SRF-MYOCD axis is the targetable driver of a well differentiated aggressive subtype of leiomyosarcomas

Elodie Darbo[1,2,3], Gaëlle Pérot[4,5], Lucie Darmusey[4,6,7], Sophie Le Guellec[4,6], Laura Leroy[4,6], Laëtitia Gaston[8], Nelly Desplat[1], Noémie Thébault[4,6], Candice Merle[4,6,7], Philippe Rochaix[4,6], Thibaud Valentin[4,9], Gwenaël Ferron[4,10], Christine Chevreau[9], Binh Bui[11], Eberhard Stoeckle[12], Dominique Ranchere-Vince[13], Pierre Méeus [14], Philippe Terrier[15], Sophie Piperno-Neumann[16], Françoise Collin[17], Gonzague De Pinieux[18], Florence Duffaud[19], Jean-Michel Coindre[1,20], Jean-Yves Blay[21,22], Frédéric Chibon (Frederic.chibon@inserm.fr)[1,4,6*]

1. INSERM UMR1218, ACTION, Institut Bergonié, 33000, Bordeaux, France.
2. CNRS UMR5800, LaBRI, 33400, Talence, France.
3. Université de Bordeaux, 33000, Bordeaux, France.
4. OncoSarc, INSERM U1037, Cancer Research Center in Toulouse (CRCT), 31000, Toulouse, France.
5. Centre Hospitalier Universitaire (CHU) de Toulouse, IUCT-Oncopole, 31000, Toulouse, France.
6. Department of Pathology, Institut Claudius Régaud, IUCT-Oncopole, 31000, Toulouse, France.
7. University of Toulouse 3, 31000, Toulouse, France.
8. Department of Medical Genetics, CHU de Bordeaux, 33000, Bordeaux, France.
9. Department of Oncology, Institut Claudius Régaud, IUCT-Oncopole, 31000, Toulouse, France.
10. Department of Surgical Oncology, Institut Claudius Régaud, IUCT-Oncopole, 31000, Toulouse, France.
11. Department of Oncology, Institut Bergonié, 33000, Bordeaux, France.
12. Department of Surgery, Institut Bergonié, 33000, Bordeaux, France.
13. Department of Pathology, Centre Léon Bérard, 69000, Lyon, France.
14. Department of Surgery, Centre Léon Bérard, 69000, Lyon, France.
15. Department of Pathology, Institut Gustave Roussy, 94800, Villejuif, France.
16. Department of Medical Oncology, Institut Curie, 75005, Paris, France.
17. Department of Pathology, Centre Georges-François Leclerc, 21000, Dijon, France.
18. Department of Pathology, Hôpital Universitaire Trousseau, 37170, Tours, France.
19. Medical Oncology Unit, Aix Marseille University, APHM Hôpital La Timone, 13000, Marseille, France.
20. Department of Pathology, Institut Bergonié, 33000, Bordeaux, France.
21. Department of Medical Oncology, Centre Léon Bérard, 69000, Lyon, France.
22. Université Claude Bernard Lyon 1, INSERM U1052, CNRS 5286, Centre Léon Bérard, 69000, Lyon, France.

* Correspondence: Frédéric Chibon, Cancer Research Center in Toulouse (CRCT), 2 avenue Hubert Curien, 31037, Toulouse, France, 0582741765, Frederic.chibon@inserm.fr

Abstract

Leiomyosarcoma is a very aggressive tumor with poor prognosis since half of the patients will develop metastasis. Today, neither chemotherapy nor targeted or immune therapies have demonstrated efficacy. These tumors show highly rearranged genomes and appear to form a heterogenous group. Understanding their complex oncogenesis either as a single disease or by subtypes has already been challenged but the underlying mechanisms driving leiomyosarcoma development which could lead to a patient care improvement remain to be deciphered. We identified a group of tightly clustered leiomyosarcomas due to their gene expression homogeneity, named "hLMS". We derived a transcriptional signature able to consistently stratify patients from two independent cohorts. In all cohorts, hLMS were preferentially carried by women, located in the internal trunk, highly differentiated, and similarly altered at the genomic level. Through the multi-omics integrative bioinformatic analysis, we show that hLMS originate from vascular smooth muscle cells presenting both contractile and synthetic characteristics, while the other group could derive from fibroblasts. We identified the MYOCD overexpression as a hLMS-specific driver candidate and functionally showed that the MYOCD/SRF axis is only essential for hLMS survival. Identification of hLMS could become standard clinical practice, leading to the development of specific effective treatments with MYOCD/SRF inhibitors.

Key words: leiomyosarcoma, differentiation, therapeutic target, transcriptional regulation, bioinformatics, high throughput

Introduction

Leiomyosarcoma (LMS) is a rare mesenchymal malignancy presenting smooth muscle differentiation and accounting for 11% of adult soft tissue sarcomas (STS) (1). LMS arises in various anatomical sites but mostly the uterus, the retroperitoneum and the limbs. LMS is one of the most aggressive STS subtypes as up to 50% of patients relapse (2) with a median survival of 12 months. However, the main treatment involves wide surgical resection for localized LMS or anthracycline-based chemotherapies for metastatic tumors, since neither targeted therapy (3) nor immunotherapy (4) have demonstrated any major therapeutic effects until now.

LMS oncogenesis is organized around a highly rearranged genome with a high number of chromosomal rearrangements leading to many copy number variations (CNV) and break-points, which are associated with poor outcome (5). However, no specific recurrent pathogenic event has been detected so far.

Stratifying patients on gene-expression profiling based on the unique prognostic and therapeutic characteristics of the tumors has been shown to improve patient outcome, thanks to the development of adapted therapies in several cancer types such as breast (6) and colon (7) cancer.

Subgroups of LMS have been identified by using different transcriptomic sequencing methods and sample collections (8–16). One subgroup was consistently highlighted with a homogeneous transcriptomic profile, expression of smooth muscle cell (SMC) differentiation markers, a higher differentiation and localized in the retroperitoneum. Nonetheless, no standardized methodology has yet emerged and the discrepancies between the reported subtypes hamper understanding of the specific biology or oncogenesis of these tumors that is essential for improving clinical care.

In an elegant study, Watson and colleagues reported that sarcomas with a strong chimeric driver oncogene have a homogenous and specific transcriptomic program (17). Therefore, we hypothesized that such a mechanism could be identified in LMS showing this characteristic. Accordingly, we conducted a systematic transcriptome-wide investigation comparing LMS (with exclusion of uterine LMS which are quite well recognizable (18)) to hundreds of other sarcomas. This highlighted two groups of LMS differing in their transcriptional homogeneity. We developed a transcriptional signature to robustly classify LMS into two groups with consistent clinical associations and gene expression in three independent cohorts. By combining “omics” data at the genome, transcriptome and micro RNA levels, we were able to establish two distinct patterns of oncogenesis that differentiate LMS into at least two pathologies originating from different cell types. One of the subgroups predicted by the signature presents a therapeutic vulnerability, which could pave the way for a new treatment option.

Results

Identification of a group of 42 LMS behaving as simple genetic sarcomas

To detect LMS molecular subtypes within sarcoma samples, we combined micro-array datasets obtained on Affymetrix (387 complex genetic sarcomas including 98 LMS) and Agilent platforms (60 GIST, 58 synovial sarcomas, 50 LPS and 87 complex genetic sarcomas) (**Figure S1A**, total = 555 samples). We selected 9066 genes (out of 17854 genes common to both platforms, **Figure S1B**) showing enough consistency to enable merging and normalization of all datasets (see methods and **Figure S1C**). We assumed that selecting modules of co-expressed genes that potentially group genes with similar functions would lead to more meaningful patient clustering. We detected 15 co-expression modules (out of 54) carrying at least 5 genes from 455 highly correlated genes. Thirteen modules were significantly associated with biological functions and cellular components (*e.g.* immune system activation, cell cycle, skeletal muscle or smooth muscle-related, adipogenesis, extracellular matrix, apical plasma membrane, genomic positional bias, **Table S1A**, **Figure S1D**). We used these 54 modules in a non-supervised approach to cluster the 555 samples and observed a subgroup of LMS clustering together, while the other LMS were mixed with other pleomorphic sarcomas. This LMS subgroup appeared to behave like sarcomas with a recurrent alteration, *i.e.* with a fairly homogenous transcriptomic program driven by a strong oncogene (17), as observed with GIST, myxoid liposarcomas and synovial sarcomas (**Figure 1A**). We thus hypothesized that this LMS subgroup (41 patients over the 98 LMSs) could be driven by a strong oncogenic program reflected by this specific gene expression profile.

To select genes that best characterized these LMS, we compared them with the remaining 57 LMS which were mixed with the other sarcomas. As the 98 LMS were all analyzed on the Affymetrix chip, we used the 22635 genes present in the chip. We identified 1672 differentially expressed genes (**Table S2A**) that we used to re-cluster the samples. Almost all

samples were classified similarly (95/98) regarding the analysis performed above on the 555 samples. We obtained 42 homogeneous LMS (hLMS) and 56 other LMS (oLMS) (**Figure 1B**).

hLMS are intra-abdominal, low-grade and metastatic LMS with homogeneous transcriptional behavior

After having confirmed that gene expression profiles within hLMS were significantly more homogeneous than within the other group (Wilcoxon's test; $P = 2.9 \times 10^{-13}$), we tested clinical feature enrichments (**Table 1**). hLMS were mostly located in the abdominal cavity ($P = 8.5 \times 10^{-9}$), developed in females ($P = 0.003$), were well differentiated ($P = 3.9 \times 10^{-9}$) and consequently were more frequently grade 1 or 2 (low grades, $P = 5.5 \times 10^{-4}$). Interestingly, despite this differentiation and grading, they had a poorer prognosis than oLMS ($P = 0.0054$, **Figure 1C**).

hLMS are characterized by both contractile and synthetic smooth muscle cell phenotypes

Functional enrichment analysis of differentially expressed (DE) hLMS/oLMS genes (**Figure 1D** and detailed in **Tables S1B** and **C**) revealed biological differences between the two groups. The transcriptional program in hLMS is strongly associated with smooth muscle cell and cell cycle activity, as evidenced by the enrichment of *E2F* and *RB1* targets, CINSARC signature, DNA replication, metabolism and mitochondrial activity in up-regulated genes. In line with these results, activating marks (H3K4me1, H3K27Ac and H3K9Ac) from ChiP-seq experiments in smooth muscles (stomach, rectum, colon, aorta) as well as ChiP-seq peaks for *SRF* and *MEF2A* were enriched in over-expressed hLMS genes (**Table S1D, Figure 1E**). On the other hand, oLMS were associated with ER-Golgi related terms, epithelial-mesenchymal transition and the TGFβ signaling pathway, while enriched histone marks in over-expressed

oLMS genes were found to be comparable to those in fibroblasts, epithelial and derived mesenchymal stem cells. These genes are under the regulation of transcription factors (TF) like *MYC*, *ETS1* or *ELK1*. Therefore, we hypothesize that hLMS and oLMS originate from distinct cell types.

To investigate the potential origin of hLMS, we analyzed the 100 most expressed hLMS genes in 7414 samples from 30 different normal tissues (TCGA GTEX dataset). Using a t-SNE approach, we observed that these genes allowed normal samples to be grouped mainly according to their tissue of origin (**Figure S2A**). Visceral smooth muscle tissues were mixed and separated from blood vessels to which the hLMS were the closest. hLMS and oLMS were well separated and oLMS showed a wider distribution between lung, adipose and breast tissues. These results support our hypothesis that the two LMS groups have a different origin and suggest that hLMS could originate from vascular smooth muscle cells.

We annotated these 100 genes (**Figure S2B**) using GSAn (19) (**Table S1E**) and found that 50 of them are part of the extracellular exosome, which are molecules (mRNA or proteins) exported to the extracellular space. This highlights the role of the extracellular matrix (ECM) and of cell-to-cell communication in hLMS pathology. Cell differentiation and migration were represented by 32 and 24 genes, respectively, which suggests the co-existence of both contractile (*MYH11, CNN1, MYL9, LMOD1)* and synthetic (*FN1, TNC, COL1A1/2, MSN, MFAP4*) phenotypes in hLMS and in blood vessels.

To complete our analysis of the genomic differences between hLMS and oLMS, we used multi-omics to analyze two additional LMS cohorts.

Gene signature identifies hLMS in two independent cohorts

To classify LMS from the ICGC (59 patients) and TCGA (75 patients) (20) cohorts, we computed the distance to hLMS and oLMS centroids based on the expression of the 1672 DE

genes from the Affymetrix cohort. When the cohorts were merged, 102 cases were strongly enough correlated with one centroid (**Figure 1F),** classifying 73 as hLMS and 29 as oLMS. Computation of clinical enrichment showed hLMS to be mainly intra-abdominal ($P = 1.5 \times 10^{-7}$), well differentiated ($P = 1.8 \times 10^{-5}$), carried by women ($P = 0.007$) and with homogeneous transcriptional profiles ($P < 2.2 \times 10^{-16}$) (**Table 1**), consistent with information from the training cohort. However, we observed no difference in metastasis-free survival between the two groups.

miRNAs adopt specific behavior in hLMS

We analyzed 475 expressed mature miRNAs from the 39 patients in the ICGC cohort (28 hLMS and 11 oLMS) and 453 in the 60 TCGA patients (43 hLMS and 17 oLMS). PCA analysis performed with all expressed miRNAs strongly differentiated hLMS and oLMS along the first principal component, which explains 37% (ICGC) and 50.2% of the variance (TCGA) (**Figure 2A**). The high correlation ($R^2 = 0.69$, **Figure 2A**) between hLMS/oLMS log-fold changes from both cohorts indicates that each group identified independently in each cohort is consistent and represents two groups of similar diseases. The results of DE analyses of both cohorts are presented in **Table S2B**.

We used the TCGA pan cancer (PANCAN) dataset to validate our hypothesis that the two groups have a different cellular origin. To this end, we used the 41 significantly differentially expressed miRNAs (35 under-expressed and 6 over-expressed in hLMS) to classify all the cancer samples (**Figure 2B**). All hLMS clustered together among 467 samples mainly from prostate adenocarcinomas (65%), digestive track tumors (stomach, colon, esophagus, rectum: altogether 14%), LMS (13 gynecological, 8 unclassified: together with hLMS, 13.7%). The most discriminative miRNAs of the cluster containing hLMS were 4 of the 6 over-expressed miRNAs in hLMS (*MIR143-3p*, *MIR145-3/5p* and *MIR1*). These miRNAs are involved in

feedforward (*MIR143/145*) (21) and negative feedback (*MIR1*) (22) loops during smooth muscle differentiation. These results corroborated our hypothesis of a smooth muscle origin of hLMS, unlike oLMS which were spread across several clusters.

Interestingly, all 87 mature miRNAs located in the DLK1-DIO3 imprinted genomic region on chromosome 14 (14q32) were repressed in hLMS. Indeed, 25 miRNAs are among the 35 significantly down-regulated in hLMS (highlighted in **Figure 2C**), 20 other show negative log-fold changes (**Figure S3A**) with very low expression in hLMS (**Figure S3B**), and 42 were not detected in any LMS groups in at least one of the cohorts. To evaluate the specificity of this global repression, we compared the expression profiles of the 72 miRNAs (among the 87 DLK1-DIO3) present in the PANCAN dataset. Most hLMS (37/42) clustered within a group of 563 patients, representing 6% of all samples, preferentially with kidney (37%), thyroid (23%), eye (11%) carcinomas and sarcomas (5.5%, 13 gynecological LMS, 1 unclassified LMS, 2 oLMS, 7 UPS, 6 myxofibrosarcomas, 3 dedifferentiated liposarcomas) (**Figure 3C**), and none of the visceral smooth muscle-related cancers (stomach, colon, rectum, etc.). These results suggest an uncommon repression that might be due to a specific mode of oncogenesis rather than to a vSMC origin.

To evaluate the putative impact of dysregulated miRNAs on hLMS biology, we analyzed their post-transcriptional regulatory network by integrating mRNA and miRNA expression data. We found 14620 significant miRNA-mRNA interactions predicted in both ICGC and TCGA cohorts (negative Pearson's correlation coefficient (PCC), adjusted $P<0.01$), of which 210 were present in at least one database (**Table S2C**). We annotated the 158 corresponding target genes (35 down- and 123 up-regulated in hLMS) with GSAn (19). Twenty-one terms with high specificity were mapped and none of them was specific to DIO3-DLK1 miRNA cluster, up- or down-regulated target genes, except "response to glucose" which was represented only by up-regulated genes (**Figure 2D,** detailed in **Table S1F**). Dysregulated genes are implicated

in major pathways, such as cell migration ("plasma membrane-bound cell projection assembly", "extracellular matrix disassembly"), cell contraction ("regulation of heart contraction", "cation channel activity", "calcium ion transport"), cell cycle and transcriptional regulation. Of note, significant interactions involving the most deregulated miRNAs were previously reported to have an impact on SMC phenotypes. Indeed, *MIR28*, which was over-expressed in hLMS, was found either to promote proliferation targeting *NME1* (average PCC: -0.61) or to inhibit it by targeting *CCND1* (average PCC: -0.48) in colon cancer (23). *MIR455*, *MIR199a* and *MIR503/MIR424*, which were under-expressed in hLMS, suppress proliferation and migration in pulmonary arterial or bladder SMC targeting *FGF7* (24) (average PCC: -0.7), *DDR1* (25) (average PCC: -0.7) and *FGF2*, respectively (average PCC: -0.63/-0.68) (26).

The transcriptional profile of miRNAs and its predicted regulatory network revealed the putative vascular smooth muscle origin and showed a strong relationship with both vSMC contractile and synthetic phenotypes. The question therefore arose whether these two types of LMS also have a distinct mode of oncogenesis.

hLMS show recurrent and specific genomic instability

LMS are characterized by their highly rearranged genome (15). Copy number alterations in hLMS appeared more homogeneous than in oLMS both in the merged cohort (**Figure 3**) and between the cohorts, which indicates highly correlated penetrance profiles (**Figure S4A**). Recurrent alterations were significantly enriched in hLMS, especially amplification of chromosome 17p12-p11.2 and loss of chr10q, chr13q14, chr17p13 (**Figure 3**). The chr17p12-p11.2 amplified region was not only significantly enriched in hLMS with 31% of hLMS showing an amplification *versus* 7-8% of oLMS, but was also the most frequently amplified region in hLMS (**Figure 4A, Table S3A**). Among the genes carried by this region, *MYOCD*

was the most frequently amplified (36% of hLMS), and was the most over-expressed gene in this region in hLMS compared to oLMS ($P < 10^{-7}$ all cohorts considered, **Figures 4A, 4B, Table S3B**). *MYOCD* expression was very high in 84% of hLMS (97/115, detailed per cohort in **Figure S4B**), whereas it was not expressed or at a very low level in oLMS, even in those with a gain or an amplification (**Figure 4B, Figure S4C**, **Table S3B**).

Interestingly, the well-known tumor suppressors *RB1*, *PTEN* and *TP53* belong to three of the eight most significantly enriched lost regions in hLMS: chr13q14 (88% *versus* 72%), chr10q23 (87% *versus* 66%) and chr17p13 (69% *versus* 13%) respectively (all cohorts considered) (**Figure 3, Table S3B**). When we used the whole genome characterization of the ICGC to further investigate these genetic variations, we found that oLMS tended to be more rearranged ($P < 0.05$, **Figure S4D**) than hLMS but that the mutational burden was similar between them ($P = 0.5$, **Figure S4D**). While no COSMIC mutational signature could be associated with the LMS groups, we found a patient-specific predicted contribution of signatures mainly related to defective DNA repair, except for LMS23, which had a disproportionate mutational burden (120 mutations/Mb *versus* less than 1 mutation/Mb for the other) and a mutational profile similar to ultraviolet light exposure, which is coherent with its location on the scalp. (**Figure S4E**). Very few genes were identified as recurrently mutated (SNV). However, by combining the different alterations, *i.e.* mutations, structural variants (SV) and losses, we found very frequently altered genes across all ICGC-LMS: *TP53* altered in 100% of cases, *RB1* in 97.4%, *PTEN* in 82%, *ATRX* in 28.2% and *DMD* in 25.6%. (**Table 2**, **Tables S4**).

We found no significant difference in *RB1* and *TP53* global alteration frequencies between the two LMS groups. However, *TP53* presented significantly different alteration patterns. Indeed, oLMS preferentially lost *TP53* completely (9/11, 82%) whereas 64.3% of hLMS (18/28) exhibited different alterations on each allele with losses, missense and frameshift

mutations (Fisher's exact test, P = 0.01, **Figure 4C**). The same trend was observed for *RB1*, without reaching significance (**Figure 4C, Table S4A**).

PTEN was almost exclusively altered by complete gene deletion, regardless of the LMS type (**Table 2**, **Table S4A**). However, although 82% of cases in both groups were altered, its protein expression loss was significantly associated with hLMS (**Figure 4E**).

ATRX mutations are described in detail in another paper from our team, in which we reported their characterization in the whole ICGC cohort (including the 39 LMS studied here). We showed that *ATRX* alteration and ATRX protein expression loss are associated with uterine LMS and the oLMS type (27). Accordingly, ATRX nuclear localization was significantly enriched in hLMS (**Table S4A, Figure 4E**).

DMD tended to be more frequently altered in oLMS than in hLMS (45.4% and 17.8%, respectively). Most *DMD* alterations involved SV, which mainly affects the *DMD* long isoforms (**Table 2**, **Table S4A and S4C**). Regardless of *DMD* genomic status, *Dp427m*, its muscle specific transcript isoform, was significantly less expressed in oLMS ($P = 1.2 \times 10^{-9}$), as was *Dp40* ($P = 1.9 \times 10^{-4}$). On the other hand, the expression of *Dp71*, a ubiquitous isoform, was similar in both LMS types (P = 0.63) **(Table S4A, Figure 4D**). Results were confirmed at the protein level, with a significant association of global DMD expression loss and particularly of Dp427 in the oLMS type **(Figure 4E**).

Therefore, despite having similar alteration frequencies of the two major suppressor genes *TP53* and *RB1*, the mechanistic differences and specific expression enrichments of the two LMS types suggest that their oncogenic processes are different. The main features underpinning this distinction are the amplification and strong expression of *MYOCD* and the loss of PTEN protein in hLMS. Indeed, these specific features of hLMS are related to *SRF/MYOCD*, the main drivers of smooth muscle cell differentiation, given that PTEN also

interacts with SRF (28). We tested the hypothesis that the SRF/MYOCD axis could be a driver of hLMS oncogenesis by investigating the therapeutic inhibition of this pathway.

hLMS can be targeted specifically with an SRF/MYOCD inhibitor

We studied the impact on cell viability of inhibitors specifically targeting the SRF/MYOCD pathway. CCG-1423, an inhibitor of the SRF/MRTF interaction (29), and CCG-100602, an inhibitor of the SRF/MYOCD interaction (30), were tested on 3 LMS (OC80: hLMS with a *MYOCD* amplification, OC48: oLMS with a *MYOCD* gain and OC88: oLMS) and 2 UPS (OC98 and OC110) cell lines (**Figure 5A**).

After 72h of treatment with increasing concentrations of each CCG, cell viability assay showed that all cell lines were sensitive to both inhibitors (IC50 ranging from 2.56±1.36 µg/mL to 21.41±3.95 µg/mL) (**Figures 5B** and **C**). Regardless of their subgroup, LMS were slightly more sensitive to CCG-1423 than UPS, with OC88 reaching significance and being more affected than OC110 (**Figure 5B**). Interestingly, responses to CCG-100602, which is specific to the SRF/MYOCD interaction, exhibited three kinds of behavior: OC80 (hLMS; IC50 = 2.85±1.15 µg/mL) was the most sensitive, OC88 (oLMS; IC50 = 6.70±0.95 µg/mL) had an intermediate response, and OC48 (oLMS; IC50 = 19.44±3.88 µg/mL) and the two UPSs (IC50 = 15.32±3.42 and 19.69±0.67 µg/mL) were the least receptive (**Figure 5C**). Overall, all cell lines had a lower IC50 with the SRF/MYOCD inhibitor than with the SRF/MRTF inhibitor. However, the higher responsiveness of hLMS compared to others when SRF/MYOCD was inhibited indicates that the oncogenic dependency to the SRF/MYOCD axis is stronger in hLMS.

Discussion

We have identified two groups of LMS with specific transcriptional and genomic alteration profiles, particular clinical features, and specific modes of oncogenesis and cellular origin. Several authors have already identified two or three subtypes of LMS consistently with one group harboring strong smooth muscle differentiation and a homogeneous transcriptomic profile that are similar to hLMS (8–16). However, our study differs from the others. First, we analyzed primary tumors and excluded uterine LMS, unlike most of the other authors (9, 11, 12, 14). Second, we report reproducible clinical descriptions of hLMS and oLMS. hLMS are highly differentiated, preferentially carried by females, low-grade and with an intra-abdominal location, while oLMS are poorly differentiated, high-grade and located in the extremities. In addition, outcome was poorer in hLMS in the Affymetrix cohort but not in the other cohorts. There are two non-exclusive possible explanations for this: i) group sizes in the ICGC and TCGA studies were not balanced with an under-representation of oLMS cases, in contrast with the Affymetrix cohort; ii) retroperitoneal LMS are known to have a longer survival than other STS in general (some over 10 years) and to develop late metastases (31), so the follow-up could be too short, leading to an underestimation of the rate of metastasis in hLMS. Furthermore, we conducted an original integrative analysis of mRNA, miRNA, copy numbers, mutations and breakpoints, so we were able to better characterize each LMS group and make new biological findings about these tumors.

Our data show that the over-expression of smooth muscle-related genes in hLMS (compared to oLMS) (e.g. *MYH11*, *TAGLN*, *ACTA2* etc) is probably triggered by MYOCD, as previously demonstrated by Pérot *et al*. (10). *MYOCD* was found to be over-expressed in more than 84% of hLMS following genetic amplification in 36% of cases. To enhance SM differentiation, MYOCD needs to co-operate with an ubiquitous transcription factor (TF), SRF (32), which

regulates the expression of targeted genes by binding to an element known as the CArG-box which is located upstream of smooth muscle (SM) contractile genes (33, 34). Accordingly, we found a high enrichment in predicted SRF binding sites (*in silico* and ChIP-seq) in over-expressed hLMS gene promoters. Moreover, Dp427, a DMD isoform, which is under the control of MYOCD (35), is nearly always expressed at the membrane in hLMS regardless of *DMD* genomic status, unlike in oLMS in which it is no longer expressed. As the longest DMD isoform loss at the membrane is highly recurrent in other myogenic diseases (36), its presence in the membrane might be needed in hLMS. Further investigations in hLMS are now necessary to better understand its role. Given the presence of smooth muscle active histone marks in over-expressed hLMS genes and their closer similarity to normal blood vessels than to viscera smooth muscles, we hypothesize that hLMS have a vascular SMC (vSMC) origin rather than arising from an acquired phenotype that triggers metastatic development (10).

The cellular origin of oLMS appears to be quite different. oLMS showed many regulatory and functional features (active histone marks, many TF binding sites predicted from a cell line showing an epithelial morphology and EMT-related terms) associated with fibroblasts, adipocytes, mesenchymal stem cells (MSC) and epithelial cells, and were spread among normal lung, adipose and breast tissues. Therefore, they could derive i) from the de-differentiation of cells from their location; ii) directly from circulating or local MSC; iii) from fusion between circulating or local MSC with a cell from the tumor site (37). Whatever their origin, this would explain the hyper-activation of the unfolded protein response observed in oLMS, which might allow the cells to adapt to the external (undifferentiated phenotype in a specialized environment) or internal (managing extra material after fusion) micro-environment (38). The heterogeneity observed among patients with oLMS and the difficulty to define a unique oncogenesis might be due to their heterogeneous cellular origin.

A novel finding was that hLMS are not only contractile and differentiated but are also proliferative, with migratory features revealing the co-existence of contractile and synthetic phenotypes. These characteristics are probably inherited from the specificity of vSMC, which have highly plastic phenotypes and can cover a wide spectrum of phenotypes from synthetic to contractile (39). Indeed, among the most expressed hLMS genes, we found markers of both phenotypes also highly expressed in blood vessel samples that could represent the natural mixture of vSMC, spanning the phenotypic continuum in these samples. The synthetic phenotype is sustained by the propensity of hLMS to proliferate via strong enhancement of the cell cycle, as suggested by the significant over-expression of *E2F1*, the enrichment of up-regulated hLMS genes in *E2F/RB1* targets, *E2F7* binding sites and cell cycle functional related terms specific to hLMS. This is contradictory with the suggestion of Hemming *et al*. (14), who considered that all LMS possess this feature. It is probably due to the low number of "other" LMS that they had in their different cohorts when comparing LMS with other STS, then highlighting what they considered to be their "conventional" characteristics. This hyper-activation arises in a context where alteration of the tumor suppressors *TP53*, *RB1* and *PTEN*, is altered equally in hLMS and oLMS but acquired through different preferential mechanisms, as shown by the genetic analysis, may boost proliferation and cell survival in a non-specific manner. Moreover, the PTEN protein is specifically totally absent in hLMS. Recently, it was demonstrated that PTEN is involved in SMC differentiation through direct interaction with SRF (28). In fact, the SRF-regulating abilities of SMC are dual: they control the expression of both smooth muscle (SM) contractile genes, depending on MYOCD, and growth-related immediate early genes (IEG) (33), depending on ELK1. Horita and colleagues showed that in the nucleus, PTEN is directly linked to SRF and helps it to link only to SM gene CArG-Boxes (28), and that PTEN is translocated into the cytoplasm upon forced SMC switch toward proliferation, which in turns increases IEG gene expression. Therefore, the total absence of

PTEN in hLMS may allow SRF to be linked with both SM genes and IEG CArG-Boxes, so SRF may express SMC contractile and proliferative genes. However, the large amount of MYOCD may compete with the other SRF interactors and lead to the observation of a significantly lower expression of targets of the ETS-family factors in hLMS than in oLMS. Nevertheless, it should be remembered that by comparing two tumor types, the activity of ETS-related transcription in hLMS may be under-estimated.

The interplay between the two phenotypes may require a fine-tuned regulation which could be partly ensured by miRNAs, as suggested by miRNA profiling analysis. Interestingly, the SRF/MYOCD complex targets all over-expressed miRNAs in hLMS, either directly upon binding to the CArG-boxes present in their promoters (22, 40) or indirectly by targeting their host gene, as in *MIR28* and *LPP*. miRNAs deregulation in hLMS could, at the same time, strengthen the contractile phenotype through a feed-forward loop (*MIR143/MIR145* (40) over-expression) and inhibition of proliferation (*MIR28-5p* over-expression as in colon cancer (23)), while maintaining the synthetic phenotype by impairing contractility (*MIR1* overexpression (22)), promoting migration (over-expression of *MIR28-3p* as in colorectal cancer (23), and of its host gene, *LPP*, as in differentiated LMS (10)) and increasing proliferation (*MIR455*, *MIR199a and MIR503/MIR424* under-expression as in pulmonary artery SMC (24, 25) and in bladder SMC (26)). We also observed a global repression of the *DLK1-DIO3* miRNAs cluster, which is highly specific to hLMS and papillary thyroid carcinomas. Its predicted target genes are involved in vasculature development and cell migration in both diseases (41), so it is probably involved in this dual phenotype. However, this cluster involves around 50 miRNAs with different functions in different cellular contexts (42), so understanding how this repression actually impacts hLMS biology will require further investigations.

Altogether, our findings suggest that hLMS originate from vSMC that do not achieve terminal differentiation, retain a remarkable degree of plasticity, probably grow in a maintained differentiated state (39, 43) and take advantage of the enhanced contractile apparatus to migrate (10). Our data show that many antagonists of vSMC differentiation are present in hLMS and that their contractile abilities appear to be essential for the oncogenesis of hLMS. In turn, this suggests that over-expression of *MYOCD* is positively selected and that it could be the event that triggers tumorigenesis.

Loss of PTEN might be a favorable event for hLMS growth, and targeting the Pi3k/AKT/mTOR pathway in LMS (mostly probable hLMS), which is physiologically inhibited by PTEN, was thought to be promising (44). Subsequently, however, it was demonstrated to induce resistance (45). Thus, targeting downstream of this pathway with an SRF/MYOCD inhibitor could overcome this resistance. We thus conducted functional inhibition assays with SRF inhibitors which showed greater efficiency on hLMS than on oLMS or UPS. Moreover, the specific inhibitor of the MYOCD/SRF interaction, which is specific to SMC, showed greater efficiency than the MRTF/SRF inhibitor, which is more specific to skeletal muscle.

In conclusion, effective treatments for hLMS and oLMS are more likely to be developed if these two entities can be properly differentiated. In this regard, the SRF/MYOCD axis is a promising target in hLMS. We now need to probe the action of an SRF/MYOCD inhibitor in hLMS to pave the way for the efficient targeted treatment of LMS.

Methods

Experimental model and subjects details

Human samples

Among samples used in the training cohort 278 out of the 387 complex genetics sarcomas (5, 15, 46, 47), the 60 GIST (48), and the 58 synovial sarcomas (49) are part of cohorts previously described (Table S5). All samples used in this cohort were part of the CRB-IB. In accordance with the French Public Health Code (articles L.1243-4 and R. 1243-61), the CRB-IB has received accreditation from the French authorities to use samples for scientific research. Samples used in the ICGC cohort were collected prospectively by the French Sarcoma Group as part of the ICGC program (International Cancer Genome Consortium). Clinico-pathological data and patient information are summarized in **Table S6**. All cases were systematically reviewed by expert pathologists of the French Sarcoma Group according to the World Health Organization guidelines (50). Patients' written informed consent was approved by the Committee of Protection of Individuals.

Cell lines and primary culture

All cell lines OC80 (LMS ; Male), OC88 (LMS ; Male), OC48 (LMS ; Female), OC98 (UPS ; Male) and OC110 (UPS ; Female) are primary cultures established as previously described (51) and were cultured in RPMI-1640 (524000-025, Life Technologies, Carlsbad, CA, USA) supplied with 10% fetal bovine serum (S1810-500, Dutscher, Brumath, France) and cells were kept at 37°C in a humidified chamber containing 5% CO_2.

Data aquisition

Expression microarray data

The 387 complex genetics sarcomas were analyzed on Human Genome U133 Plus 2.0 array (900466, Affymetrix, Santa Clara, CA, USA), according to the manufacturer's procedures. For GIST, synovial sarcomas, LPS and 87 complex genetics sarcomas, gene expression

analysis was carried out by Agilent Whole Human 44K Genome Oligo Array (G4112A, Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol.

Copy number data

CGH from Affymetrix cohort (Table S5): Array-CGH for 53 cases was performed using BAC-array as described in Chibon et al., 2010 (5), and with Genome-Wide Human SNP 6.0 arrays (901153, Affymetrix, Santa Clara, CA, USA) for 31 cases, according to the manufacturer's protocol with 500 ng DNA as input.

Sequencing data

DNA, total RNA and miRNA were extracted from frozen samples of the ICGC cohort and sequenced using Illumina technologies (Illumina Inc., San Diego, CA, USA) HiSeq2000 for DNA and RNA samples (paired-end) and HiSeq2500 for miRNA samples (single-end). Extraction, library preparation and sequencing protocols are detailed in **Supplementary methods**.

Sequencing data analysis

RNA sequencing (RNA-seq)

Alignment and expression quantification were performed as previously described (47). Fusion transcripts were detected with Defuse v0.6.1 (52) as previously described (53).

miRNA sequencing (miRNA-seq)

Reads were then trimmed for adaptors using cutadapt version 1.10 (54) with -q 30 and -m 18 parameters. Sequencing quality was assessed using FastQC from the Babraham Institute (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We then aligned the reads with mature miRNA sequences from the miRbase Sequence database (55) according to the recommendations in (56). We first used the BWA-aln algorithm version 0.7.17 (57) with -n 1 -o 0 -e 0 -k 1 -t 8 parameters and evaluated mapping quality using Qualimap version 2.2.2b

(58). We then used Samtools version 1.9 (59) to discard the reads with a mapping quality under 30 and to count mapped reads with the reference sequences (*samtools idxstats*).

Whole Genome Sequencing (WGS)

DNA reads were trimmed of the 5' and 3' low quality bases (phred cut-off 20, max trim size 30 nt) and sequencing adapters were removed with Sickle2 (60) (https://github.com/najoshi/sickle) and SeqPrep3 (https://github.com/jstjohn/SeqPrep), respectively. Then, DNA-curated sequences were aligned using Bowtie v2.2.1.0 (61), with the very sensitive option, on the Human Genome version hg19. Thus, aligned reads were filtered out if their alignment score was less than 20 or if they were duplicated PCR reads, with SAMtools v0.1.19 (59) and PicardTools v1.118 (62) (http://broadinstitute.github.io/picard/), respectively.

Single nucleotide and structural variants detection

Single Nucleotide Variant (SNV)

SNV were detected in RNA-seq and WGS data using samtools mpileup (SAMtools v0.1.19 (59)), with a minimum of 20 as phred quality score (-Q 20), and bcftools (SAMtools v0.1.19 (63)) with options view –cvg for RNA-seq data and call –Am for WGS data. RNA-seq detected variants with fewer than 5 coverage reads were filtered out. The variants detected in normal, tumor DNA and tumor RNA were merged in the same file. Then, somatic variants were extracted with: (i) a minimum coverage of 14 reads in the tumor and 8 in the normal and (ii) a minimal allelic fraction of 0.3 in tumor and 0 in normal. Variants were annotated using the Annovar v20160314 tool (64). Variants were selected whose alternative allele frequency (AF) in the Caucasian population (CEU) is lower than 0.1%, as reported in the 1000Genome database (65). Finally, variants were kept if they were localized in coding regions and were non-synonymous.

Structural variants

Breakpoints were detected from WGS data. Paired-end reads were aligned using Bowtie v2.2.1.0 (61), very sensitive local option allowing soft-clipped sequences. The algorithm has three main steps: i) identification of potential breakpoints, ii) characterization of the second side of the breakpoints, and iii) selection of high-confidence breakpoints. All parameters were set to analyze 60X tumor and 30X normal sequencing depth. Very conservative filters were used to minimize false positive detection. Details are available in the **Supplementary methods**.

Copy number variants (CNV): Genome-Wide Human SNP 6.0 arrays were analyzed as previously described in (66). Genes absent in more than a third of the patients were discarded. From ICGC WGS paired tumor/normal data were processed using the cn.MOPS R package (67) with default parameters and a 500-nucleotide window. Regions were intersected with TxDb.Hsapiens.UCSC.hg19.knownGene R package version 3.2.2 (68) gene models. Regions with an estimated copy number of 128 were discarded.

For both datasets, genes overlapping segments with different predicted copy numbers were attributed with the lowest number of copies.

Experimental validation

Fluorescent in situ hybridization

FISH assay was performed on tissue microarrays using the Histology FISH accessory kit (K579911-2, Dako, Agilent Technologies, Santa Clara, CA, USA) according to the manufacturers' instructions. Thirty-eight tumors from the ICGC cohort were analyzed. Each case was represented by three spots 4 µm-thick and 1mm in diameter. FISH assay was performed using a commercially available MYOCD FISH probe labeled in spectrum orange and a chromosome 17 control probe labeled in FITC (EG-MYOCD-CHR17-20ORGR, Empire Genomics, Williamsville, NY, USA). *MYOCD* and control probe enumeration was

performed with a Nikon Eclipse 90i fluorescent microscope with appropriate filters. Pictures were captured using a Pannoramic 250 Flash II Digital Slide Scanner and analyzed with the Pannoramic Viewer (3DHISTECH Ltd., Budapest, Hungary). A case was considered as interpretable when almost 80% of cells presented a signal for both probes. A loss was when only one copy of *MYOCD* was observed in the majority of cells; a normal status was when two copies of *MYOCD* gene were detected in the majority of cells; a gain or polysomy was when 3 to 5 copies of *MYOCD* or both *MYOCD* and control probe were detected; and amplification was when the number of *MYOCD* signals was equal to or greater than 6, especially when clustered signals were observed.

Alterations verification strategy

For the ICGC cohort, *ATRX*, *TP53*, *RB1*, *PTEN* and *DMD* sequences for each case obtained by whole genome sequencing were entirely screened using the Integrative Genomics Viewer (IGV version 2.6.3 (69)) to search for alterations possibly missed by the detection algorithms used. All SV were verified on gDNA by PCR and Sanger sequencing. MS/NS mutations not found in either WGseq or RNAseq and all FS were verified at both DNA and RNA levels by PCR and RT-PCR, respectively, followed by Sanger sequencing. For samples with enough material left, fusion transcripts detected by RNA-seq were verified by RT-PCR and Sanger sequencing.

For screening of mutations on genomic DNA, PCR primers were designed using the Primer 3 program (70)(https://bioinfo.ut.ee/primer3-0.4.0/). All PCR were performed on 50ng of DNA using AmpliTaqGold® DNA polymerase (4311820, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. PCR program for *TP53* mutations validation is described in (71) For other PCR, the PCR program used is a Touch-down 60°C program (TD 60°C) (Hybridization temperatures: 2 cycles at a temperature of 60°C, followed

by 2 cycles at 59°C, 2 cycles at 58°C, 3 cycles at 57°C, 3 cycles at 56°C, 4 cycles at 55°C, 4 cycles at 54°C, 5 cycles at 53°C and finally 10 cycles at 52°C).

Total RNA was first reverse-transcribed using random hexamers and the High Capacity cDNA Reverse Transcription Kit (4368814, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. All primers used were designed using the Primer 3 program (70)(https://bioinfo.ut.ee/primer3-0.4.0. All PCR were performed as previously described for PCR on genomic using TD60°C PCR program.

Sanger Sequencing

PCR products were purified using an ExoSAP-IT PCR Purification Kit (US78200, GE Healthcare, Piscataway, NJ, USA) and sequencing reactions were performed with the Big Dye Terminator V1.1 Kit (4336805, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's recommendations. Samples were purified using the Big Dye XTerminator Purification kit (4376486, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions and sequencing was performed on a 3730xl Genetic Analyzer for cohort 1 or 3130xl Genetic Analyzer for cohort 2 (Applied Biosystems, Foster City, CA, USA). Sequences were then analyzed using the Sequencing analysis V5.3.1 and the SeqScape V2.6 software (Life Technologies, Carlsbad, CA, USA). FinchTV software (V1.4.0) was also used (Geospiza, Seattle, WA, USA).

Immunohistochemistry

IHC assays were performed on tissue microarrays. Immunohistochemistry for PTEN detection was performed on a BenchMark Ultra instrument (Ventana, Washington D.C, USA). Antigen retrieval was performed using a CC1 protocol for 4 min at 100°C (Ventana, Washington D.C, USA). The anti-PTEN antibody (1:200, 9559, RRID:AB_390810, Clone 138G6, Cell Signaling Technology, Danvers, MA, USA) was diluted in Prep kit 26 (783-

2876, Roche, Basel, Switzerland) and incubated for 1h. Antibody detection was performed with the Optiview detection kit for 12 min (860-099, Ventana, Washington D.C, USA). Immunohistochemistry for P53 detection was performed on a Bond-III (Leica Microsystems, Wetzlar, Germany) using the clone DO-7 monoclonal antibody (GA61661-2, Dako Omnis, ready-to-use, incubation 20 min, Agilent Technologies, Santa Clara, CA, USA). Antibody detection was performed using EnVision FLEX/HRP (GV80011-2, Dako, Agilent Technologies, Santa Clara, CA, USA). Immunohistochemical pictures were taken with a Pannoramic 250 Flash II Digital Slide Scanner and analyzed with the Pannoramic Viewer (3DHISTECH Ltd., Budapest, Hungary).

Immunofluorescence

Immunofluorescence was performed on tissue microarrays. First, tissues were de-paraffinized in three xylene baths for 5 minutes and then rehydrated in successive baths of ethanol from 100% to 70%. For heat-induced epitope retrieval, slides were incubated for 20 minutes in a microwave oven in DAKO Target Retrieval pH6 (S203130-2, DAKO, Agilent Technologies, Santa Clara, CA, USA). Then, they were incubated with primary antibodies overnight at 4°C in a humidity chamber and after with secondary antibodies for 1 hour at room temperature. Slides were then mounted with Vectashield mounting medium plus DAPI (H-1200-10, Vector Laboratories, Burlingame, CA, USA). Images were acquired on a Zeiss Cell Observer Microscope (Zeiss, Oberkochen, Germany) or a confocal microscope LSM 780 (Zeiss, Oberkochen, Germany). Primary antibodies used for Dp427 was the anti-human Dystrophin form Leica Biosystems (1:25, Dy4/6D3, NCL-DYS1, RRID:AB_442080, Leica Biosystems, Wetzlar, Germany) and for all the dystrophin isoforms, the Dystrophin antibody from Abcam (1:100, ab15277, RRID:AB_301813, Abcam, Cambridge, UK). Secondary antibodies used were goat anti-mouse Alexa Fluor 488 (1:400, A-11001, RRID:AB_2534069, Thermo Fisher

Scientific, Waltham, MA, USA) and goat anti-rabbit Alexa Fluor 594 (1:400, A-11072, RRID:AB_2534116, Thermo Fisher Scientific, Waltham, MA, USA).

Cytotoxicity analysis

LMS and UPS cells were seeded in 96-well microplates at a density of 2.10^3 cells per well in 100µL RPMI-1640 medium (524000-025, Life Technologies, Carlsbad, CA, USA). After 24h incubation at 37°C, 100µL of medium were added with 2X of final concentration (100 to 1.5 µg/mL) of either CCG-1423 (10010350, CAS:285986-88-1, Bertin Bioreagent, Montigny le Bretonneux, France) or CCG-100602 (10787, CAS:1207113-88-9, Bertin Bioreagent, Montigny le Bretonneux, France). Cells were incubated at 37°C for 72h, after which 20µl (5mg/mL) solution of MTT (M2128, 5mg/mL, Sigma, St Louis, MO) dissolved in water were added. After 2h of incubation, media were removed and the MTT metabolic product formazan was dissolved in 100µl DMSO (5879, Sigma, St Louis, MO, USA). Absorbance at 570nm and 650nm was measured with the Clariostar plate reader (BMG Labtech, Ortenberg, Germany) and cell viability was analyzed as followed: $((A_{sample(570nm)} - A_{sample(650nm)})/(A_{Control(570nm)} - A_{Control(650nm)})) * 100$. IC_{50} was calculated with GraphPad Prism (GraphPad Software, San Diego, CA, USA) using non-linear regression. Each experiment was made 3 times in triplicate.

Bioinformatics and statistical analysis

Normalization of Affymetrix and Agilent micro-arrays and gene selection

We used 87 samples analyzed on Agilent and Affymetrix platforms (Table S5). We selected the genes with Pearson's correlation coefficient (PCC) with itself between chips over 0.8 or better than with any other genes in both experiments. We then normalized gene expression first in separate experiments and then on a merged dataset by applying quartile normalization (preprocessCore R package version 1.48.0 (72)). We harmonized the expression between the

platforms by gene expression median centering in each experiment and then adding the mean of the experiment medians. The method is illustrated in Figure S1 and detailed in Supplementary methods.

Gene module clustering

To define groups of co-expressed genes, we computed pairwise Pearson's correlation coefficient (PCC) of gene expression with variance > 2 across patients. We built a graph of co-expression with correlated genes (PCC > 0.7) and search for communities using edge.betweenness.community from igraph R package version 1.2.5 (73).

Sample clustering and PCA analysis

In all cases concerning unsupervised clustering, samples were clustered using *PCA* and *HCPC* functions from the FactorMiner R package version 2.3 (74) and were visualized using the pheamap R package version 1.0.12 (75). PCA analysis on ICGC and TCGA miRNA transcriptome were computed using the *prcomp* R function and visualized using the ggbiplot R package (76). We used the R package Rtsne version 0.15 (77) to visualize GTEX data with parameters dims=2 perplexity=100 and max_iter=1000.

Patient classification

We computed centroids with citccmst R package version 1.0.2 (7) for hLMS and oLMS using the 1672 differentially expressed genes detected in the Affymetrix cohort. Distance to centroids was computed as 1 - Spearman's correlation coefficient for each patient from all three cohorts. Patient classification was performed using the mclust R package version 5.4.6 (78). We selected the Gaussian mixture distribution estimation that best fitted the hLMS centroid distance distribution (maximization of Bayesian Information Criterion).

Clinical enrichment

Clinical enrichment significance was performed using two-tailed Fisher's exact test for categorical data comparing one category to the others and Wilcoxon's test for continuous data.

Survival analysis

Survival analysis was performed using the survival R package version 3.1-12 (79) by fitting a simple Kaplan Meier model and we set the significance at log-rank test p-value < 0.01. Survival curves were plotted with survminer version 0.4.6 (80).

Differential expression analysis

mRNA: Differential expression (DE) analysis was performed using two-tailed Welch Student's test corrected for multi-testing using Holm's method and t-scores were stored as a measure of hLMS/oLMS expression. Expression of the *MYOCD* gene was set to high if its expression was over the third quartile of the global expression distribution (from all genes in all samples) separately in each experiment.

ICGC miRNA : We applied the edgeR R package version 3.28.1 (81) on raw counts to normalize data and define significantly differentially expressed genes with a generalized linear model fitting (*glmQLFit*), correcting the returned p-values using Holm's method. We kept as expressed any miRNAs (*filterByExpr*) with a summed raw count over all samples > 10 and represented by a minimum of five reads in at least one sample.

TCGA miRNA: We computed the differential expression on TCGA miRNAs using the miRComb R package with *limma* method (82). We kept miRNAs with a median normalized count >1 in at least one of the LMS groups in both cohorts, having a hLMS/oLMS absolute logFC > 1, and a Holm's adjusted *limma* p-value < 0.01. We used the *lm* function from the stats R package to compute the R^2 value between ICGC and TCGA logFC.

Functional enrichment and mapping

Modules of co-expressed genes were analyzed using the *enricher* function from the ClusterProfiler R package version 3.14.3 (83) with the Molecular Signatures Database version v6 (MSigDB)(84). Significance threshold of hypergeometric test FDR adjusted p-value: < 0.05.

Differentially expressed genes were analyzed using command line version of GSEA software (version 3.0)(84) and MSigDB v6. We submitted the gene list ranked by hLMS/oLMS t-scores to the *xtools.gsea.GseaPreranked* function with default parameters. Significance threshold on permutation test FDR-adjusted p-value: 0.05. For the sake of clarity, only terms with adjusted p-values < 0.01 are reported. Enrichments for regulatory elements in groups of over- and under- expressed genes were performed on the iCistarget webserver (https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget/)(85). Significance threshold of normalized enrichment score (NES) > 3 by default. Annotations related to position weight matrix predictions from the same transcription factors were regrouped.

We used the GSAn webserver (https://gsan.labri.fr/)(19) with default parameters to exhaustively annotate genes with the most precise Gene Ontology term.

miRNA-mRNA interaction analysis

The MiRComb R package (82) was used to integrate miRNA, mRNA expression data and experimentally validated miRNA-mRNA interactions from miRecords v4 (86) and miRTarBase v7.0 (87). We retrieved 6262 known interactions which represent interactions between 30 DE pre-miRNAs and 3850 genes. Pre-miRNAs expression was estimated by averaging signals from derived mature miRNA. We kept interactions for which mRNA and miRNA had a hLMS/oLMS absolute log Fold Change (logFC) > 1, a *limma* p-value < 0.01, a significant Pearson's anti-correlation (adjusted p-value < 0.01) and were described in at least one of the two databases.

CNV recurrence analysis

Alteration recurrence was estimated by computing the frequency of each event (homozygous, heterozygous deletion, gain of one copy and amplification, *i.e.* gain of four copies or more), i.e. the number of a given event divided by the total number of patients (missing data being discarded). To evaluate alteration enrichment in each LMS group, losses were grouped in homo- and heterozygous deletions and gains, gains and amplifications (one-tailed Fisher's exact test p-value < 0.01). We computed enrichment for each type of event in the 291 cytobands by comparing, the number of significantly altered genes defined above for each band (one-tailed Fisher's exact test corrected with Holm's method < 0.01). If more than one type of event was enriched, the most significant was kept.

Mutational pattern analysis

We analyzed somatic mutations patterns using the MutationalPatterns R package version 1.12.0 (88). We first generated a 96 tri-nucleotide mutation count matrix per patient which we compared to the 30 COSMIC signatures v3.1 (89). We kept signatures with cosine similarity > 0.75 and computed the optimal contribution that best explained the observed mutational profiles in patients.

Tumor mutation burden was computed using the total number of somatic variants divided by the total length of human genome version hg19 (22 autosomal and 2 sexual chromosomes).

Data Availability

There are restrictions to the availability of human tumor samples due to patient consent. This study did not generate new unique plasmids or reagents.

Genomic and expression arrays for the other sample will be available on Gene Expression Omnibus (GEO) under accession GSE159847 and GSE159848 on the 2021-10-29.

ICGC cohort Whole-Genome sequencing and RNA sequencing data for the 67 LMS are available at https://dcc.icgc.org/projects/LMS-FR. miRNA data are available on Gene

Expression Omnibus (GEO) under accession GSE159849. Raw files will be available on the 2021-10-29 on Sequence Read Archive under accessions SRP288162.

Correspondence between previously published sample identifiers in Gene Expression Omnibus (GSE40021, GSE21050, GSE23980, GSE71118, GSE154591) and in ArrayExpress (E-MTAB-373) datasets and identifiers used in this paper is presented Table S5.

The code is under verification in OCEANCODE (90) (provisional DOI:10.24433/CO.0299110.v1)

Authors' Contributions:

Conceptualization, E.D, L.D, G.P, F.C; Methodology, E.D, F.C; Validation, E.D., L.D, G.P; Formal analysis, E.D, L.D, G.P, L.G, N.D, S.LG, L.L, C.M, F.C; Investigation, E.D, L.D, G.P; Resources, P.R, T.V, G.F, C.C, B.B, E.S, D. RV, P.M, P.T, S.PN, F.C, G.DP, F.D, J.M.C, J.Y.B ; Data Curation, E.D, G.P, L.D; Writing – Original Draft, E.D, L.D, G.P, F.C; Writing – Review & Editing, E.D, L.D, G.P, F.C; Visualization: E.D, L.D, G.P; Supervision: F.C; Funding Acquisition: J.M.C, J.Y.B, F.C, Other (data and expertise): The French Sarcoma Group and the International Cancer Genome Consortium

Acknowledgements

The authors would like to thank the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) for WG and RNA sequencing services and the Genomics Unit at the Centro de Regulación Genómica (CRG, Barcelona, Spain) for assistance with the smallRNAseq services. The results shown here are partly based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We are grateful to the French Sarcoma Group for tumor banks and associated clinical annotations and to Jean-Baptiste Courrèges. The following French cancer centers also participated in this study: Centre Paul Papin (Angers), Centre Oscar Lambert (Lille), Institut Paoli Calmettes (Marseille). Bioinformatics analyses were performed on the Core Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013).

Conflict of interest: The authors declare no potential conflicts of interest.

Financial support: The Instituts Thematiques Multiorganismes (ITMO) Cancer and the Claudius Regaud Institute supported this work.

Figure legends

Figure 1: Transcriptional analysis and patient classification: **A-B**. Heatmap showing clustering of 555 sarcoma patients and 455 genes (A) and 98 LMS patients and 1672 differentially expressed genes between hLMS and oLMS (B). Patients were clustered using HCPC method and genes are grouped by co-expression modules. Color scheme. Histotype: green forest: leiomyosarcomas, red: GIST, pink: undifferentiated sarcomas, orange: myxoid liposarcomas, blue: dedifferentiated sarcomas, grey: synovial sarcomas, turquoise: other sarcomas. Location: yellow: extremities, green: internal trunk, black: trunk wall. Sex: green: female, blue: male. Grade and differentiation: yellow: 1, green: 2, black: 3. Mitotic count: blue to red: from low to high: A – 0 to 120, B - 0 to 60. Cluster: green: hLMS, blue: oLMS. **C**. Kaplan-Meier metastasis-free survival analysis in hLMS and oLMS. Number indicates the Log-Rank test p-value. **D**. GSEA analysis on z-scores obtained from hLMS / oLMS gene expression comparison. Each dot is an enriched term (FDR < 0.01); size corresponds to number of genes involved; x-axis contains mean t-score of all genes annotated in given term and y-axis corresponds to GSEA NES score. Related terms colored the same way. **E**. i-Cistarget analysis of 843 under- and 800 over-expressed genes in hLMS relative to oLMS. The x-axis represents NES score obtained for over-expressed genes from 0 toward right and for under-expressed genes from 0 to left. The left and right parts are independent; the enriched features were clustered on the y-axis according to the cell type or tissue they were analyzed from. Histone modifications are only active marks of transcription (H3K4me1, H3K4me3, H3K27ac and H3K9ac). Detailed legends for E-F are available in **Table S1**. **F**. Top panel: distance distribution to centroids (x-axis) computed from transcriptional signature for ICGC and TCGA patients (bars on x-axis). Colors correspond to cluster assignation: patients with a distance lower than 0.6 to one of the centroids were assigned to corresponding centroid

(green: hLMS, dark blue: oLMS), while patients with intermediate value were not classified (light grey). Middle and bottom panels: PCA analysis using transcriptional signature genes in ICGC and TCGA cohorts. Each point is a patient, green: hLMS, dark blue: oLMS and light grey: not classified. X-axis and y-axis represent respectively principal components 1 and 2 and their associated representation of variance.

Figure 2: Analysis of miRNAs expression **A.** Left panel: PCA obtained from expression of 484 mature miRNAs in 39 ICGC patients (top) and of 475 mature miRNAs in 60 TCGA patients (bottom). Colors correspond to hLMS (green, ICGC: 28 patients, TCGA: 43 patients) and oLMS (blue, ICGC: 11 patients, TCGA: 17 patients). First two principal components are shown with percentage of variance they capture. Right panel: Scatterplot showing correlation between hLMS/oLMS LogFC in ICGC (y-axis) with TCGA (x-axis). Each dot represents a mature miRNA (347 expressed in both cohorts) and red color indicates 71 significant mature DE miRNAs in both cohorts. Line represents linear regression with interval confidence in shaded grey. **B**. HCPC clustering on the 41 mature miRNAs differentially expressed in LMS subtypes across the 9564 PANCAN samples. Heatmap showing median-centered miRNA expression (low: blue to high: red). Column annotation represents histotype of samples for which colors are specified at bottom of figure. Composition in histotype of clusters is detailed in bar plot below heatmap. The y-axis corresponds to the proportion. **C.** Heatmap showing differentially expressed miRNAs (rows) in ICGC (top, 55 miRNAs) and TCGA (bottom, 243 miRNAs). Column annotation corresponds to hLMS (green) and oLMS (blue). Expression values are median-centered (low: blue to high: red). Black rectangles highlight mature miRNAs from DIO3-DLK1 miRNA cluster (ICGC: 26, TCGA: 63, 25 in common). **D.** Functional terms mapped to 158 miRNA targeted genes. The x-axis indicates number of down-regulated (toward left in blue) and up-regulated (toward right in dark red if targeted

with only miRNAs from DIO3-DLK1 cluster, medium red if targeted by both miRNAs from DIO3-DLK1 cluster and other miRNAs and light red if targeted by other miRNAs) genes annotated with the term (y-axis).

Figure 3. Copy number analysis of 7479 genes in LMS merged cohort (84 Affymetrix (CGH), 39 ICGC (WGS) and 62 TCGA (CGH) patients). **A**. Heatmap showing copy number of genes (columns) in each patient (rows). Patients grouped according to LMS type (hLMS: green, oLMS: blue). Annotations above heatmap show chromosomes from 1 to 22 alternating grey and black and arms (p: green, q: blue). Annotation below shows significantly enriched events in hLMS, Fisher's Exact test p-value < 0.01. Color scheme is same for copy number and enrichment: homozygous deletion: dark blue, heterozygous deletion: light blue, normal: white, light red: gain of one copy, dark red: gain of 4 or more copies. **B**. Penetrance plot. Percentage (y-axis) of gain (red) and loss (blue) events are represented in hLMS (top panel) and oLMS (bottom panel). Each position on x-axis is a gene that corresponds to genes in A.

Figure 4: Zoom on genes of interest. **A.** Zoom on chr17p13-11/q11 genomic region (x-axis) penetrance profile containing *MYOCD*. Left y-axis indicates percentage of loss (light blue), deletion (dark blue), gain (light red) and amplification (dark red); right y-axis shows hLMS/oLMS median t-scores (from the three cohorts). Each gene represented by bar (penetrance) and dot (t-score). **B and D**. Violin plots showing *MYOCD* gene expression and *DMD* isoform expression (RNA level, y-axis) in hLMS and oLMS in the three cohorts respectively. **** indicates a t-test p-value $<10^{-7}$, *** p-value $< 10^{-3}$. **C**. Distribution of *TP53* and *RB1* allele status in hLMS and oLMS. Dot sizes correlate with percentage of patients in LMS group harboring defined status. Cases with biallelic inactivation of *TP53* are compared between groups (2L+2MS: only one mechanism altering both alleles *versus*

MS/FS+MS/L+FS/L: two different mechanisms altering each allele): red oval indicates Fisher test p-value < 0.01. L: loss, MS: missense, FS: false sense, MUT: MS or FS, WT: wild-type, if 2 is specified, both alleles are concerned. **E**. Cellular distribution of DMD, PTEN and ATRX proteins in hLMS and oLMS. For DMD, localization of its Dp427 isoform is also presented. A: absent, N: nuclear, M: membranous, C: cytoplasmic. Dot sizes correlate with percentage of patients in LMS group harboring defined localization. Red circle indicates Fisher's Exact test p-value < 0.01 (bold) and < 0.05 (thin).

Figure 5: SRF/MYOCD inhibitor can specifically target hLMS **A.** Distance to centroids determining h/oLMS status on 1672 genes of 3 LMS cell lines. **B**. Cytotoxicity curves of CCG-1423, inhibitor of SRF/MRTF axis, on 3 LMS and 2 UPS cell lines, using MTT assay after 72h of treatment at increasing concentrations (from 1.5 to 100µg/mL). First graph represents one of three experimentations used to determine IC50 with GraphPad. Second graph shows IC50 (mean ± s.d.; N = 3 independent assays). **C.** Cytotoxicity curves of CCG-100206, inhibitor of SRF/MYOCD axis, on same cell lines, using MTT assay after 72h of treatment at same increasing concentrations. First graph represents one of three experimentations used to determine IC50 with GraphPad. Second graph shows IC50 (mean ± s.d.; N = 3 independent assay). *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, p-value was calculated with unpaired t-test for **B** and **C**.

Tables

Table 1: Clinical enrichment and gene expression homogeneity in 42 hLMS vs 56 oLMS from Affymetrix cohort and 73 hLMS vs 29 oLMS in combined ICGC (28 vs 11) and TCGA (45 vs 18). (%) indicates that numbers in hLMS and oLMS columns are percentages of patients annotated with first feature (Well: well differentiated, Low: grade 1 + 2, F: female, Internal trunk). M: Male, other: Extremities, Trunk wall, limbs. The p-value was computed using Fisher's Exact test. Otherwise the median is reported, and the p-value was obtained with Wilcoxon's test. (+) next to p-values indicates a significant enrichment in hLMS while (-) indicates a significant enrichment in oLMS.

		Affymetrix			**ICGC + TCGA**		
Feature	**Test**	**hLMS**	**oLMS**	**p-value**	**hLMS**	**oLMS**	**p-value**
Differentiation (%)	Well (vs Poor)	88	24	3.87×10^{-9} (+)	84	41	1.83×10^{-5} (+)
Grade (%)	Low (vs High)	58	24	6.51×10^{-4} (+)			
Sex (%)	F (vs M)	76	48	0.003 (+)	68	41	0.007 (+)
Location (%)	Internal trunk (vs other)	60	7	8.48×10^{-9} (+)	82	27	1.53×10^{-7} (+)
Mitotic counts (median)	Ranks	17	24.5	0.009 (-)	11	35	0.0006 (-)
Gene expression variance (median)	Ranks	0.7	1	2.90×10^{-13} (-)	**ICGC**		
					0.25	0.36	$< 2.2 \times 10^{-16}$ (-)
					TCGA		
					0.45	0.9	$< 2.2 \times 10^{-16}$ (-)

Table 2: Summary of genetic alterations in 39 ICGC patients for *TP53*, *RB1*, *PTEN*, *ATRX* and *DMD*. Alterations are categorized as follows: mutation: missense, nonsense, frameshift, non-FS, splicing, SV: structural variant, loss: loss of at least one allele, total: number of patients carrying at least one alteration. Numbers indicate percentage of patients harboring the alteration; the actual numbers are reported between brackets.

Group	Alterations	*TP53*	*RB1*	*PTEN*	*ATRX*	*DMD*
hLMS	**mutation**	60.7 (17)	21.4 (6)	0 (0)	7.1 (2)	3.6 (2)
oLMS		18.2 (2)	9 (1)	0 (0)	27.3 (3)	9 (1)
all		48.7 (19)	17.9 (7)	0 (0)	12.8 (5)	5.1 (3)
hLMS	**SV**	25 (7)	35.7 (10)	3.6 (1)	7.1 (2)	14.3 (4)
oLMS		36.4 (4)	36.4 (4)	0 (0)	0(0)	36.4 (4)
all		28.2 (11)	35.9 (14)	2.6 (1)	5.1 (2)	20.5 (8)
hLMS	**loss**	89.3 (25)	92.9 (26)	82.1 (23)	7.1 (2)	3.6 (1)
oLMS		90.9 (10)	81.8 (9)	81.8 (9)	18.1 (2)	0 (0)
all		89.7 (35)	89.7 (35)	82 (32)	10.2 (4)	2.6 (1)
hLMS	**total**	100 (28)	100 (28)	82.1 (23)	21.4 (6)	17.8 (5)
oLMS		100 (11)	90.9 (10)	81.8 (9)	45.5 (5)	45.5 (5)
all		100 (39)	97.4 (38)	82 (32)	28.2 (11)	25.6 (10)

`

SUPPLEMENTAL EXCEL TABLE LEGENDS

Tables S1 related to Figure 1 and Figure 2: Results of functional annotations and enrichment in co-expressed modules, differentially expressed genes in Affymetrix cohort, top hundred most expressed genes in hLMSs and miRNA target genes.

Tables S2 related to Figure 2: Table summarizing differential expression analysis for mRNA in Affymetrix, ICGC and TCGA, for miRNA in ICGC and TCGA and miRComb miRNA-mRNA interaction analysis.

Tables S3 related to Figure 3: Table summarizing copy number alteration analysis by cytogenic band and by genes in merged cohort.

Tables S4 related to Figure 4: A. Table summarizing genetic alterations occurring for *TP53*, *RB1*, *PTEN*, *DMD*, *ATRX* and *MYOCD* in each sample of ICGC cohort as well as expression data of the six genes. **B.** Table presenting mutations detected in *TP53*, *RB1*, *DMD* and *ATRX*. **C.** Table presenting structural variants detected in *TP53*, *RB1*, *DMD*, *PTEN* and *ATRX*. **D.** Tables presenting primers used on genomic DNA to validate mutations and structural variants. **E.** Table describing primers used on cDNA to validate mutations and fusion transcripts.

Table S5 related to methods: Presentation of cohort of 555 samples analyzed on micro-arrays and 84 Affymetrix cohort samples for which copy number data were analyzed.

REFERENCES

1. WHO Classification of Tumours Editorial Board. *Soft Tissue and Bone Tumours [Internet]*. 2020:

2. Judson I et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. *The Lancet Oncology* 2014;15(4):415–423.

3. van der Graaf WT et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet* 2012;379(9829):1879–1886.

4. Ben-Ami E et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study: Nivolumab for Uterine Leiomyosarcoma. *Cancer* 2017;123(17):3285–3290.

5. Chibon F et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med* 2010;16(7):781–787.

6. Perou CM et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747–752.

7. Marisa L et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5):e1001453.

8. Nielsen TO et al. Molecular characterisation of soft tissue tumours: a gene expression study. *The Lancet* 2002;359(9314):1301–1307.

9. Baird K et al. Gene Expression Profiling of Human Sarcomas: Insights into Sarcoma Biology. *Cancer Res* 2005;65(20):9226–9235.

10. Perot G et al. Strong Smooth Muscle Differentiation Is Dependent on Myocardin Gene Amplification in Most Human Retroperitoneal Leiomyosarcomas. *Cancer Research* 2009;69(6):2269–2278.

11. Beck AH et al. Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling. *Oncogene* 2010;29(6):845–854.

12. Guo X et al. Clinically Relevant Molecular Subtypes in Leiomyosarcoma. *Clinical Cancer Research* 2015;21(15):3501–3511.

13. Lee Y-F et al. Gene expression profiling identifies distinct molecular subgroups of leiomyosarcoma with clinical relevance. *Br J Cancer* 2016;115(8):1000–1007.

14. Hemming ML et al. Oncogenic gene expression programs in leiomyosarcoma and characterization of conventional, inflammatory and uterogenic subtypes. *Mol Cancer Res* 2020;molcanres.0197.2020.

15. Gibault L et al. New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics. *J. Pathol.* 2011;223(1):64–71.

16. Italiano A et al. Genetic Profiling Identifies Two Classes of Soft-Tissue Leiomyosarcomas with Distinct Clinical Characteristics. *Clinical Cancer Research* 2013;19(5):1190–1196.

17. Watson S et al. Transcriptomic definition of molecular subgroups of small round cell sarcomas: Molecular classification of sarcoma subtypes. *J. Pathol* 2018;245(1):29–40.

18. Amant F et al. Clinical management of uterine sarcomas. *The Lancet Oncology* 2009;10(12):1188–1198.

19. Ayllon-Benitez A et al. GSAn: an alternative to enrichment analysis for annotating gene sets. *NAR Genomics and Bioinformatics* 2020;2(2):lqaa017.

20. The Cancer Genome Atlas Research Network et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* 2017;171(4):950–965.e28.

21. Cordes KR et al. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* 2009;460(7256):705–710.

22. Jiang Y, Yin H, Zheng X-L. MicroRNA-1 inhibits myocardin-induced contractility of human vascular smooth muscle cells. *J. Cell. Physiol.* 2010;225(2):506–511.

23. Almeida MI et al. Strand-Specific miR-28-5p and miR-28-3p Have Distinct Effects in Colorectal Cancer Cells. *Gastroenterology* 2012;142(4):886–896.e9.

24. Zhou C et al. Mir-455-3p-1 represses FGF7 expression to inhibit pulmonary arterial hypertension through inhibiting the RAS/ERK signaling pathway. *Journal of Molecular and Cellular Cardiology* 2019;130:23–35.

25. Hashemi Gheinani A et al. MicroRNA MiR-199a-5p Regulates Smooth Muscle Cell Proliferation and Morphology by Targeting WNT2 Signaling Pathway. *J. Biol. Chem.* 2015;290(11):7067–7086.

26. Kim J et al. An endothelial apelin-FGF link mediated by miR-424 and miR-503 is disrupted in pulmonary arterial hypertension. *Nat Med* 2013;19(1):74–82.

27. Darmusey L et al. ATRX alteration contributes to tumor growth and immune escape in pleomorphic sarcomas [Internet]. *BioRxiv* [published online ahead of print: 2020]; doi:10.1101/2020.10.23.352112

28. Horita H et al. Nuclear PTEN functions as an essential regulator of SRF-dependent transcription to control smooth muscle differentiation. *Nat Commun* 2016;7(1):10830.

29. Evelyn CR et al. CCG-1423: a small-molecule inhibitor of RhoA transcriptional signaling. *Mol Cancer Ther* 2007;6(8):2249–2260.

30. Zhou N et al. Rho Kinase Regulates Aortic Vascular Smooth Muscle Cell Stiffness Via Actin/SRF/Myocardin in Hypertension. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2017;44(2):701.

31. Miettinen M. Smooth muscle tumors of soft tissue and non-uterine viscera: biology and prognosis. *Modern Pathology* 2014;27(S1):S17–S29.

32. Miano JM. Myocardin in biology and disease. *J Biomed Res* 2015;29(1):3–19.

33. Miano J. Serum response factor: toggling between disparate programs of gene expression. *Journal of Molecular and Cellular Cardiology* 2003;35(6):577–593.

34. Treisman R. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* 1986;46(4):567–574.

35. Turczyńska KM et al. Regulation of Smooth Muscle Dystrophin and Synaptopodin 2 Expression by Actin Polymerization and Vascular Injury. *Arterioscler Thromb Vasc Biol.* 2015;35(6):1489–1497.

36. Wang Y et al. Dystrophin is a tumor suppressor in human cancers with myogenic programs. *Nat. Genet.* 2014;46(6):601–606.

37. Lartigue L et al. Genome remodeling upon mesenchymal tumor cell fusion contributes to tumor progression and metastatic spread. *Oncogene* 2020;39(21):4198–4211.

38. Yadav RK et al. Endoplasmic Reticulum Stress and Cancer. *J Cancer Prev* 2014;19(2):75–88.

39. Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol. Rev.* 2004;84(3):767–801.

40. Vacante F et al. The function of miR-143, miR-145 and the MiR-143 host gene in cardiovascular development and disease. *Vascular Pharmacology* 2019;112:24–30.

41. Geraldo MV, Nakaya HI, Kimura ET. Down-regulation of 14q32-encoded miRNAs and tumor suppressor role for *miR-654-3p* in papillary thyroid cancer. *Oncotarget* 2017;8(6):9597–9607.

42. Benetatos L et al. The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cell. Mol. Life Sci.* 2013;70(5):795–814.

43. Alexander MR, Owens GK. Epigenetic Control of Smooth Muscle Cell Differentiation and Phenotypic Switching in Vascular Development and Disease. *Annu. Rev. Physiol.* 2012;74(1):13–40.

44. Cuppens T et al. Potential Targets' Analysis Reveals Dual PI3K/mTOR Pathway Inhibition as a Promising Therapeutic Strategy for Uterine Leiomyosarcomas—an ENITEC Group Initiative. *Clin Cancer Res* 2017;23(5):1274–1285.

45. Fourneaux B et al. Dual inhibition of the PI3K/AKT/mTOR pathway suppresses the growth of leiomyosarcomas but leads to ERK activation through mTORC2: biological and clinical implications. *Oncotarget* 2016;8(5):7878–7890.

46. Gibault L et al. From PTEN loss of expression to RICTOR role in smooth muscle differentiation: complex involvement of the mTOR pathway in leiomyosarcomas and pleomorphic sarcomas. *Mod Pathol* 2012;25(2):197–211.

47. Lesluyes T et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur. J. Cancer* 2016;57:104–111.

48. Lagarde P et al. Mitotic Checkpoints and Chromosome Instability Are Strong Predictors of Clinical Outcome in Gastrointestinal Stromal Tumors. *Clinical Cancer Research* 2012;18(3):826–838.

49. Lagarde P et al. Chromosome Instability Accounts for Reverse Metastatic Outcomes of Pediatric and Adult Synovial Sarcomas. *JCO* 2013;31(5):608–615.

50. Fletcher C et al. *WHO Classification of Tumours of Soft Tissue and Bone*. Lyon: IARC Press; 2013:

51. Lagarde P et al. Stable Instability of Sarcoma Cell Lines Genome Despite Intra-Tumoral Heterogeneity: A Genomic and Transcriptomic Study of Sarcoma Cell Lines. *Austin J Genet Genomic Res* 2015;2(2):1014.

52. McPherson A et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput. Biol.* 2011;7(5):e1001138.

53. Delespaul L et al. Recurrent TRIO Fusion in Nontranslocation-Related Sarcomas. *Clin. Cancer Res*. 2017;23(3):857–867.

54. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j*. 2011;17(1):10.

55. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 2011;39(Database):D152–D157.

56. Tam S, Tsao M-S, McPherson JD. Optimization of miRNA-seq data preprocessing. *Briefings in Bioinformatics* 2015;16(6):950–963.

57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754–1760.

58. García-Alcalde F et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 2012;28(20):2678–2679.

59. Li H et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078–2079.

60. Joshi N, Fass J. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Internet]*. 2011:

61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012;9(4):357–359.

62. Broad Institute. *Picard Toolkit [Internet]*. Broad Institute, GitHub repository; 2019:

63. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011;27(21):2987–2993.

64. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.

65. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526(7571):68–74.

66. Lesluyes T et al. Genomic and transcriptomic comparison of post-radiation versus sporadic sarcomas. *Mod Pathol* 2019;32(12):1786–1794.

67. Klambauer G et al. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Research* 2012;40(9):e69–e69.

68. Carlson M. *TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s*. 2015:

69. Robinson JT et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29(1):24–26.

70. Untergasser A et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 2012;40(15):e115.

71. Manie E et al. High Frequency of TP53 Mutation in BRCA1 and Sporadic Basal-like Carcinomas but not in BRCA1 Luminal Breast Tumors. *Cancer Research* 2009;69(2):663–671.

72. Bolstad B. *preprocessCore: A collection of pre-processing functions [Internet]*. 2019:

73. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006;Complex Systems:1695.

74. Lê S, Josse J, Husson F. **FactoMineR**: An *R* Package for Multivariate Analysis [Internet]. *J. Stat. Soft.* 2008;25(1). doi:10.18637/jss.v025.i01

75. Kolde R. *Pheatmap: pretty heatmaps [Internet]*. 2019:

76. Vu VQ. *ggbiplot: A ggplot2 based biplot [Internet]*. 2011:

77. van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. *Mach Learn* 2012;87(1):33–55.

78. Scrucca L et al. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal* 2016;8(1):289.

79. Therneau TM. *A Package for Survival Analysis in R [Internet]*. 2020:

80. Kassambara A, Kosinski M, Biecek P. *survminer: Drawing Survival Curves using "ggplot2" [Internet]*. 2019:

81. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–140.

82. Vila-Casadesús M, Gironella M, Lozano JJ. MiRComb: An R Package to Analyse miRNA-mRNA Interactions. Examples across Five Digestive Cancers. *PLoS ONE* 2016;11(3):e0151127.

83. Yu G et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* 2012;16(5):284–287.

84. Mootha VK et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34(3):267–273.

85. Imrichová H et al. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res* 2015;43(W1):W57–W64.

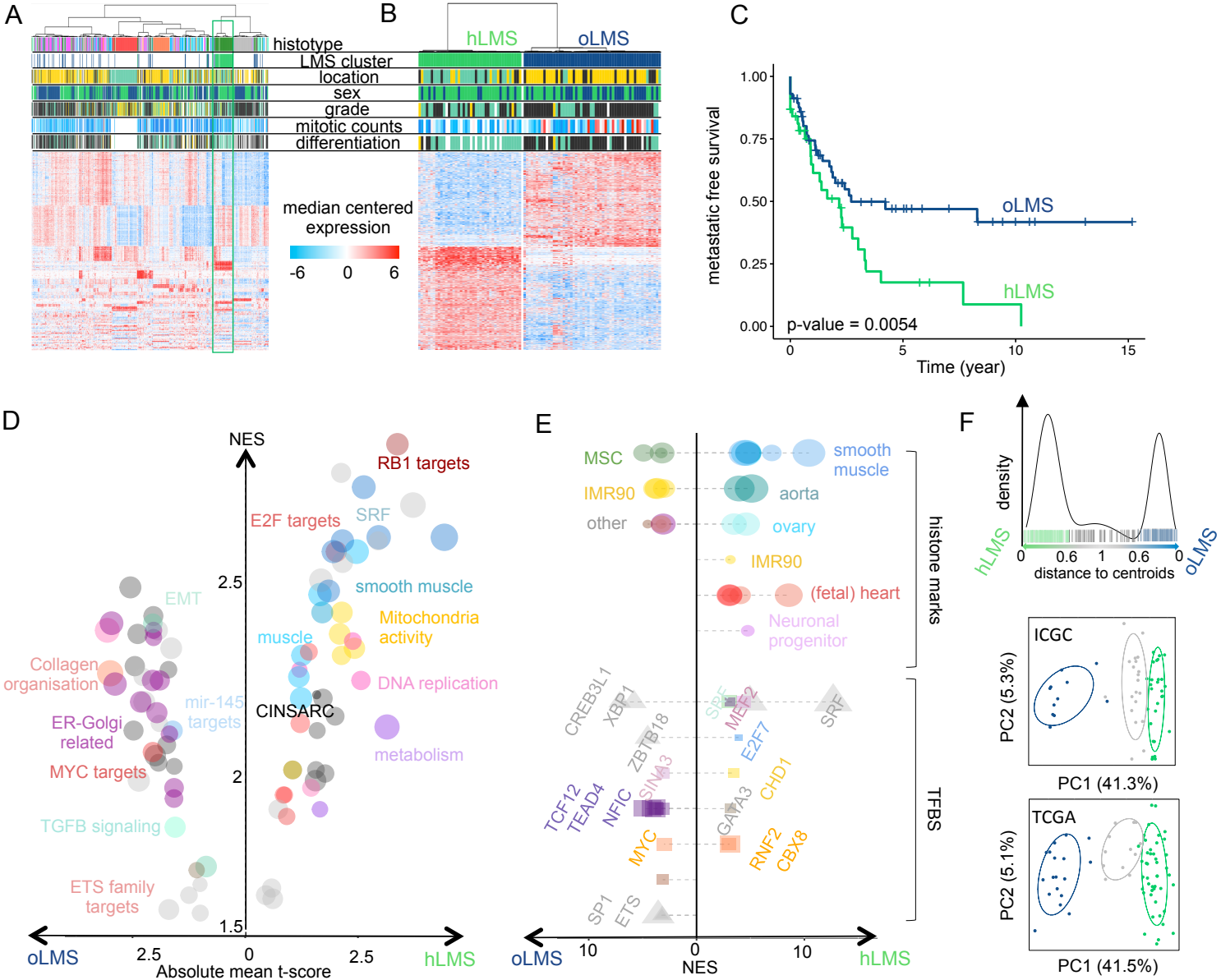
86. Xiao F et al. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research* 2009;37(Database):D105–D110.

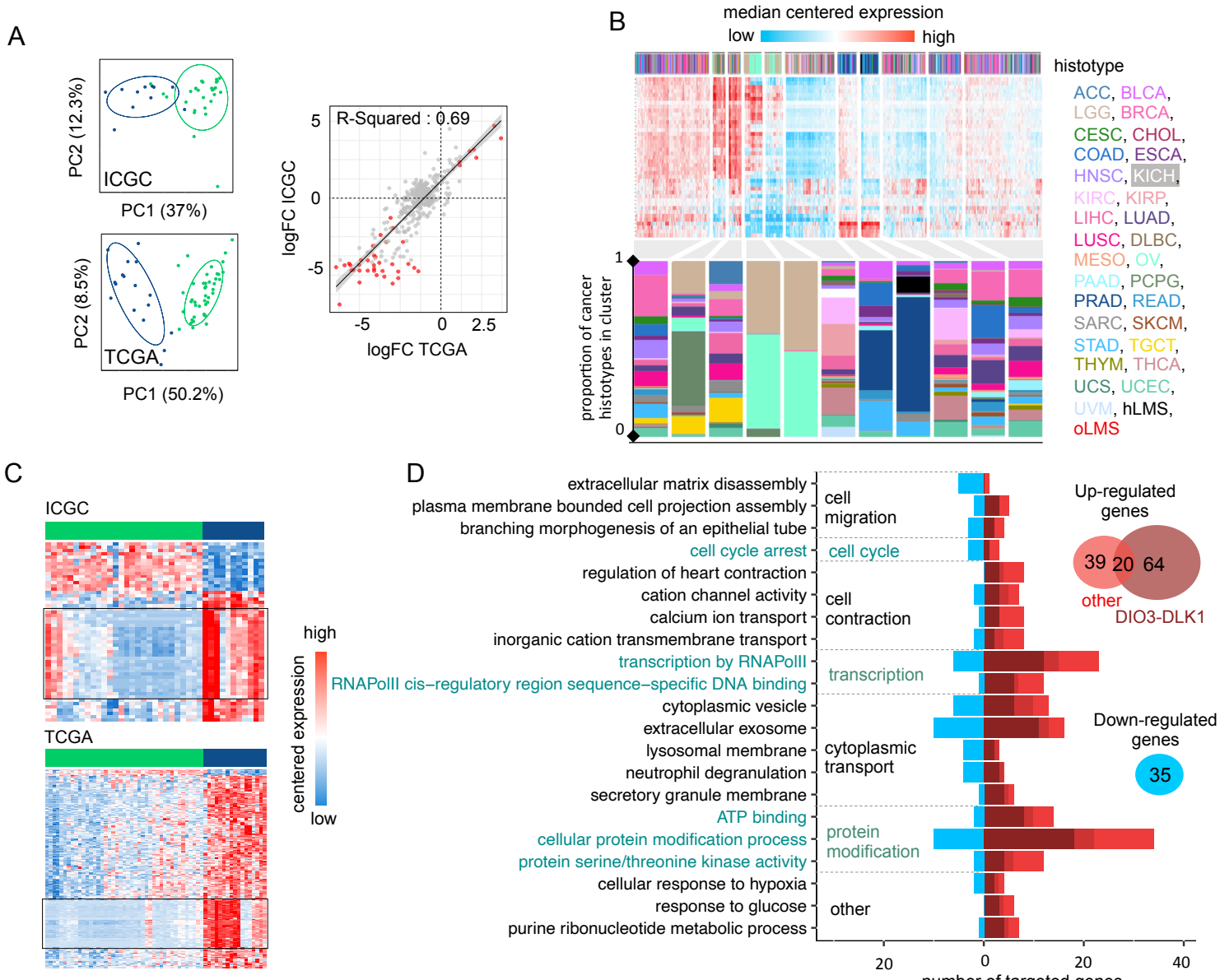
87. Chou C-H et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* 2018;46(D1):D296–D302.

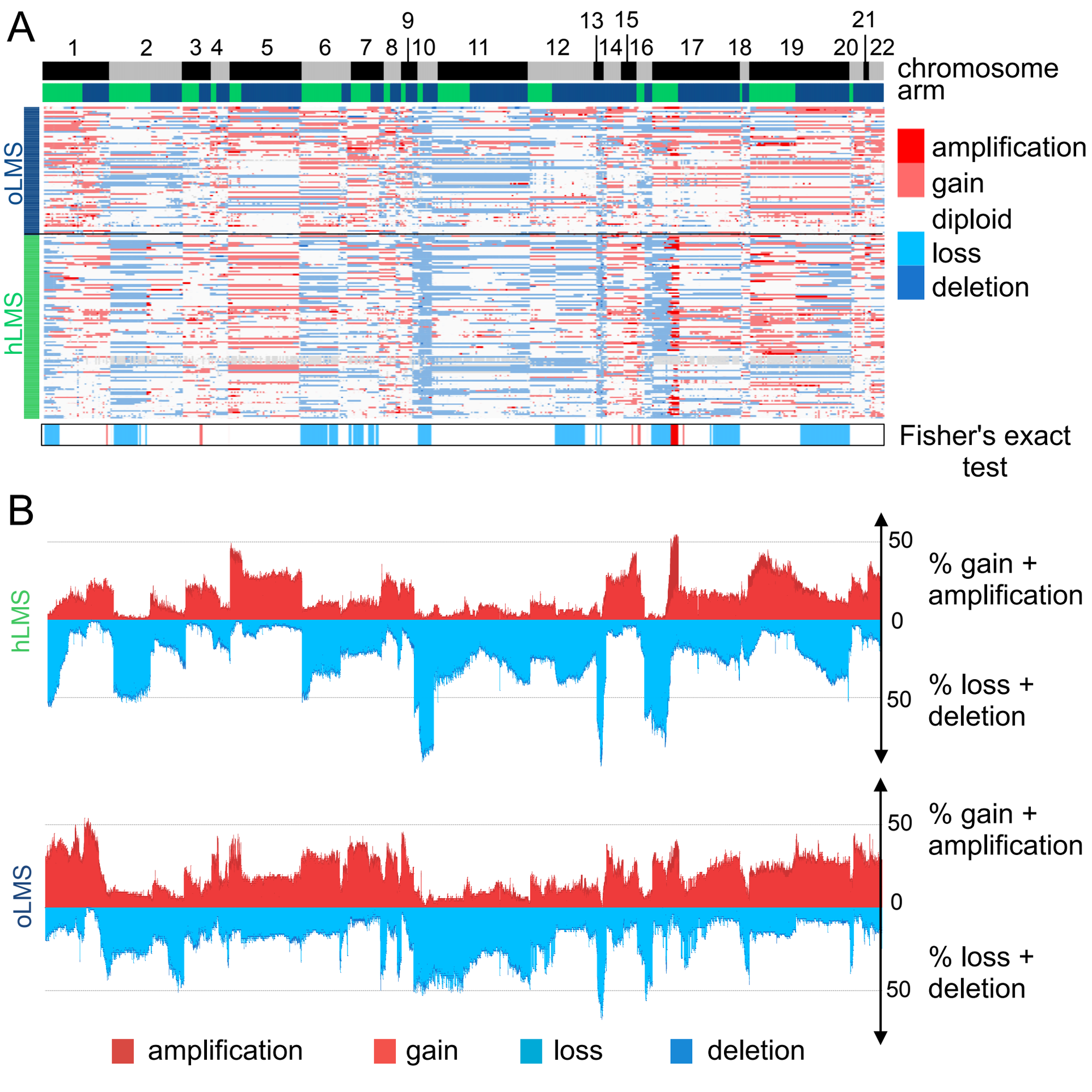
88. Blokzijl F et al. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* 2018;10(1):33.

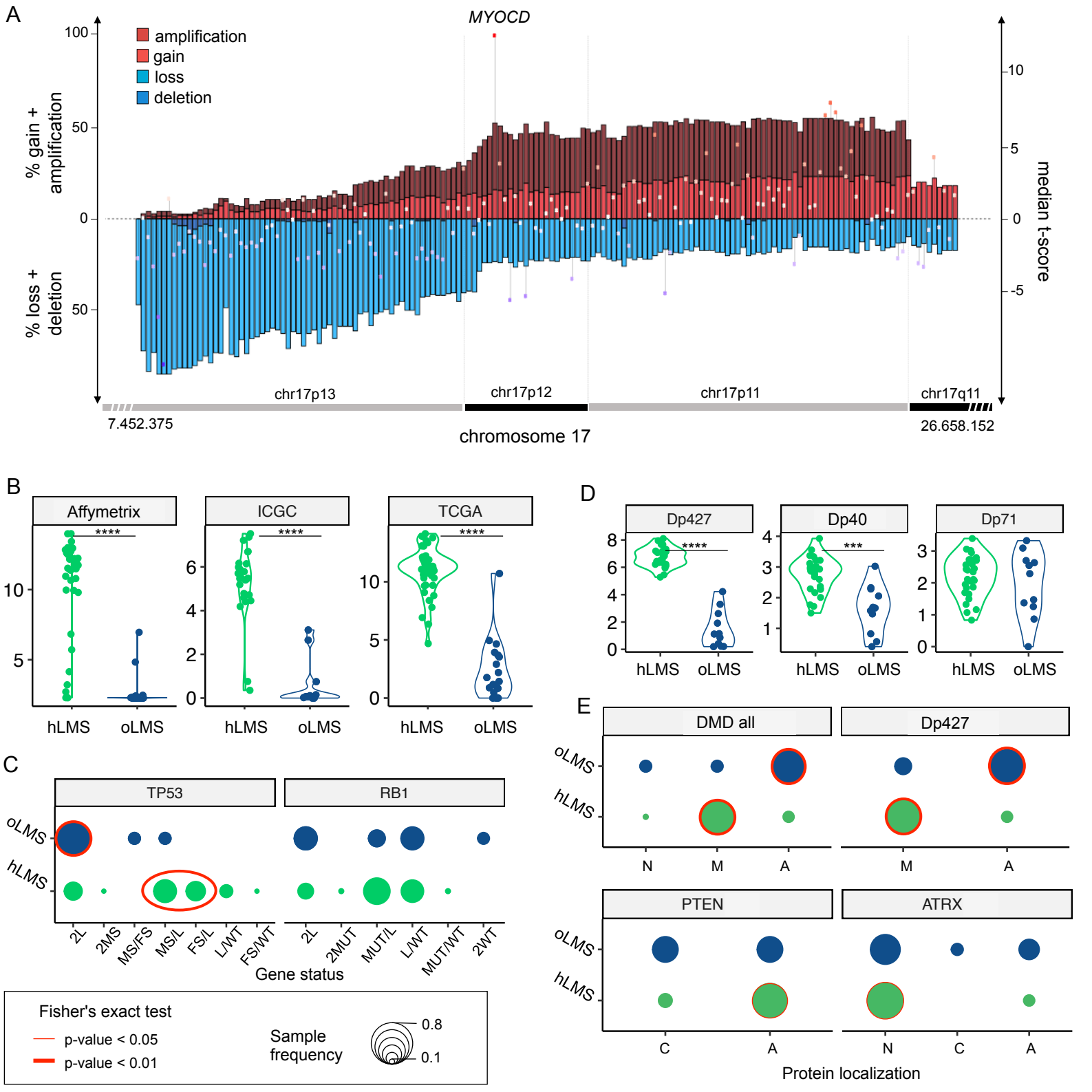
89. Tate JG et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research* 2019;47(D1):D941–D947.

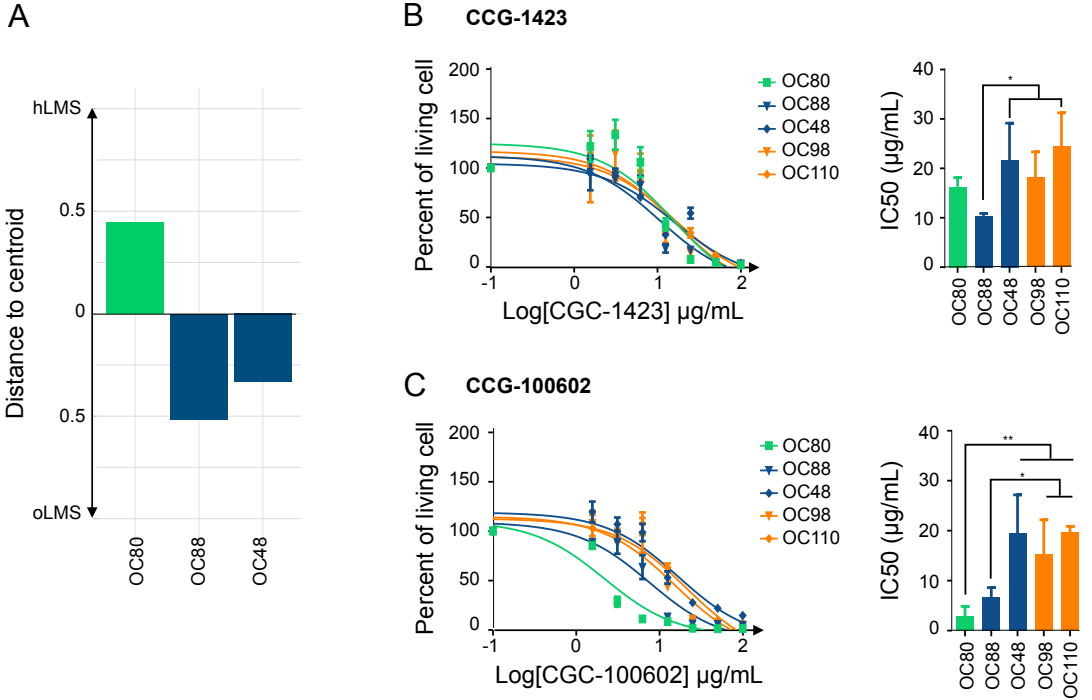
90. Clyburne-Sherin A, Fei X, Green SA. Computational Reproducibility via Containers in Psychology [Internet]. *Meta-Psychology* 2019;3. doi:10.15626/MP.2018.892


A
B
C
D
E
F
histotype
LMS cluster
location
sex
grade
mitotic counts
differentiation
hLMS
oLMS
median centered expression
-6
0
6
metastatic free survival
1.00
0.75
0.50
0.25
0.00
p-value = 0.0054
Time (year)
NES
RB1 targets
E2F targets
SRF
smooth muscle
Mitochondria activity
muscle
DNA replication
CINSARC
metabolism
EMT
Collagen organisation
mir-145 targets
ER-Golgi related
MYC targets
TGFB signaling
ETS family targets
Absolute mean t-score
MSC
IMR90
other
smooth muscle
aorta
ovary
IMR90
(fetal) heart
Neuronal progenitor
histone marks
CREB3L1
XBP1
ZBTB18
SINA3
TCF12
TEAD4
NFIC
MYC
SP1
ETS
SRF
MEF2
E2F7
CHD1
GATA3
RNF2
CBX8
TFBS
density
distance to centroids
ICGC
TCGA
PC1 (41.3%)
PC2 (5.3%)
PC1 (41.5%)
PC2 (5.1%)


A
PC2 (12.3%)
ICGC
PC1 (37%)
PC2 (8.5%)
TCGA
PC1 (50.2%)
R-Squared : 0.69
logFC ICGC
logFC TCGA
5
0
-5
-5
0
2.5
B
median centered expression
low
high
histotype
ACC, BLCA, LGG, BRCA, CESC, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, DLBC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THYM, THCA, UCS, UCEC, UVM, hLMS, oLMS
proportion of cancer histotypes in cluster
1
0
C
ICGC
TCGA
high
centered expression
low
D
extracellular matrix disassembly
plasma membrane bounded cell projection assembly
branching morphogenesis of an epithelial tube
cell cycle arrest
regulation of heart contraction
cation channel activity
calcium ion transport
inorganic cation transmembrane transport
transcription by RNAPolII
RNAPolII cis–regulatory region sequence–specific DNA binding
cytoplasmic vesicle
extracellular exosome
lysosomal membrane
neutrophil degranulation
secretory granule membrane
ATP binding
cellular protein modification process
protein serine/threonine kinase activity
cellular response to hypoxia
response to glucose
purine ribonucleotide metabolic process
cell migration
cell cycle
cell contraction
transcription
cytoplasmic transport
protein modification
other
Up-regulated genes
39
20
64
other
DIO3-DLK1
Down-regulated genes
35
20
0
20
40
number of targeted genes


A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
chromosome arm
oLMS
hLMS
amplification
gain
diploid
loss
deletion
Fisher's exact test
B
hLMS
50
0
50
% gain + amplification
% loss + deletion
oLMS
50
0
50
% gain + amplification
% loss + deletion
amplification
gain
loss
deletion


A
MYOCD
amplification
gain
loss
deletion
% gain + amplification
% loss + deletion
median t-score
chr17p13
chr17p12
chr17p11
chr17q11
7.452.375
26.658.152
chromosome 17
B
Affymetrix
ICGC
TCGA
hLMS
oLMS
C
TP53
RB1
oLMS
hLMS
2L
2MS
MS/FS
MS/L
FS/L
L/WT
FS/WT
2MUT
MUT/L
MUT/WT
2WT
Gene status
Fisher's exact test
p-value < 0.05
p-value < 0.01
Sample frequency
0.8
0.1
D
Dp427
Dp40
Dp71
E
DMD all
PTEN
ATRX
N
M
A
C
Protein localization


A
hLMS
oLMS
Distance to centroid
0.5
0
0.5
OC80
OC88
OC48
B
CCG-1423
Percent of living cell
Log[CGC-1423] µg/mL
OC80
OC88
OC48
OC98
OC110
IC50 (µg/mL)
*
C
CCG-100602
Percent of living cell
Log[CGC-100602] µg/mL
OC80
OC88
OC48
OC98
OC110
IC50 (µg/mL)
**
*