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.) were sampled. Where no frozen material was available to match with the specimen, formalin fixed, paraffin embedded tissue blocks were retrieved from the histopathology archives of the Queen Elizabeth Hospital, Birmingham.

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 5mm$^3$ piece of tissue was immersed in buffer ATL overnight with 20mg proteinase K. The resulting lysate was processing using the Qiagen DNEasy Blood & Tissue kit (Qiagen UK). Extracted DNA was eluted into buffer AL and was quantified using Nanodrop spectrophotometry for purity and Qubit (Thermo Scientific, UK) fluorimetry for DNA concentration.

Because of limited volumes of tissue, RNA extraction was performed on formalin fixed, paraffin embedded sections. Three 10µM FFPE sections were cut using a microtome then processed using the Purelink FFPE RNA extraction kit (Invitrogen) as per manufacturer’s instructions. Extracted RNA was quantified for purity using Nanodrop spectrophotometry and for quantity using Qubit fluorimetry.

**Whole genome sequencing (WGS) & targeted sequencing**

Shotgun WGS was performed using 500ng 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 as per manufacturers guidelines. For six samples (associated normal), no fresh frozen material was available and therefore a PCR based whole genome sequencing library preparation was used (Illumina FFPE) using 1µg of DNA. Sequencing 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 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 in where no control normal was available, a panel of normals were 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. A long range PCR (conditions available on request) was performed and PCR amplicons cleaned and ligated using an Oxford Nanopore ligation sequencing kit (LSK-109). Library was then loaded onto a ONT MinIon 9.5.1 flow cell and run for 36 hours. Data was base called using Guppy and variant calling performed by MiniMap2. No polishing was performed.

Whole genome sequencing data for the well differentiated liposarcoma line 93T449 was downloaded from the NCBI SRA archive. Downloaded data was 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 1µg 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) and purified. The final libraries were sequenced on an Illumina HiSeq 2000 to an average of 35 million reads with a read length of 90bp.

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 Human Methylation450 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 manufacturers’ 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 exported. Differential methylation was quantified using eBayes/limma and differentially methylated regions 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 Files (VCF) to MAF (Mutation Annotation File) using vcf2maf. In patients were 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 intrachromosomal 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 Exon Primer with a Tm of 60°C. Primer specificity was verified using UCSC in-silico PCR and primers were optimised with gradient PCR before use. 10ng of tumour DNA underwent PCR using the Qiagen Multiplex kit using primers against FOXD4L3 (supplementary table 6). 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 forwards and reverse primers at 0.4 µM, 25 µL LongAmp Taq 2X Master Mix (New England Biolabs, USA) and nuclease-free water up to a total of 50 µL. Thirty cycles of long range PCR (with a 2.5 minute 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 hours. Raw sequence FAST5 files were demultiplexed and converted to FASTQ using the Guppy Base caller. Called FASTQ files underwent two rounds of polishing with Canu followed by a round of polishing with Medaka. Sequence data was aligned using
minimap2 to the GRCh38 genome followed by variant calling using FreeBayes (command line `freebayes -C 2 -O 0 -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.

**Whole genome sequencing reveals recurrent mutations in FOXD4L3**

Mutational rate in all tumours was between 1.65-1.8 mutations/mega base, making WDLPS a relatively infrequently mutated tumour to others studied in the TCGA. Analysis of recurrent mutations using MutSigCV2 within the samples demonstrated significant mutations (Table 2, 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 table 1) 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). We also analysed recurrently mutated genes within the Cancer Gene Census and those previously reported in liposarcoma...
(11) using the Illumina Encore pipeline (Supplementary figure 1) including immune evasion (HLA-A, 5/8 samples), Wnt signalling (FAT1; 4/8 samples), double strand 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 details and found a recurrent mutation with 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. The mutation was just upstream of the Forkhead box domain of the protein (Supplementary Figure 2), suggesting that this mutation may have functional consequences due to loss of this protein binding domain.

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 and 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. We found that the variant allele frequency 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 (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 6), demonstrated the recurrent copy number gain in *HGMA2*, the binding partner of *MDM2* (28), in six out of eight cases. Recurrent amplification was seen (Figure 3) 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 Figure 5) of small scale chromosomal translocation across the tumour types. Re-arrangement of both between and within chromosomes 6 and 12, was seen (Supplementary figure 6), 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 table 6). The most differentially expressed gene was *TNS1* (Tensin 1, normalised counts 20.54 vs. 110.9, p=2x10^-6, Q=0.03), which is part of the PTEN signalling pathway. Pathway analysis however demonstrated dysregulated pathways in “Pathways in Cancer” (hsa05200, ES
11.61, p=9.1x10^{-6}), MAPK signalling pathway (hsa04010, ES 6.49, p=1.52x10^{-3}) and PI3K-Akt signalling (hs04151, ES 5.26, p=5.21x10^{-3}).

Analysis of alternative splicing was carried out using STAR (Supplementary table 7), with FN1 (fibronectin 1) being the top ranked alternatively spliced gene (Normalised counts tumour = 10.28, normal=0.41; p=4.52x10^{-9}, Q=8.25x10^{-6}. Pathway analysis demonstrated the focal adhesion pathway was significantly alternatively spliced (hsa04510, ES=18.22, p=1.22x10^{-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 table 8).

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

Gene fusion prediction (Table 3) 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 4 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 table 9) 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 HumanMethylation 450 array. The top differentially methylated gene (Supplementary table 4) was ANKAR (cg06479433, %meth\textsubscript{tumour} = 95.0, %meth\textsubscript{normal} = 74.1%, p\textsubscript{adj} = 2.59x10^{-10}, BF = 19.01. Analysis of differentially methylated regions demonstrated six differentially methylated regions that met a genome wide significance level of 1x10^{-6} (table 4). The most significant DMR was TIMP2 (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 TIMP2 becomes inactivated, thus dysregulating Wnt. Two DMRs were in CpG islands approximately 5kb 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 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=2.94x10^{-7}. 

19
CONCLUSIONS

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*, 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 paralogs of FOXD4 (approximately > 95%) the principle difference between paralogs is the presence of an indel leading to 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, however 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, however 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 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 tumorigenesis.

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 WDLPS, however multiple potential target pathways for therapy have now been identified.
REFERENCES


Figure 1: Spider chart of mutation type in whole genome sequencing data
Figure 2: Chromosomal ideogram showing chromosomal gains/losses in chromosome 12
### Table 1: Table of patient characteristics

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Sample ID</th>
<th>Type</th>
<th>Gender</th>
<th>Age</th>
<th>Tumour location</th>
<th>Tumour histological type</th>
<th>Patient alive?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P000111</td>
<td>S000532</td>
<td>TUMOUR</td>
<td>Male</td>
<td>59</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Alive</td>
</tr>
<tr>
<td>P000214</td>
<td>S000565</td>
<td>TUMOUR</td>
<td>Male</td>
<td>76</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Deceased</td>
</tr>
<tr>
<td>P001190</td>
<td>S038759</td>
<td>TUMOUR</td>
<td>Female</td>
<td>63</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Alive</td>
</tr>
<tr>
<td>P001193</td>
<td>S038757</td>
<td>TUMOUR</td>
<td>Female</td>
<td>81</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Deceased</td>
</tr>
<tr>
<td></td>
<td>S038758</td>
<td>NORMAL</td>
<td>Female</td>
<td>81</td>
<td>Retroperitoneum</td>
<td>Normal fat</td>
<td>Deceased</td>
</tr>
<tr>
<td>P001207</td>
<td>S010732</td>
<td>TUMOUR</td>
<td>Female</td>
<td>60</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Deceased</td>
</tr>
<tr>
<td></td>
<td>S044139</td>
<td>NORMAL</td>
<td>Female</td>
<td>60</td>
<td>Retroperitoneum</td>
<td>Normal fat</td>
<td>Deceased</td>
</tr>
<tr>
<td>P001212</td>
<td>S038764</td>
<td>TUMOUR</td>
<td>Male</td>
<td>58</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Alive</td>
</tr>
<tr>
<td>P001224</td>
<td>S010732</td>
<td>TUMOUR</td>
<td>Female</td>
<td>63</td>
<td>Retroperitoneum</td>
<td>Well differentiated liposarcoma</td>
<td>Deceased</td>
</tr>
<tr>
<td></td>
<td>S038770</td>
<td>NORMAL</td>
<td>Female</td>
<td>63</td>
<td>Retroperitoneum</td>
<td>Normal fat</td>
<td>Deceased</td>
</tr>
</tbody>
</table>
Table 2: List of mutations ranked by significance

<table>
<thead>
<tr>
<th>Rank</th>
<th>Gene</th>
<th>Gene name</th>
<th>p-value</th>
<th>q-value</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPRN</td>
<td>Shadow Of Prion Protein Homolog</td>
<td>8.21E-09</td>
<td>1.55E-04</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>ADH1C</td>
<td>Alcohol dehydrogenase 1C</td>
<td>1.52E-07</td>
<td>1.43E-03</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>FOXD4L3</td>
<td>Forkhead box D3 Like 3</td>
<td>2.57E-07</td>
<td>1.62E-03</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>KRTAP2-2</td>
<td>Keratin Associated Protein 2-2</td>
<td>2.04E-06</td>
<td>9.63E-03</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>LILRB3</td>
<td>Leukocyte Immunoglobulin-Like Receptor, Subfamily B</td>
<td>3.37E-06</td>
<td>1.27E-02</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>GYPE</td>
<td>Glycophorin-E</td>
<td>8.26E-05</td>
<td>2.60E-01</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>GOLGA8B</td>
<td>Golgin A8 Family, Member B</td>
<td>1.22E-04</td>
<td>3.28E-01</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>OR2A42</td>
<td>olfactory receptor family 2 subfamily A member 42</td>
<td>2.05E-04</td>
<td>4.82E-01</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>IL7R</td>
<td>Interleukin 7 Receptor</td>
<td>2.52E-04</td>
<td>4.82E-01</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>C12orf77</td>
<td>Chromosome 12 Open Reading Frame 77</td>
<td>2.68E-04</td>
<td>4.82E-01</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3: List of gene fusions as predicted by WGS

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
<th>Gene start</th>
<th>Gene end</th>
<th>no. fusion</th>
<th>no. sample</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMGA2</td>
<td>12</td>
<td>66217911</td>
<td>66360075</td>
<td>4</td>
<td>4</td>
<td>S000565, S038759, S038764, S038761</td>
</tr>
<tr>
<td>HLA-A</td>
<td>6</td>
<td>29909037</td>
<td>29913661</td>
<td>7</td>
<td>4</td>
<td>S000532, S000565, S038764, S038761</td>
</tr>
<tr>
<td>SDHA</td>
<td>5</td>
<td>218356</td>
<td>256815</td>
<td>11</td>
<td>4</td>
<td>S010732, S000565, S038759, S038764</td>
</tr>
<tr>
<td>Rank</td>
<td>Gene ID</td>
<td>Gene feature</td>
<td>Chr</td>
<td>Start of DMR</td>
<td>End of DMR</td>
<td>Size of DMR in bp</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>--------------</td>
<td>-----</td>
<td>--------------</td>
<td>------------</td>
<td>------------------</td>
</tr>
<tr>
<td>TSPAN31</td>
<td>12</td>
<td>58131796</td>
<td>58143994</td>
<td>3</td>
<td>2</td>
<td>S010732, S038761</td>
</tr>
<tr>
<td>ESR1</td>
<td>6</td>
<td>151977826</td>
<td>152450754</td>
<td>4</td>
<td>2</td>
<td>S038758, S010732</td>
</tr>
<tr>
<td>CDC73</td>
<td>1</td>
<td>193091147</td>
<td>193223031</td>
<td>2</td>
<td>1</td>
<td>S038761</td>
</tr>
<tr>
<td>ERC1</td>
<td>12</td>
<td>1099675</td>
<td>1605099</td>
<td>2</td>
<td>1</td>
<td>S038761</td>
</tr>
<tr>
<td>NCOR2</td>
<td>12</td>
<td>124808961</td>
<td>125052135</td>
<td>2</td>
<td>1</td>
<td>S000565</td>
</tr>
<tr>
<td>SUFU</td>
<td>10</td>
<td>104263744</td>
<td>104393292</td>
<td>2</td>
<td>1</td>
<td>S038761</td>
</tr>
<tr>
<td>MYH11</td>
<td>16</td>
<td>15797029</td>
<td>15950890</td>
<td>3</td>
<td>1</td>
<td>S000565</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>9</td>
<td>139388896</td>
<td>139440314</td>
<td>3</td>
<td>1</td>
<td>S038761</td>
</tr>
<tr>
<td>TPR</td>
<td>1</td>
<td>186280784</td>
<td>186344825</td>
<td>3</td>
<td>1</td>
<td>S038761</td>
</tr>
<tr>
<td>MDM2</td>
<td>12</td>
<td>69201956</td>
<td>69239214</td>
<td>5</td>
<td>1</td>
<td>S000565</td>
</tr>
<tr>
<td>WIF1</td>
<td>12</td>
<td>65444406</td>
<td>65515346</td>
<td>5</td>
<td>1</td>
<td>S000565</td>
</tr>
<tr>
<td>AFF3</td>
<td>2</td>
<td>100162323</td>
<td>100759201</td>
<td>17</td>
<td>1</td>
<td>S038761</td>
</tr>
</tbody>
</table>

Table 4: List of significantly differentially methylated regions in WDLPS
Table 5: Recurrent copy number events

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